Locus-wide studies into the transcriptional regulation of *Runx1* in developmental hematopoiesis

Andrew Jarratt

Trinity College

MRC Molecular Hematology Unit
Weatherall Institute of Molecular Medicine

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy
University of Oxford

Trinity Term 2011

Locus-wide studies into the transcriptional regulation of *Runx1* in developmental hematopoiesis

Andrew Jarratt (Trinity College)

Doctor of Philosophy - Trinity Term 2011

Abstract

Developmental hematopoiesis sees the generation of the first blood cells and definitive blood during embryonic development. The founding cell of definitive hematopoiesis, the hematopoietic stem cell (HSC), gives rise to all adult blood lineages throughout the life span of an organism. It is expected that future *ex-vivo* manipulation of HSCs for therapeutic uses will benefit from a thorough understanding of the mechanisms, both cellular and genetic, that give rise to HSCs. One of the most critical regulators of HSC emergence in the embryo is the transcription factor (TF) Runx1. One aim of our lab is to decipher what controls the *cis*-regulation of *Runx1* to understand better how it exerts its function in the emergence of HSCs. In this thesis, chromatin assays were used to identify putative enhancers within the 1.3 Mb *Runx1* syntenic region. Seven novel enhancers were identified that mediate reporter gene expression in discrete patterns of *Runx1*-specific hematopoietic expression in transient transgenic embryos. Characterization of the cells marked by one of these enhancers, the +110 enhancer in a transgenic mouse line, showed that it is active in clonogenic progenitors at E11.5, but, interestingly, not HSCs. Finally, chromosome conformation capture (3C) assays showed physical interactions between the *Runx1* P1 and P2 promoters and between the *Runx1* P1 and P2 promoters and putative regulatory elements in the 1.3 Mb syntenic region. Together, these data increase our understanding of the complexity of *Runx1 cis*-regulation during development and provide a starting point for characterizing what upstream *trans*-acting factors converge on *Runx1* to specify blood.

Acknowledgements

This thesis is dedicated to the loving memory of Dr. Thomas Bee, my best friend.

First and foremost, I would like to thank my supervisor, Dr. Marella de Bruijn for everything she has done for me over the past years. As an employer, supervisor and guidance councillor (!), she has always supported me and helped me to achieve my potential. For this I will always be thankful. I would also like to thank my parents, Peta and Dave, my brothers Michael and Nicholas, and my best friends Megan, Matthew, Anneliese, Louise, Tessa, Eoin, Kirsti-Ann, Catherine, Muriel, Mark and Bethan for their constant love and encouragement particularly during the inevitable 'difficult days'. Special mention must also go to Dr. Wade Nottingham and Dr. Alix Kwasniewska who have been my friends at the WIMM since day one. I would also like to thank the many members of the de Bruijn lab, past and present, including my students Jodie, Chris and Adam who all helped contribute to the work in this thesis and Dr. Gemma Swiers and John O'Rourke for crucial help with mouse experiments and for keeping me on the right track! As for other people at the WIMM, I would like to thank Drs. Doug Vernimmen, Jim Hughes and Marco de Gobbi for teaching me everything I know about chromatin assays. In addition the help of Jackie Sloane-Stanley, Pik-Shan Li and Chrissy Rode in generating transgenic embryos and Kevin Clark for FACS experiments, cannot go without mention. I would not have been able to do this work without their help. Finally, I would like to give a big thanks to members of the Porcher and Vyas labs, particularly Julia, Angela, Amy, Sarah, Terri, Nicolas and Kate for all the tea-break and lunch time chats and many a good night at the pub.

List of Abbreviations

AC = Angiogenic cord

AGM = Aorta Gonad Mesenephros

AVU = Aorta + Vitelline and Umbilical arteries

BA = Branchial arches

BI = Blood Island

BMP = Bone morphogenic protein

3C = Chromosome conformation capture

cDNA = Complementary DNA

CFU-C = Colony forming unit in culture

CFU-S = Colony-forming unit-spleen

ChIP = Chromatin Immunoprecipitation

ChIP-chip = Chromatin Immunoprecipitation on microarray

ChIP-seq = Chromatin Immunoprecipitation followed by high-throughput sequencing

CML = Chronic myelogenous leukemia

CNE = Conserved non-coding element

CNS = Central nervous system

DA = Dorsal Aorta

DAPI = 4',6-diamidino-2-phenylindole

Dhh = Desert hedgehog

DMEM = Dulbecco's modified eagle medium

DNA = Deoxyribonucleic acid

DNAseI = Deoxyribonuclease I

DRG = Dorsal root ganglia

EE = Embryo equivalent

EHT = Endothelial to hematopoietic transition

Endo. = Endoderm

EryP = Primitive erythrocyte

ES = Embryonic stem

EtOH = Ethanol

EtBr = Ethidium bromide

F0 embryo = transient (harvested before birth)

FACS = Fluorescence activated cell sorting

FCS = Fetal calf serum

FL = Fetal liver

FLV = Friend Leukemia Virus

FRP = Forward primer + Reverse Primer + Probe

gDNA = Genomic DNA

GFP = Green fluorescent protein

GS = Gemma Swiers

HE = Hemogenic endothelium

HMBA = Hexamethylene bisacetamide

HP = Hematopoietic progenitor

HSC = Hematopoietic stem cell

Ihh = Indian hedgehog

I-MEL = Induced mouse erythroleukaemia

IMDM = Iscove's Modified Dulbecco's Media

JOR = John O'Rourke

JSS = Jackie Sloane-Stanley

KI = Knock-in

KO = Knock-out

LacZ = β-galactosidase

LB = Lysogeny broth

LF = Long fragment

LH2 = LIM-homeobox gene 2

LIF = Leukaemia inhibitory factor

MdB = Marella de Bruijn

MEL = Mouse erythroleukaemia

MEP = Megakaryocyte-erythroid progenitor

mES = Mouse embryonic stem

MTG = Monothioglycerol

O/N = Overnight

PAS = Para-aortic splanchnopleura

PB = Peripheral blood

PBS = Phosphate buffered saline

PCR = Polymerase chain reaction

Ptch = Patched

qPCR = Quantitative real-time PCR

RFP = Red fluorescent protein

RNA = Ribonucleic acid

RNase = Ribonuclease

RPM = Revolutions per minute

RPMI = Roswell park memorial institute medium

RT = Room temperature

RUNX1 = human gene

RUNX1 = human protein

Runx1 = mouse gene

Runx1 = mouse protein

runx1 = zebrafish gene

SDS = Sodium dodecyl sulfate

Shh = Sonic hedgehog

Sih = Silent heart

Smo = Smoothend

SNAP = S-nitroso-N-acetyl- penicillamine

Sp = Somite pair

TBE = Tris/borate/EDTA buffer

TF = Transcription factor

TG = Transgenic

TPO = Thrombopoietin

TSS = Transcription start site

UCSC = University of California at Santa Cruz

U-MEL = Un-induced mouse erytholeukaemia cell

UV = Ultraviolet

VA = Vitelline artery

VEGF = Vascular endothelial growth factor

VU = Vitelline + Umbilical arteries

WT = Wild type

YS = Yolk sac

List of Tables

Table 2.1: Primers used for generating sequences for cloning 43

Table 2.2: Primers used for genotyping 47

Table 2.3: Antibodies used in ChIP experiments 56

Table 2.4: Taqman probes and primers used in ChIP qPCR reactions 57

Table 2.5: Mouse α-globin Taqman probes and primers used in ChIP experiments 58

Table 2.6: Coordinates (UCSC build NCBI37/mm9) within which probes for each gene indicated were selected on 2 x 105K Agilent ChIP-chip custom arrays 60

Table 2.7: Coordinates (UCSC build NCBI37/mm9) within which probes for each gene indicated were selected on 4 x 44K Agilent ChIP-chip custom arrays. 61

Table 2.8: Mouse *α-globin* and *xpb* Taqman probes and primers used in 3C experiments. 66

Table 2.9: Antibodies used in FACS experiments 70

Table 3.1: Summary of enrichment of H3K4Me2/3 peaks in E11.5 AGM, E12.5 FL and 416B cells and DNaseI-seq peaks in 416B cells at 15 selected regions thoughout the *Runx1* syntenic region 92

Table 3.2: Summary of binding of 10 hematopoietic transcription factors to 15 selected regions thoughout the *Runx1* syntenic region in HPC7 cells. 92

Table 4.1: Summary of expression seen in sections of transient transgenic (F0) hsp68*LacZ*-enhancer embryos 111

Table of Contents

Abstract ..... **ii**

Acknowledgements ..... **iii**

List of Abbreviations ..... **iv**

List of Tables ..... **viii**

Chapter One: Introduction **1**

1.1 Adult Hematopoiesis 2

1.2 Developmental Hematopoiesis 5

1.2.1 Primitive Hematopoiesis 5

1.2.2 Transient definitive hematopoiesis 8

1.2.3 Definitive Hematopoiesis 9

1.3 Is there a common progenitor of primitive and definitive blood? 10

1.4 What is the cellular origin of the HSC? Hemangioblast vs. hemogenic endothelium 13

1.5 The role of Runx1 in developmental hematopoiesis 19

1.5.1. The Runx Transcription Factors 19

1.5.2. The roles of Runx1, Runx2 and Runx3 in development 20

1.5.3. Runx1 activity and expression during developmental hematopoiesis 21

1.5.4. The transcriptional regulation of Runx1 during developmental hematopoiesis 26

1.6 What upstream factors regulate *Runx1* activity? 29

1.6.1 Signalling Pathways and Transcription Factors 29

1.6.2 Notch Signalling 32

1.6.3 Notch signalling and Gata2 33

1.6.4. The role of Scl 34

1.6.5 The Hedgehog, VEGF and BMP signalling pathways 35

1.6.6 The effect of blood flow on Runx1 expression 37

1.7 Chapter Summary 38

1.8 Project Aim 38

Chapter Two: Materials and Methods 40
2.1 Bioinformatics 41
2.1.1 Comparative Genomics 41
2.1.2 ChIP-Chip, ChIP-Seq and DNaseI data 41
2.2 DNA preparation and manipulation 41
2.2.1 Polymerase Chain Reaction (PCR) 41
2.2.2 Ligation of DNA into vectors 42
2.2.3 Transformation of bacteria 42
2.2.4 Growth of bacterial colonies for DNA isolation 42
2.2.5 Digestion of DNA with restriction enzymes 44
2.2.6 Determination of DNA fragment size by agarose gel electrophoresis 44
2.2.7 DNA extraction from agarose gels 45
2.2.8 DNA sequencing 45
2.2.9 Cloning of DNA fragments into reporter constructs 45
2.2.10 Genotyping using adult mouse and embryonic tissues 46
2.2.11 Isolation of Genomic DNA 46
2.2.12 RNA isolation 48
2.2.13 cDNA synthesis 49
2.3 Cell Culture 50
2.3.1 Mouse 416B myeloid cell line culture 50
2.3.2 Human K562 erythroleukemia cell line culture 50
2.3.3 Mouse HPC7 ES-derived progenitor cell line culture 51
2.3.4 Mouse erythroleukemia (MEL) cell line culture 51
2.3.5 J1 mESC cell line culture 51
2.3.6 Megakaryocyte differentiation from E14.5 mouse fetal livers 52
2.4 Luciferase Assay 53
2.4.1 Electroporation of cells 53
2.4.2 Analysis of luciferase activity 54
2.5 Chromatin Techniques 54
2.5.1 Chromatin Immunoprecipitation (ChIP) 54
2.5.2 Quantitative Real-Time PCR (qPCR) 55
2.5.3 Design of custom Agilent ChIP-chip microarrays 59

2.5.4 Labelling and amplification of ChIP material 59
2.5.5 ChIP-chip Hybridization 62
2.5.6 Chromosome Conformation Capture (3C) 63
2.6 Analysis of 3C products 64
2.6.1 TBE gel electrophoresis 64
2.6.2 3C qPCR 64
2.7 Mice 65
2.7.1 Generation and analysis of LacZ-reporter F0 transgenic embryos 65
2.7.2 Generation and analysis of GFP-reporter transgenic embryos 67
2.7.3 Animal husbandry 67
2.7.4 Preparation of embryonic tissues for cell sorting 68
2.7.5 Preparation of adult tissues for cell sorting 69
2.7.6 Flow Cytometry 69
2.7.7 Methylcellulose colony assay (CFU-C) 69
2.7.8 Long-term reconstitution analysis 71
2.8 Statistics 71

Chapter Three: Identification and initial analysis of candidate *Runx1 cis*-regulatory elements 72
3.1 Chapter Aim 73
3.2 The identification and characterization of *Runx1 cis*-regulatory elements 73
3.2.1 The +23 enhancer was found in a screen of a 270 Kb genomic region spanning the Runx1 locus 73
3.2.2 Identifying novel Runx1 enhancers: a 1.3 Mb screen in the Runx1 syntenic region 76
3.2.3 Histone modifications at known Runx1 CNEs 77
3.2.4 Interim Conclusions 81
3.3 ChIP-chip: a screen for novel *cis*-regulatory elements 83
3.3.1 Establishing the ChIP-chip method 83
3.3.2 Additional control experiments prior to Runx1 ChIP-chip 84
3.3.3 Runx1 ChIP-chip, ChIP-seq and DNaseI-seq 85

3.4 *In vitro* enhancer activity of novel elements within the *Runx1* sytenic region 94
3.5 Chapter Discussion 98

Chapter Four: *In vivo* analysis of candidate *Runx1 cis*-regulatory elements 100
4.1 Chapter Aim 101
4.2. Generation of transient transgenic (F0) embryos 103
4.3. Analysis of reporter gene expression in hsp68*LacZ*+3 F0 transgenic embryos 109
4.4. Analysis of +110 activity in hsp68*LacZ*+110 and hsp68*GFP*+110 F0 transgenic embryos 110
4.5. Analysis of reporter gene expression in hsp68*LacZ*+205 F0 transgenic embryos 122
4.6 Analysis of reporter gene expression in hsp68*LacZ*-480-450-370 F0 transgenic embryos 127
4.7 Analysis of reporter gene expression in hsp68*LacZ*-480 F0 transgenic embryos 128
4.8 Analysis of reporter gene expression in hsp68*LacZ*-450 F0 transgenic embryos 128
4.9 Analysis of reporter gene expression in hsp68*LacZ*-370 F0 transgenic embryos 131
4.10 Chapter Discussion 131
4.10.1 Enhancers within the Runx1 locus: the +3, +23, +110 and +205 132
4.10.2 Enhancers in the Runx1 gene desert: the -480-450-370 composite and the -480, -450 and -370 individual enhancers. 134
4.10.3 Concluding remarks 135

Chapter Five: Functional analysis of the cells in which the +110 enhancer is active 137
5.1 Chapter Aim 138
5.2 Generating an hsp68*GFP*+110 mouse line 139

5.3 Characterization of cells to which the +110 enhancer mediates GFP expression during developmental hematopoiesis 143
5.4 Clonogenic potential of +110 targeted cells: CFU-C assays 147
5.5 Long-term reconstitution ability of cells in which the +110 is active 151
5.6 Megakaryocyte differentiation from E14.5 hsp68*GFP*+110 fetal livers 154
5.7 Analysis of +110 enhancer activity in adult hematopoietic organs ... 155
5.8 Chapter Discussion 157

Chapter Six: Chromosome Conformation Capture (3C)

Physical interactions of regulatory elements with the P1 and P2 promoters 161
6.1 Chapter Aim 162
6.2 The Chromosome Conformation Capture (3C) assay 162
6.3 *Runx1* 3C 168
6.3.1 Rationale for use of different cell types in 3C assays 168
6.3.2 Testing Runx1 3C Taqman probes and primers by qPCR 171
6.4 Elements interacting with the P1 promoter HindIII fragment 175
6.5 Interactions of elements with the 3'P2 promoter HindIII fragment... 182
6.6 Chapter Discussion 189

Chapter Seven: General discussion and future directions 193
7.1 Summary of results 194
7.2 Which *Runx1* enhancers are required for *Runx1* expression? 195
7.3 What controls the distinct activities of *Runx1* enhancers? 197
7.4 How many more *Runx1 cis*-regulatory elements are there? 199
7.5 *Runx1* enhancers as experimental tools 200
7.6 Concluding remarks 201

Appendix I. 202

Appendix II. 203

References 206

Chapter One: Introduction

Blood is essential for life. Since the earliest days of medicine, its importance to the balance between health and illness has been appreciated, being included as one of Hippocrates' 'Four Humors' in the fifth century B.C. With the advancement of the scientific process through the centuries, the composition of blood has been established as a highly complex organ of different cell types. From oxygen-carrying erythrocytes and pathogen-engulfing macrophages, to the B and T-cells of the adaptive immune system, the functions of blood in sustaining and protecting the life of an organism are complex and varied.

It is no surprise therefore that the study of blood formation, or hematopoiesis, is an intensive area of research. Understanding the origin of hematopoietic cells and the genes vital for distinct cellular phenotypes has been the goal of scientists and clinicians for many years. Through understanding what controls normal hematopoiesis, we hope to gain insight into the causes of hematopoietic malignancies such as leukaemia with the goal of creating better therapies and eventually cures for such conditions.

1.1 Adult Hematopoiesis

In the adult mammal the generation of all blood cell types occurs in the bone marrow. Steady-state hematopoiesis sees the generation of millions of different types of blood cell every second. Indeed, in an adult man weighing 70 Kg, this turnover of cells has been estimated to be close to 1×10^{12} cells per day (Ogawa, 1993). Hematopoiesis proceeds in a hierarchical order,

resulting in the formation of myeloid cells (erythrocytes, megakaryoctes/platelets, eosinophils, neutrophils, basophils, monocytes, and macrophages) and lymphoid cells (B-cells and T-cells). These mature cell types represent the end point in the hematopoietic hierarchy (Figure 1.1).

The founding cell of the hierachy is the long-term hematopoietic stem cell (LT-HSC), a rare, pluripotent cell type, first described by Till and McCullough in the 1960's (Till and Mc, 1961; Till et al., 1964). The LT-HSC has the ability to divide asymmetrically reproducing itself and also producing a more committed hematopoietic progenitor. As detailed in figure 1.1, the pathway from LT-HSC to unipotential, mature blood cell types, involves cellular intermediates. Based on phenotypic analyses of BM cell types and their ability to form different blood lineages, it is thought that a process of lineage restriction occurs, with cells progressively losing multipotency as this process proceeds (Luc et al., 2008). Previously, lineage restriction was believed to occur via a common myeloid progenitor (CMP) and a common lymphoid progenitor (CLP) (Akashi et al., 2000; Kondo et al., 1997). Current data suggest a slightly more complex process, whereby ST-HSCs (defined by having less self-renewal ability than the LT-HSC) give rise to a CMP but also a lymphoid-primed multi-potent progenitor (LMPP) (Adolfsson et al., 2005). Both of these intermediates then undergo further lineage restriction to form a granulocyte-monocyte progenitor (GMP), the progenitor of basophils, eosinophils, neutrophils, and monocytes/macrophages. The CMP can also undergo further lineage restriction to form the megakaryocyte-erythroid progenitor (MEP), which gives rise to erythrocytes and megakaryocytes/

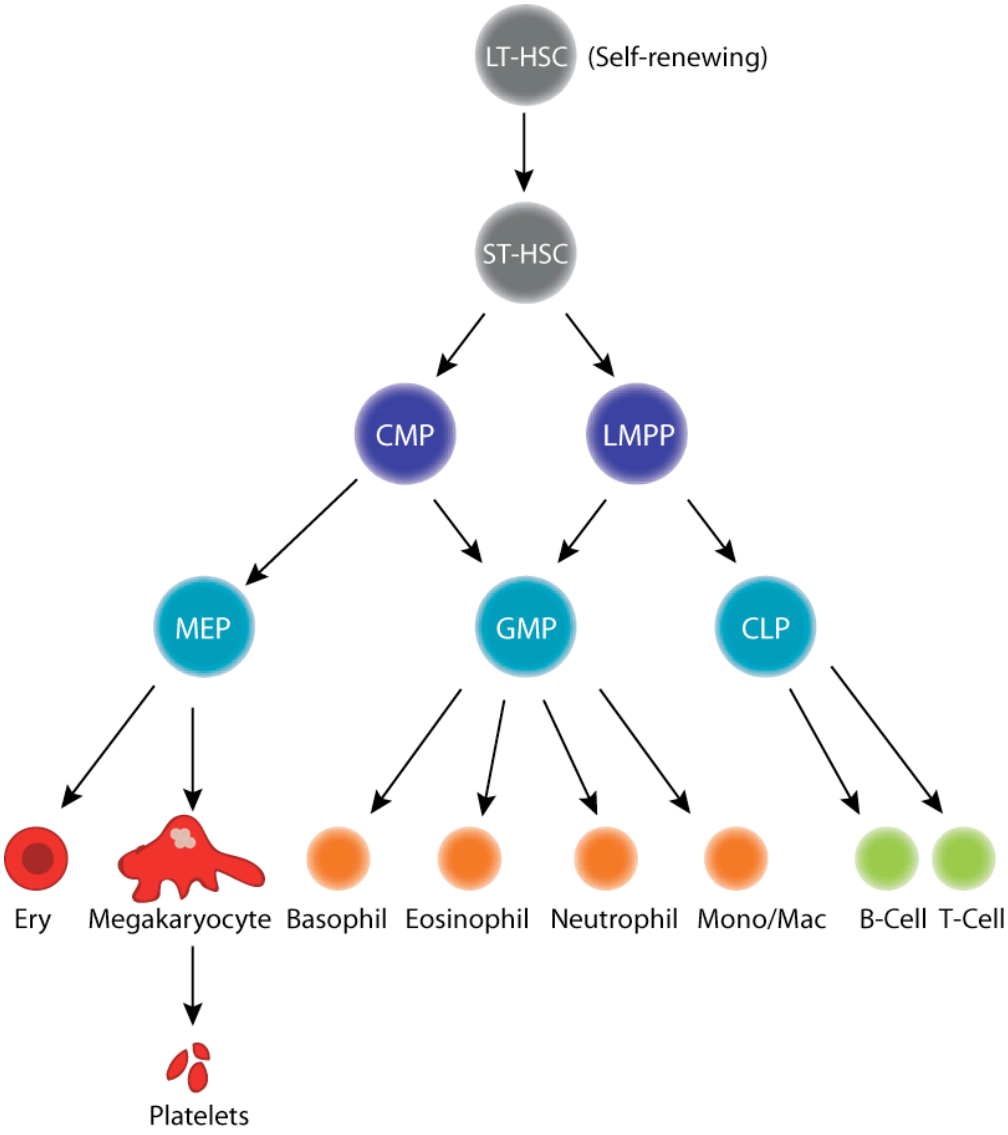


Figure 1.1. Schematic representation of the hematopoietic hierarchy. The model proposes that the hematopoietic stem cell undergoes lineage restriction via a CMP and LMPP. The CMP is able to give rise to MEPs and GMPs whilst the LMPP can also give rise to the GMP as well as a CLP. The MEP, GMP and CLP are the final intermediates before the last lineage restrictions to unipotent, mature blood cell types. Abbreviations: LT-HSC, long-term hematopoietic stem cell; ST-HSC, short-term hematopoietic stem cell; CMP, common myeloid progenitor; LMPP, lymphoid-primed multipotent progenitor; MEP, megakaryocyte-erythroid progenitor; GMP, granulocyte/monocyte progenitor; CLP, common lymphoid progenitor; Ery, erythrocyte; Mono/Mac, monocyte/macrophage. Adapted from (Iwasaki et al,, 2007; Orkin and Zon, 2008) and Dr. Nicolas Goardon, personal communication.

platelets. To generate lymphoid cells, the LMPP undergoes further lineage restriction, generating the CLP, which will further commit to give rise to B and T-cells.

1.2 Developmental Hematopoiesis

Developmental hematopoiesis refers to the generation of the first blood cells and the definitive hematopoietic system in the embryo. In the mouse embryo, this process results in the formation of blood cells at different times and in different anatomical locations, both extra- and intra-embryonic. Of particular interest to developmental biologists is the genesis of the very first HSCs that will support hematopoiesis post-natally and throughout adult life (described in Section 1.1). The formation of HSCs however, is just one facet to the process of developmental hematopoiesis, which occurs in three overlapping waves, defined by the type of hematopoietic cells produced. These waves are i) primitive hematopoiesis, ii) transient definitive hematopoiesis and iii) adult definitive hematopoieisis, explained below (Figure 1.2 A and B).

1.2.1 Primitive Hematopoiesis

Primitive hematopoiesis in the mouse embryo occurs from E7.5 of development and sees the generation of primitive erythrocytes in the extra-embryonic yolk sac (YS) as well as primitive macrophages and megakaryocytes (Figure 1.2 A and B) (Haar and Ackerman, 1971; Moore and

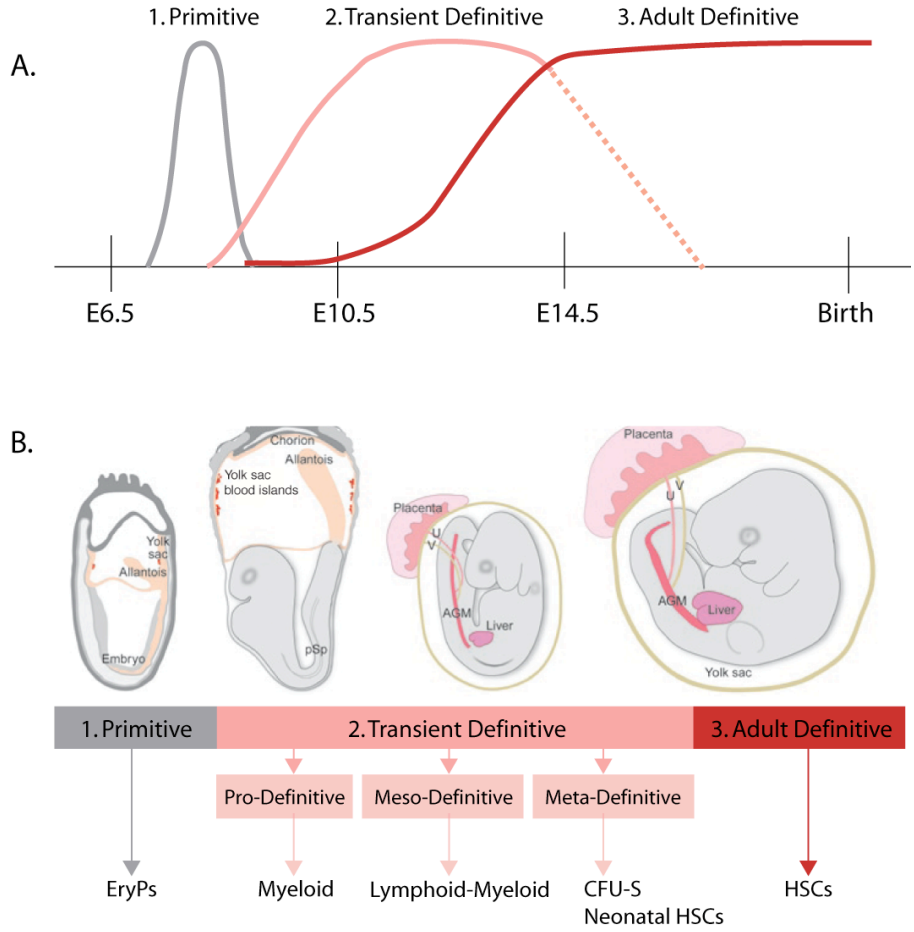


Figure 1.2. The generation of hematopoietic cells in the mouse conceptus. A) The three waves of hematopoiesis and the length of time of cell production within each: 1. Primitive, from around E7.5 to E9.5 (EryP-CFCs cease to be produced by E9.5); 2. Transient Definitive, from E7.5-8 to just before birth and; 3. Adult definitive from E10.5, which continues post-natally in the bone marrow. B) Schematic representation of mouse embryonic development at E7.5, E8.25, E9 and E10.5 (left to right) in relation to the waves of hematopoiesis in '1.2 A'. Primitive erythrocytes (EryPs) are generated during the primitive wave. During this wave the E7.5 and E8.25 concepti show the outgrowing allantois that will fuse with the chorion to form the placenta. The transient definitive wave also occurs from the time of the primitive wave and can be sub-divided into i) Pro-definitive (myeloid cells produced), ii) Meso-definitive (lymphoid-myeloid cells produced) and iii) Meta-definitive (CFU-S and neonatal HSCs produced). During this wave, the E9 embryo has turned and is enveloped in the yolk sac. By late E9 the fetal liver begins to be colonized by hematopoietic progenitors. The adult definitive wave occurs from E10.5. In the E10.5 conceptus, HSCs are found in clusters of hematopoietic cells budding from the luminal walls of the dorsal aorta in the AGM region, the vitelline (V) and umbilical (U) arteries. CFU-S, colony-forming unit-spleen. 'B' adapted from (Dzierzak and Speck., 2008).

Metcalf, 1970; Palis et al., 1999; Tober et al., 2007). During this wave, 'primitive' erythrocytes (EryPs) form in close association with vascular endothelial cells lining the yolk sac blood islands (BIs). Both of these cell types are thought to be derived from a common mesodermal precursor known as the 'hemangioblast', which is found in the primitive streak before the formation of the BIs (Huber et al., 2004) reviewed in (Cumano and Godin, 2007). Upon generation, EryPs are morphologically and biochemically distinct from later adult, or 'definitive', erythrocytes, being six-times larger, nucleated and expressing different combinations of globin genes (Leder et al., 1992; Steiner and Vogel, 1973). The function of EryPs is to provide for the oxygen needs of the early embryo after entering the circulation at 4-6 somite pairs (sp) (E.8.25).

Importantly, the generation of EryPs is transient in nature, continuing only until the 20 sp stage (E.9) of development when the progenitor cells of the EryPs, the EryP-CFCs, stop expanding and cease to be generated (Kingsley et al., 2004; Palis et al., 1999) (Figure 1.2 A). EryPs do however persist in the circulation until E12.5 when their numbers plateau. Until this stage, the cells accumulate increasing amounts of hemoglobin and undergo morphological changes, becoming smaller with greater nuclear condensation. In addition, work by the Palis group has shown that primitive erythroblasts enucleate between E12.5 and E16.5 and are present in the blood of mice up to five days after birth (Kingsley et al., 2004) (reviewed in (McGrath and Palis, 2005)).

1.2.2 Transient definitive hematopoiesis

The transient definitive wave of hematopoiesis results in the formation of the first adult-type progenitors and some adult cell types, in the absence of HSCs. This wave can be further subdivided into the, i) pro-definitive wave, ii) meso-definitive wave and iii) meta-definitive wave, which occur from the time of primitive hematopoiesis at E7.5 up to E9 (Dzierzak and Speck, 2008)) (Figure 1.2 A and B).

During the pro-definitive wave, erythroid-myeloid progenitors are generated in the allantois (an extra-embryonic tissue that will later fuse with the chorion to form the placenta) (E7.5-E8), the YS (E8.25-E9.5) and the placenta (E.9) (Alvarez-Silva et al., 2003; Corbel et al., 2007; Lux et al., 2008; Palis et al., 1999; Rampon and Huber, 2003; Zeigler et al., 2006). Meso-definitive hematopoiesis sees the generation of both erythroid-myeloid-lymphoid progenitors and multipotent low-level repopulating progenitors from E7.5-E8 in the para-aortic splanchnopleura (PAS) of the embryo proper (the precursor of the AGM) (Cumano et al., 1996; Cumano et al., 2001). Meta-definitive hematopoiesis occurs in both the YS and AGM at E9 and results in the emergence of progenitors with potent colony-forming unit-spleen ability (CFU-S) and progenitors known as neonatal repopulating HSCs. The latter cell type are able to confer multi-lineage engraftment when injected into the fetal livers of neonatal recipient mice, but are unable to engraft irradiated adult mice (Medvinsky et al., 1996; Medvinsky et al., 1993; Yoder et al., 1997).

1.2.3 Definitive Hematopoiesis

Definitive hematopoiesis results in the formation of adult-type or 'definitive' HSCs, a trait unique to this wave of hematopoiesis. HSCs are autonomously generated in the aorta-gonad-mesonephros (AGM) region of the embryo proper from E10.5 as part of clusters of hematopoietic cells that bud from the ventral wall of the dorsal aorta (Figure 1.2 A and B) (Medvinsky and Dzierzak, 1996; Muller et al., 1994). This phenomenon also occurs in vitelline and umbilical arteries (VU) (de Bruijn et al., 2000; North et al., 2002).

Shortly after their emergence at these sites, HSCs are also found in the YS (Medvinsky et al., 1996; Muller et al., 1994; Samokhvalov et al., 2007) and placenta (Gekas et al., 2005; Mikkola et al., 2005; Ottersbach and Dzierzak, 2005). Indeed, by E11.5, the AGM, YS and placenta have each been estimated to contain one HSC respectively (Kumaravelu et al., 2002; Ottersbach and Dzierzak., 2010). HSCs then enter the circulation at E12.5 and seed the fetal liver (FL), where they proliferate to produce more HSCs in preparation for seeding the thymus, spleen and ultimately the bone marrow (reviewed in (Medvinsky et al., 2011)). Whether the YS and placenta are capbable of *de novo* generation of HSCs remains to be formally demonstrated as it is unclear whether HSCs seed these organs via the bloodstream from the embryo proper. To address this issue, the Mikkola group used the *Ncx1*$^{-/-}$ mouse model (Rhodes et al., 2008). *Ncx1* encodes a sodium-calcium exchanger membrane. The absence of this protein in *Ncx1*$^{-/-}$ embyos results in

a lack of circulation due to heartbeat failure. Clusters of hematopoietic cells were found in the placentae of *Ncx1*$^{-/-}$ embyos, but transplantation of these cells into lethally irradiated adult recipients to test for long-term, multi-lineage reconstitution potential has not yet been done. These cells can therefore not be formally defined as HSCs.

1.3 Is there a common progenitor of primitive and definitive blood?

How do blood cells of the primitive and transient definitive waves form in the absence of the HSC, which is not generated until after they are produced? Two models have been proposed to try and explain this phenomenon (Figure 1.3, reviewed in (Medvinsky et al., 2011)). The first model proposes that primitive/transient definitive blood forms via a mesodermally derived progenitor to provide for the early needs of the embryo. Conversely, definitive blood is generated from a separate progenitor. In the latter pathway, the progenitor cell would give rise first to a pre-HSC cell type that would subsequently form an HSC (Figure 1.3 A). This model is supported by data from studies using *Xenopus* embryos, where primitive and definitive blood types derive from different blastomeres of the blastula stage embryo (Ciau-Uitz et al., 2000).

In the second model, a common hematopoietic progenitor (CHP) of both primitive and definitive blood is proposed. This CHP would first give rise to a precursor of both the primitive and the transient definitive blood (i.e. to produce EryPs and transient adult blood cell types). The progenitor would

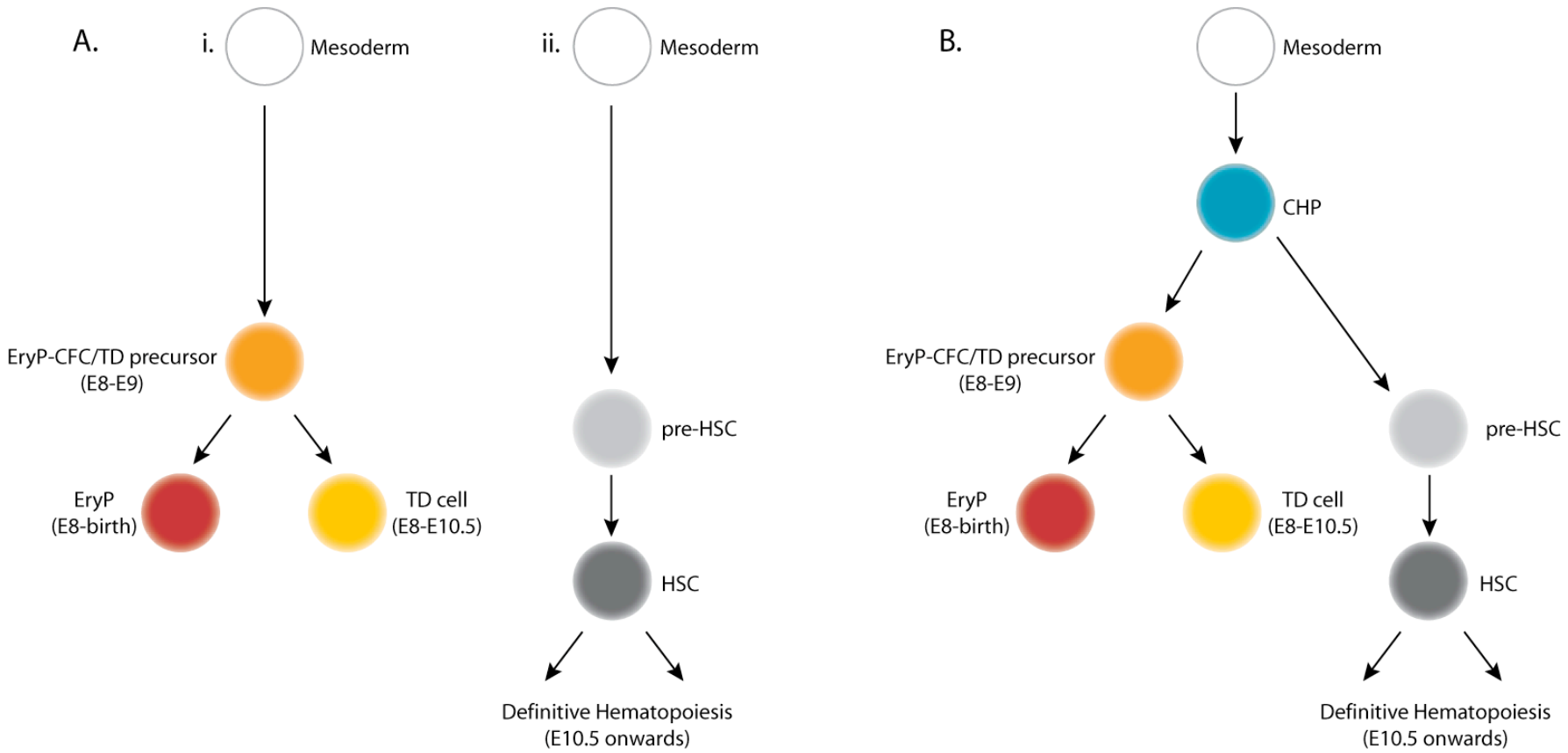


Figure 1.3. For figure legend see overleaf.

Figure 1.3. Two proposed models for the generation of primitive and definitive blood during developmental hematopoiesis. A) Model one: i) Primitive erythrocytes (EryPs) and transient definitive (TD) blood cells are formed from a mesoderm-derived common precursor (EryP-CFC/TD precursor), ii) Hematopoietic stem cells (HSCs) are produced from a separate progenitor, the pre-HSC who's origin is also mesodermal. B) Model two: a mesoderm-derived common hematopoietic progenitor (CHP) first gives rise to the EryP-CFC/TD precursor cell type. When the generation of primitive and transient definitive blood is no longer required, the CHP gives rise to the pre-HSC to begin the formation of adult definitive blood via the HSC.

then cease the production of these transient cells and generate a pre-HSC, which would become an HSC, at the onset of definitive hematopoiesis (E10.5) (Figure 1.3 B).

Model one is supported by the data highlighted in Section 1.2 and strongly argues the case for their being two progenitors, one in the YS which functions as the major site of primitive/transient definitive hematopoiesis and a separate progenitor which leads to the formation of HSCs in the AGM and VU arteries. Further evidence in support of Model One is presented below in Section 1.4. Model two would argue that the YS is the common origin of all blood, with intra-embryonic sites of HSC emergence acting as intermediate locations for cells derived from the YS. The Nishikawa group investigated this model by tracing Runx1-labelled YS cells from E7.5 of development. They propose that these early Runx1-labelled cells in the YS develop into fetal lymphoid progenitors and also adult HSCs (Samokhvalov et al., 2007). However, these results remain inconclusive due to several caveats. The most important of these caveats is that *Runx1*-expressing cells are not restricted to the YS, so the use of *Runx1* expression as a label to lineage trace cells of a YS origin is flawed from the outset.

1.4 What is the cellular origin of the HSC? Hemangioblast vs. hemogenic endothelium.

The cellular origin of the HSC and the molecular controls that initiate its emergence has been an intensive area of research, particularly in the last

fifteen years (reviewed in (Swiers et al., 2010a)). Conflicting ideas exist as to which cell type/types HSCs originate from. As mentioned, HSCs are present in clusters of hematopoietic cells that bud from the ventral wall of the dorsal aorta (DA) in the AGM region of the mouse embryo starting from E10.5 (34sp) as well as in the VU arteries. The close connection of these clusters with the arterial endothelium has therefore led to the idea that blood may form from this endothelium.

Two major theories exist regarding the cellular origin of the HSC. The 'hemangioblast' theory proposes a mesodermal progenitor cell that gives rise to both hematopoietic and vascular endothelial cell types (Huber et al., 2004). So far, *in vivo*, the hemangioblast has only been found as a cell type that precedes the formation of EryPs and transient definitive blood in the YS blood islands. Whether such a cell type precedes the formation of the HSC is not clear. The second theory, or 'hemogenic endothelium' theory, proposes that HSCs are formed from a phenotypically distinct endothelium which has the capacity to generate hematopoietic cell types including the HSC (Jaffredo et al., 1998; Nishikawa et al., 1998; North et al., 1999). Importantly, the two theories are not mutually exclusive (Dieterlen-Lievre, 1998; Lancrin et al., 2009).

The 'hemogenic endothelium' theory is the most probable model for HSC and hematopoietic progenitor cell emergence, supported by data from several independent groups. These data suggest that a hemangioblast cell type first gives rise to hemogenic endothelial cells and that definitive hematopoietic

cells, in turn, arise from these endothelial cells. In an *in vitro* embryonic stem (ES) cell differentiation model which followed the development of ES cells through to the generation of definitive hematopoietic cells, the Lacaud group determined that a $Flk1^{+}$brachyury^{+} hemangioblast-type cell gave rise to adherent hemogenic endothelial-like cells ($Tie2^{hi}$c-$kit^{+}CD41^{-}$) dependent on the action of the TF Scl. These cells in turn gave rise to definitive hematopoietic cells ($CD41^{+}CD45^{+}$), which detached from the hemogenic endothelial-like blast colonies. The generation of the hemogenic endothelial-like cells was dependant on the TF SCL, whilst the generation of blood cells was dependant on Runx1 (Lancrin et al., 2009).

More direct visual data were obtained by the Schroeder group, who followed the differentiation of single cells from an early, ES-derived mesoderm population ($Flk1^{+}$E-Cadherin^{+}) *in vitro*, using time-lapse microscopy. The group labelled cells with cell surface and molecular markers specific to endothelial and/or hematopoietic cell types. They were then able to continuously track alterations in these single cells to determine any change in state from endothelial to hematopoietic cell during culture on OP9 stromal cells. It was found that the $Flk1^{+}$E-Cadherin^{+} mesoderm cells first gave rise to adherent endothelial colonies and from these, the emergence of hematopoietic cells could be tracked. Endothelial cells were seen to become hematopoietic cells from within the population that expressed both hematopoietic and endothelial markers, the hemogenic endothelium (Eilken et al., 2009).

In addition to these *in vitro* studies, two *in vivo* studies (Bertrand et al., 2010; Kissa and Herbomel., 2010) and an *ex vivo* study (Boisset et al., 2010) published in 2010 furthered the ES-system work and, for the first time, provided visual evidence by live imaging of the emergence of hematopoietic cells from the hemogenic endothelium of the DA (reviewed in (Swiers et al., 2010b)). In the Bertrand and Kissa and Herbomel studies, the zebrafish was used as model organism, which is amenable to such studies due to being transparent. Also, similar to the mouse, the first HSCs that form in the zebrafish bud from the ventral wall of the dorsal aorta, which is a common trait of definitive hematopoiesis in vertebrate embryos (reviewed in (Bertrand and Traver, 2009; Dzierzak and Speck, 2008; Medvinsky et al.; Swiers et al., 2010a)). It should be noted that, although called 'HSCs' in the zebrafish literature, these cells have not been formally shown to be HSCs by transplantation assays in zebrafish.

Both the Bertrand and the Kissa and Herbomel studies used the promoter of the gene *kdrl* that encodes the receptor of vascular endothelial growth factor (VEGF), to drive a fluorescent reporter gene in vascular endothelial cells and their resulting progeny. Reporter gene expression was mediated to these endothelial cells and their long-term progeny were shown to be definitive hematopoietic cells. The Bertrand study therefore concluded that blood is derived from the hemogenic endothelium of the DA in this system. One caveat to these claims is that the *kdrl* promoter may also be active in the first blood cells too, so the origin of blood may not necessarily be of endothelial origin in this study. More striking, visual evidence was provided in the study

by Kissa and Herbomel, who were able to show that hematopoietic cells emerge from the endothelium via a specific mechanism named the 'Endothelial to Hematopoietic Transition' (EHT). The investigators tracked GFP^+ endothelial cells in the DA, which gradually altered their morphology so that they exited the endothelium as a hematopoietic cell to enter the circulation via the cardinal vein, without damage to the wall of the DA (Kissa and Herbomel., 2010) (Figure 1.4).

The emergence of blood cells from the aortic endothelium in the mouse embryo was also demonstrated in real-time in the *ex vivo* Boisset study (Boisset et al., 2010). Using the transgenic reporter line *Ly-6A*-GFP (Ly-6A marks endothelial cells and the first emerging HSCs) (de Bruijn et al., 2002; Ma et al., 2002), thick sections of E10.5 transgenic embryo AGM regions were sustained in culture. Using an antibody against CD31 (endothelial marker), the investigators recorded GFP^+ cells emerging from the $CD31^+$ endothelium and entering the lumen of the DA (Figure 1.4). Further labelling with antibodies against c-Kit and CD41 demonstrated the hematopoietic nature of these previously endothelial cells.

In addition to these live-imaging studies, work in the mouse embryo by the Speck lab has shown that essentially all HSCs are derived from cells that had once expressed the vascular endothelial marker, VE-Cadherin (Chen et al., 2009). These data are supported by a similar fate-tracing study that also found HSCs to be derived from VE-$Cadherin^+$ endothelial cells (Zovein et al.,2008). Importantly the formation of hematopoietic cells from the VE-$Cadherin^+$

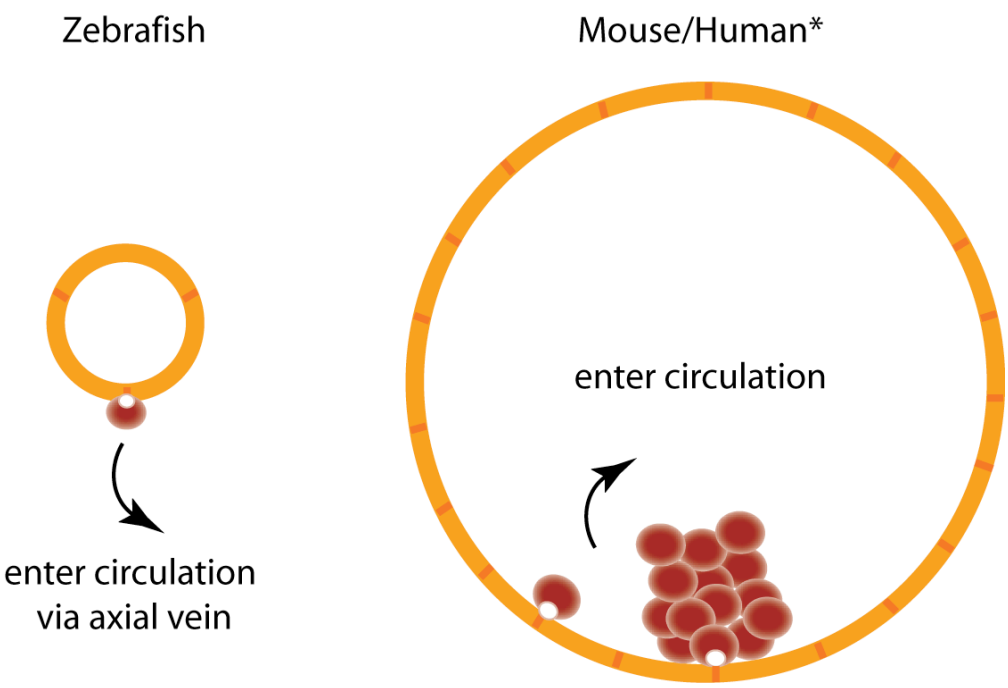


Figure 1.4 Schematic representation of the origin of definitive blood cells in the dorsal aortae of zebrafish and mouse/human. Endothelial cells (orange) have been shown to undergo a physical transformation such that they leave the endothelial layer as hematopoietic cells (red), in a process known as endothelial to hematopoietic transition or 'EHT' (Bertrand et al, 2010; Kissa and Herbomel, 2010). In the zebrafish, hematopoietic cells enter the circulation via the axial vein without the formation of clusters. *In the mouse/human hematopoietic cells bud in the direction of the aortic lumen where they form clusters of cells which eventually enter the circulation. Figure adapted from (Swiers et al, 2010b). Not to scale.

endothelial cells in the Speck lab study was found to be critically dependent on the TF Runx1 (Chen et al., 2009), as is also the case in the Kissa and Herbomel and Boisset studies. Collectively therefore, these studies provide extensive evidence for an endothelial origin of the definitive hematopoietic wave of blood generation. The molecular pathways that control this process are discussed below in Section 1.5 and 1.6, with particular focus on the role of Runx1. Finally, it should be noted that an alternative cellular origin of hematopoietic stem and progenitor cells has also been proposed. Studies in mouse embryos have suggested that the clusters of hematopoietic cells budding from the ventral wall of the DA may originate in the mesenchyme underlying the ventral wall of the DA, in so-called 'sub-aortic patches' (Bertrand et al., 2005; Godin and Cumano, 2002). In this model, cells are thought to migrate from the mesenchyme and integrate into the ventral wall of the DA. However, as so much data indicate that the origin of hematopoietic cells is endothelial, the mesechymal theory is not widely accepted (although can't be formally excluded).

1.5 The role of Runx1 in developmental hematopoiesis

1.5.1. The Runx Transcription Factors

RUNX1 (also known as *AML1* or *Pebp2*α) was first discovered at the t(8;21) breakpoint on human chromosome 21, which is the most frequent translocation in acute myelogenous leukaemia, being implicated in 12% of

cases (Look, 1997; Miyoshi et al., 1991). Point mutations in *RUNX1* have also been found in patients with myelodysplastic syndromes, which can cause increased pre-disposition for acquisition of additional mutations leading to leukaemia (Imai et al., 2000). The protein encoded by *RUNX1* is a member of the 'core binding factor' family of heterodimeric transcription factors (CBFs) or 'Runx' proteins. The mammalian core-binding factor family comprises three DNA-binding α-subunits (Runx1, Runx2 and Runx3) (Bae et al., 1993; Levanon et al., 1994; Ogawa et al., 1993b), and a non DNA-binding β-subunit, CBFβ, which increases the binding affinity of the α-subunits to DNA (Ogawa et al., 1993a; Tahirov et al., 2001; Wang et al., 1993; Warren et al., 2000). Runx proteins bind DNA via a 128 amino acid evolutionarily conserved sub-unit known as the 'runt' domain, so called as the domain has significant homology with the Drosophila pair-rule gene, *runt* (Erickson et al., 1992). This domain also facilitates dimerization with CBFβ (Kagoshima et al., 1993).

1.5.2. The roles of Runx1, Runx2 and Runx3 in development

Each Runx protein has distinct spatio-temporal expression patterns during embryogenesis, and is involved in the generation of different tissues and organs. Knockout studies have shown that Runx1 is essential for the establishment of the definitive hematopoietic programme in the developing mouse embryo (Okuda et al., 1996; Wang et al., 1996) (see Section 1.5.3). Runx2 is essential for osteogenesis. Indeed, in the *Runx2*$^{-/-}$ mouse model, embryos develop and are born, but die shortly after birth due to a complete

lack of ossification of the skeletal system caused by an arrest in the maturation of osteoblasts (Komori et al., 1997; Otto et al., 1997). *Runx3* is involved critical in neurogenesis where it is involved in the formation of proprioceptive dorsal route ganglion neurons (DRG) neurons. Also, in the *Runx3*$^{-/-}$ knock-out model, mice have severe motor defects and lack of coordination due to a failure of a specific sub-population of DRG neurons (TrkC neurons) to regulate axonal projections to target sites in the spinal cord and muscle (Inoue et al., 2002; Levanon et al., 2002). *Runx3* is also involved in the development of CD8 T cells during thymopoiesis (Levanon et al., 2003; Taniuchi et al., 2002; Woolf et al., 2003). Runx3 was previously reported to act as a tumour suppressor in gastric epithelium (Li et al., 2002), but it was recently unambiguously shown that this claim is incorrect (Levanon et al., 2011).

1.5.3. Runx1 activity and expression during developmental hematopoiesis

As mentioned in sections 1.4 and 1.5.2, *Runx1* is essential for definitive hematopoiesis in the developing vertebrate embryo, especially during EHT in the DA (Boisset et al., 2010; Cai et al., 2000; Kissa and Herbomel., 2010; North et al., 1999). The latter poin demonstrated most visually in real time in the Kissa study described in section 1.4. Using a morpholino against *runx1* transcripts (to knock-down the gene's activity), morphant embyos generated far fewer EHT events than WT embryos, and when hematopoietic cells dislodged from the aortic endothelium, they exploded into fragments, due to apoptosis (Kissa and Herbomel., 2010). In the mouse AGM, the Boisset study

(Boisset et al., 2010) also found the emergence of hematopoietic cells to be dependent on the presence of functional Runx1 in the embryo. Indeed, the emergence of hematopoietic cells from the hemogenic endothelial-like cells in the ES differentiation system is also Runx1 dependent (Lancrin et al., 2009).

The critical importance of Runx1 to the generation of definitive blood is not a new paradigm however. Gene knockout studies by Okuda and also Wang in the mid 1990's demonstrated that *Runx1*$^{-/-}$ embryos die at E12.5 due to a complete block of definitive hematopoiesis and severe haemorrhaging of primitive blood in the central nervous system (CNS) (Okuda et al., 1996; Wang et al., 1996). Vitally, these embryos fail to produce HSC-containing intra-aortic clusters and therefore the definitive wave cannot occur (Cai et al., 2000; North et al., 1999). *Runx1*$^{-/-}$ embryos were found to maintain relatively normal primitive hematopoiesis in the YS however (Okuda et al., 1996; Wang et al., 1996) with EryPs being produced in similar numbers to WT embryos, but with an abnormal morphology and a decrease in the erythroid markers Ter119, EKLF and GATA-1 (Yokomizo et al., 2008).

To gain insight into where *Runx1* is expressed in the developing mouse embryo Nancy Speck's group engineered a *Runx1-LacZ-KI* (*Runx1*$^{lz/+}$) mouse, where the last two exons of *Runx1* are replaced with *lacZ* coding sequences (North et al., 1999). The resulting Runx1-β-galactosidase fusion protein is expressed in all domains where *Runx1* is normally expressed, both hematopoietic and non-hematopoietic. Using this model, *Runx1* expression is

evident via X-gal staining at E7.5 in the endoderm and extra-embryonic mesoderm. At E8, it is seen in a tight band circling the YS (Figure 1.5, A v) and strong staining is seen in the emerging EryPs (Figure 1.5, A vi) after which point expression decreases significantly. By E8.5, expression is no longer detectable in these cells.

In the embryo proper at E8.5, staining is seen in the endothelial cells of the vitelline artery and the ventral walls of the paired DA (Figure 1.5, A vii). At E10.5, expression is seen in several hematopoietic sites, including the VU arteries, the FL, and the AGM (Figure 1.5, B v-viii) At this time point, the expression of *Runx1* is seen in cell types in the AGM region including a subset of endothelial cells in the wall of the DA, para-aortic mesechymal cells underlying the DA (Figure 1.5, B v) and a few cells in the urogenital ridges. Importantly, it is during this stage of development that *Runx1*$^{+}$ hematopoietic clusters harbouring HSCs emerge in association with the *Runx1*-expressing endothelial cells. Indeed, in *Runx1*$^{lz/rd}$ embryos (which lack a functional *Runx1* allele), endothelial expression of *Runx1* in the AGM at E10.5 is significantly decreased as compared to the *Runx1*$^{lz/+}$ embryos and clusters are no longer seen either in the DA or the VU arteries (North et al., 1999). Collectively therefore, these early data were already pointing to a role for *Runx1* in the transition of the hemogenic endothelium to a hematopoietic fate.

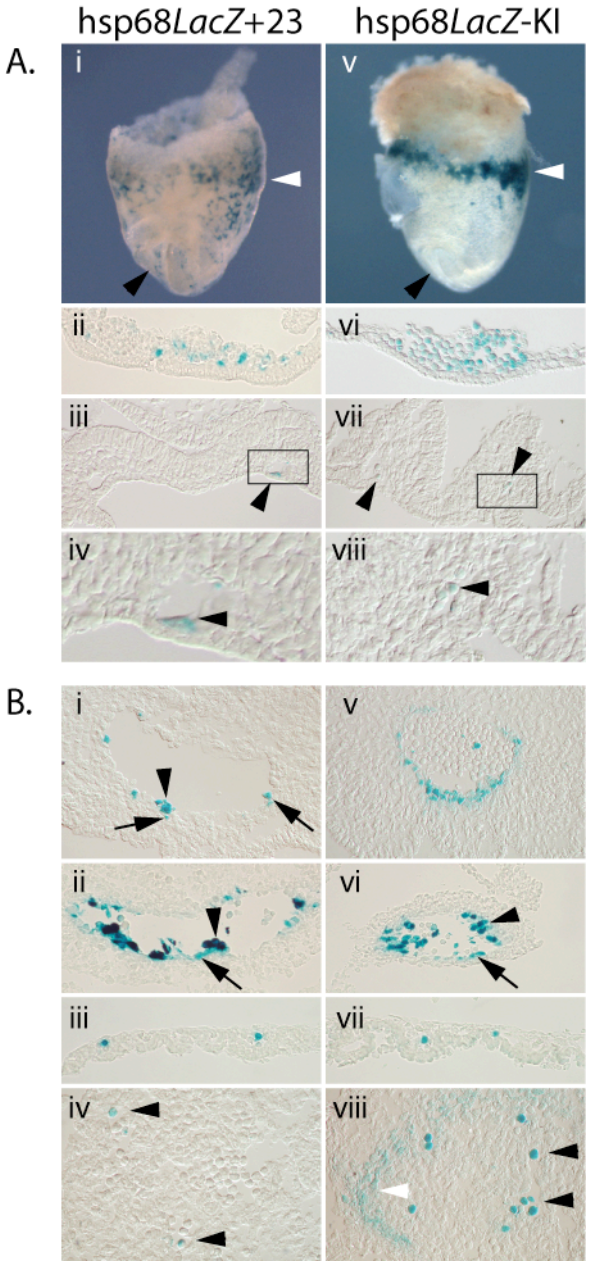


Figure 1.5. Comparison of +23 driven *LacZ* expression to *Runx1-LacZ*KI expression at hematopoietic tissues of the developing mouse embryo. (A) *LacZ* expression in early E8 (5-6 sp) *hsp68LacZ+23* F0 transgenic (i-iv) and *Runx1-LacZ*KI (v-viii) embryos. (i,v) *LacZ* expression in the whole mounts. White arrowheads point at the area of the BI: black arrowheads point at the anterior paired dorsal aortae, outside of the hematopoietic PAS region. (ii,vi) *LacZ* expression in the YS BI. (iii,vii) Expression in the posterior paired dorsal aorta in the PAS region. Arrowheads indicate expression in the wall of the dorsal aortae. (iv and viii) Enlargement of boxed areas in panes iii and vii, respectively. Images in panels i and v were made using a 1.6x objective lens; panels ii-iv and vi-viii: 20x Normarski objective. (B) *LacZ* expression in E10 hsp68*LacZ*+23 F0 transgenic (i-iv) and *Runx1-LacZ*KI (v-viii) embryos. (i,v) Expression in the aorta in the AGM region (dorsal = up). Arrowhead in panel Bi indicates +23 enhancer activity in a cluster of hematopoietic cells. *LacZ* expression is also seen in a few individual cells in the endothelial layer and the mesenchyme (arrows). *LacZ* expression in

the *Runx1-LacZ*KI (v) also includes a wider population of endothelial and mesenchymal cells. (ii,vi), Expression in hematopoietic cells clusters of the umbilical artery (arrowheads) and cells in the endothleial wall (arrows). (iii, vii) Expression in YS blood cells. (iv,viii) *LacZ* expression in FL. Black arrowheads indicate expression in individual hematopoietic cells. The white arrowhead points at the FL capsule. Original magnification, x200. Figure adapted from (Nottingham et al, 2007).

1.5.4. The transcriptional regulation of Runx1 during developmental hematopoiesis

Owing to the vital role that *Runx1* plays in the generation of HSCs from the hemogenic endothelium (both *in vitro* and *in vivo*), our lab has initiated studies into the transcriptional regulation of the gene to gain further understanding of the molecular mechanisms that dictate the emergence of HSCs during development.

Runx1 is located on chromosome 16 in the mouse and is 224kb in length, consisting of 7 alternatively spliced exons (Figure 1.6) (Bee et al., 2009b). The gene is transcribed from 2 alternative promoters, the proximal P2 and the distal P1, which lie 128kb apart (Figure 1.6) (Bee et al., 2009b). We have carried out a systematic analysis of the activity of both promoters in cell populations from the developing mouse embryo and found that *Runx1* expression in primitive erythrocytes is largely P2-derived as is the case in the cells of the putative hemogenic endothelium. Slightly later in development both promoters are transcriptionally active in definitive hematopoietic stem and/or progenitor cells from the YS or AGM and vitelline and umbilical arteries. After cells have migrated to the fetal liver, a promoter switch occurs and the P1 gradually becomes the main hematopoietic promoter and remains so into adulthood (Bee et al., 2009b).

Importantly, neither promoter, in isolation, is able to confer tissue-specific *Runx1* expression *in vitro* or *in vivo* in F0 transgenic (TG) embryos (Bee et

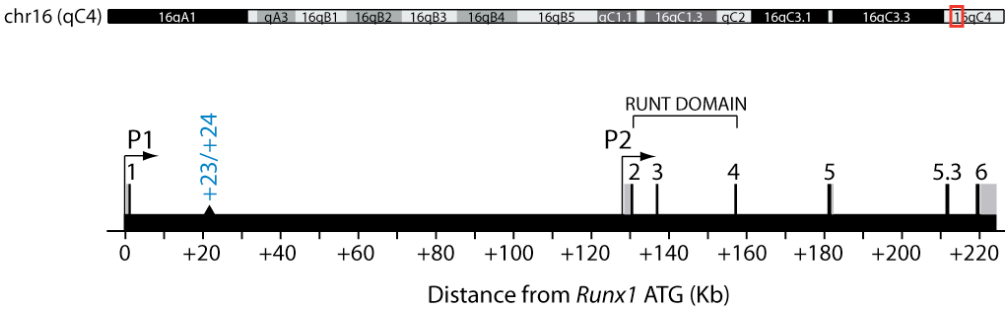


Figure 1.6. Schematic representation of the *Runx1* locus. At the top of the figure is a chromosome ideogram indicating the region represented below (red box). This region represents the 224 Kb *Runx1* locus, including the P1 and P2 promoters, 7 alternatively spliced exons (numbered black points) and UTRs (grey). The exons coding for the DNA-binding Runt domain are indicated. The +23 hematopoietic-specific enhancer and the +24 'booster' element are depicted as a black triangle. The x-axis shows length of the gene in Kb, starting from the start of exon 1 (0).

al., 2009a; Ghozi et al., 1996), indicating that *Runx1* is regulated by *cis*-regulatory elements elsewhere in the locus. Therefore, to identify *Runx1 cis*-reglatory elements, we have used several approaches including comparative genomics in combination with chromatin accessibility assays.

Comparative genomics is a powerful tool to identify conserved non-coding elements (CNEs). CNEs present across divergent species are thought to be retained in genomes due to common regulatory function, for example as gene promoters, silencers, insulators or enhancers (Boffelli et al., 2004; Hughes et al., 2005), although non-conserved, species-specific *cis*-regulatory elements are also evident in genomes (reviewed in (Bulger and Groudine., 2011)). The initial screen for *Runx1* CNEs involved a Gumby analysis (Brudno et al., 2003; Frazer et al., 2004; Prabhakar et al., 2006) of a 6-way Multi-LAGAN alignment of 270 kb of mouse genomic sequence spanning *Runx1* and the corresponding genomic sequences of human (*Homo sapiens*), dog (*Canis familiaris*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), and frog (*Xenopus tropicalis*).

This screen revealed CNEs with significant likelihood of *cis*-regulatory function, including the P1 and P2 (Nottingham et al., 2007). To select only those CNEs that are accessible to TF binding in hematopoietic cells, DNAseI hypersensitive analysis was performed in E12 FL cells and 416B myeloid progenitor cells. This led to the identification of the +23 CNE (Figure 1.6), which was shown to mediate *LacZ* reporter gene expression to the first emerging HSCs in a *Runx1* specific pattern in F0 transgenic mouse embryos.

The +23 enhancer acted in this way in combination with the heterologous *hsp68* promoter (Figure 1.5, B i,ii) (Nottingham et al., 2007) and also the P1 and P2 *Runx1* promoters (Bee et al., 2009a).

In addition to the +23, we have also characterized a mouse-opposum CNE, the +24, which lies 700bp downstream of the +23 in exon 1. Data show that the element has no classical enhancer activity on its own in either transiently transfected 416B cells or in hsp68*LacZ*+24 F0 TGs. However, when cloned together with the +23 to make the +23+24 element (hsp68*LacZ*+23+24 or hsp68*GFP*+23+24), the +24 acts to boost the activity of the +23 both *in vitro* and *in vivo* where it increases the frequency of TG embryos expressing *LacZ/GFP* at the sites of hematopoiesis (Wade Nottingham and Marella de Bruijn, unpublished data). Indeed, an hsp68*GFP*+23LF mouse line is currently being used to help delineate the cells in which the +23 is active during developmental hematopoiesis (Gemma Swiers, John O'Rourke).

1.6 What upstream factors regulate *Runx1* activity?

1.6.1 Signalling Pathways and Transcription Factors

The expression of *Runx1* is regulated by several upstream factors, including the Notch, VEGF, Hedgehog and BMP signalling pathways and the TFs Scl, Gata2, and also Ets factors (reviewed in (Swiers et al., 2010a)) (Figure 1.7). Auto-regulation of *Runx1* has also been suggested by virtue of highly

conserved Runx binding sites being present in or near to the P1 and P2 promoters. Indeed a positive auto-regulatory loop was found to act on the P1 promoter in a myeloid progenitor cell line (Bee et al., 2009a; Levanon and Groner, 2004; Pimanda et al., 2007b; Swiers et al., 2010a). However, we have shown that mutation of the two highly conserved (to frog) RUNX binding sites in the P1 promoter does not affect reporter-mediated expression in $P1_{mut}$*LacZ*+23 transient transgenic embryos (Bee et al., 2009a) suggesting that auto-regulation is conferred by P2 RUNX sites or those located in *cis* elsewhere in the locus (Swiers et al., 2010a).

We have also shown that +23-mediated reporter gene expression in the developing embryo is likely to depend upon the binding of Gata2 and Ets factors to the +23 enhancer during developmental hematopoiesis (Nottingham et al., 2007). Mutation of the Gata site found within the +23 enhancer sequence caused a complete loss of *LacZ* reporter gene expression in E10.5 hsp68*LacZ*+23^{Gata} mutant F0 TGs in the DA. Mutation of three conserved Ets binding sites within the +23 core sequence also caused a significant decrease in *LacZ* expression in hsp68*LacZ*+23^{Ets} mutant F0 TG embryos at E10.5 (Nottingham et al., 2007). We also showed that Scl (and its binding partners Lmo2 and Ldb1) bind the +23 in the 416B myeloid progenitor cell line, suggesting a role in the regulation of *Runx1* via the +23. Scl has also been found to bind to +23, in YS and FL ChIP experiments (Landry et al., 2008) as well as in the HPC7 ES-derived progenitor cell line (Wilson et al., 2010).

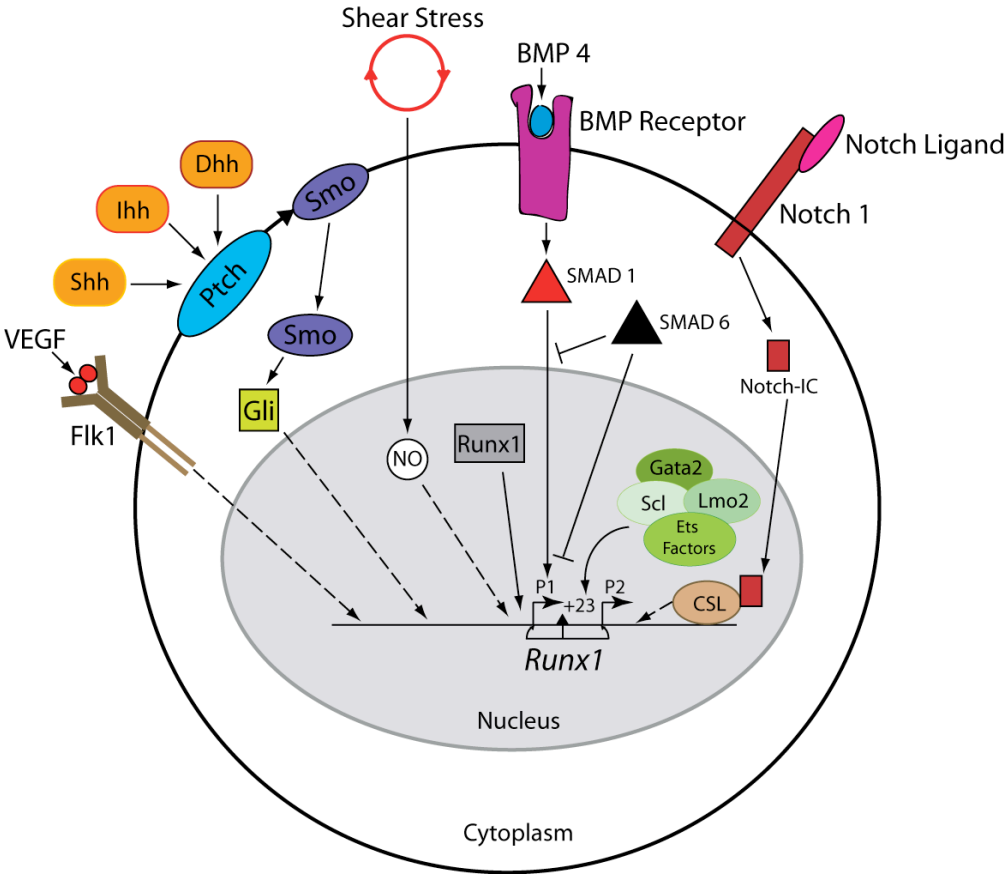


Figure 1.7. Schematic representation of upstream signalling pathways and transcription factors that affect *Runx1* expression during developmental hematopoiesis. Depicted is a cell in which *Runx1* is transcribed in the nucleus (dark grey oval). The P1 and P2 promoters and the +23 enhancer are represented. Dashed lines indicate an indirect/unknown direct effect on expression. Black lines represent a known regulatory effect via binding to a specific region of the gene, e.g. SMAD1/SMAD6 binding to the P1 promoter. Signalling pathway abbreviations: VEGF, vascular endothelial growth factor; Flk1, fetal liver kinase 1; Shh, sonic hedgehog; Ihh, Indian hedgehog; Dhh, desert hedgehog; Ptch, patched (hedgehog receptor); Smo, smoothened (receptor that mediates activation of the hedgehog signalling pathway after activation of Smo); BMP, bone morphogenic protein; Notch-IC, Notch intracellular domain; NO, nitric oxide.

More information regarding these signalling pathways and TFs is detailed below in section 1.6.2.

1.6.2 Notch Signalling

Notch signalling occurs when a ligand (Jagged 1, Jagged 2 or Delta 1, 3 or 4) presented on the surface of a signal-inducing cell, binds to the adjacent signal-receiving cell via one of the four Notch receptor molecules (Notch1, Notch2, Notch3, Notch4). This binding of ligand to receptor induces proteolytic cleavage events that cause the intracellular Notch (Notch-IC) domain to translocate to the nucleus (Figure 1.7). Here the Notch-IC protein associates with the TF CSL, which is already bound at specific DNA sequences. CSL binding at these sites recruits transcriptional repressors, which are displaced upon the binding of Notch-IC to the CSL protein. This removal of repressor complexes and a subsequent exchange for activatory complexes, leads to the activation of Notch1 target genes (reviewed in (Bigas et al., 2010)).

Notch signalling is critical for endothelium formation, arterial specification and HSC formation (Burns et al., 2005). The importance of Notch signalling in the emergence of HSCs has been demonstrated in chimeric mouse studies using *Notch1*$^{-/-}$ ES cells in combination with WT ES cells. In the resulting chimeric mice there is no contribution to the definitive blood population from the *Notch1*$^{-/-}$ ES cells (Hadland et al., 2004). In addition, work from the Hirai group showed that PAS and YS tissues taken from *Notch1*$^{-/-}$ embryos at E9.5

were unable to form the definitive blood lineages when injected into newborn recipients (Kumano et al., 2003). Importantly, Runx1 and Notch1 are thought to act in the same genetic pathway. Runx1 is believed to be downstream of Notch1 in this pathway as enforced expression of *Runx1* in *Notch1*$^{-/-}$ mutants is able to partly rescue AGM hematopoiesis. It remains unclear however if *Runx1* is directly regulated by effectors of Notch1 signalling as no CSL-bound *cis*-regulatory elements have yet been identified in the *Runx1* locus (Wang et al., 2011).

1.6.3 Notch signalling and Gata2

The Notch pathway and the expression of *Gata2* are closely linked. Like Runx1, Gata2 is thought to act in the same genetic pathway as Notch1, as enforced expression of *Gata2* also partly rescues AGM hematopoiesis in *Notch1*$^{-/-}$ and *Jagged1*$^{-/-}$ mutants (Nakagawa et al., 2006; Robert-Moreno et al., 2008). Notch signalling is known to directly regulate the expression of *Gata2* through binding to the *Gata2* promoter (Robert-Moreno et al., 2005; Robert-Moreno et al., 2008). Gata2, like Runx1, is critical for HSC emergence in the AGM (Tsai et al., 1994; Tsai and Orkin, 1997). *Gata2*$^{-/-}$ embryos die around E10.5-11.5 due to anaemia caused by a massive reduction in the number of EryPs formed (Tsai et al., 1994). Of note, both *Notch1* (and some of its ligands and effectors) and *Gata2* are expressed in the emerging clusters of hematopoeitic cells and the hemogenic endothelium in the ventral wall of the DA (Robert-Moreno et al., 2005; Robert-Moreno et al., 2008), both sites of *Runx1* expression. Evidence from Xenopus studies using

morpholinos against *Gata2* demonstrated that *Runx1* expression was lost in these morphants, showing that *Gata2* acts upstream of *Runx1*. We have also demonstrated that Scl binds to the *Runx1* +23 hematopoietic-specific enhancer and that +23-enhancer mediated *LacZ* expression in hsp68*LacZ*+23 TG F0 embryos is abolished when the GATA site within the enhancer is mutated (Nottingham et al., 2007).

1.6.4. The role of Scl

The TF Scl is also expressed at E10.5 in the emerging clusters of the DA, and is well characterized as a master regulator of hematopoiesis (Lecuyer and Hoang, 2004; Porcher et al., 1996; Shivdasani et al., 1995). *Scl*$^{-/-}$ embryos die earlier than *Runx1*$^{-/-}$ embryos (between E8.5-10.5) due to a block in the primitive wave of hematopoiesis, which results in extreme anaemia and tissue necrosis (Shivdasani et al., 1995). This early death also precludes the formation of HSCs and hence the definitive wave. The role of Scl in hematopoiesis was further demonstrated *in vitro* using the ES cell differentiation system, which showed that Scl is required for the transition of the hemangioblast cell type to hemogenic endothelial-like cells (Lancrin et al., 2009). Hemangioblast populations derived from *Scl*$^{-/-}$ ES cells are unable to undergo this transition, meaning that hemogenic endothelial-like cells (from which blood cells form dependent on the action of *Runx1*) are not generated (Lancrin et al., 2009).

In addition, work in the zebrafish has shown that Scl is critical for the emergence of HSCs. In zebrafish morphant embryos in which Scl transcripts have been knocked-down, there is a complete lack of both primitive and definitive blood (Patterson et al., 2005), as is the case in the mouse knockout model (Shivdasani et al., 1995). Interestingly in zebrafish Scl morphant embryos, the DA does not form, meaning that no hemogenic endothelium is present from which HSCs could emerge. The latter point mirrors the *in vitro* data described in the above paragraph, where blood cells could not form from hemangioblasts derived from *Scl*$^{-/-}$ ES cells, as hemogenic endothelium cannot form. Importantly, Scl is known to bind to the *Runx1* +23 enhancer (Landry et al., 2008; Nottingham et al., 2007) suggesting that *Runx1* is a direct downstream target of Scl.

1.6.5 The Hedgehog, VEGF and BMP signalling pathways

The Hedgehog, VEGF, and BMP signalling pathways are also known to affect *Runx1* expression during developmental hematopoiesis (reviewed in (Swiers et al., 2010a)). Hedgehog signalling acts via the binding of hedgehog ligands known as Sonic (Shh), Indian (Ihh) and Desert (Dhh) hedgehog to the transmembrane receptor Patched (Ptch), which in turn activates the transmembrane receptor Smoothened (Smo). Smo then translocates into the cell to the primary cilium, where it activates Gli TFs. The Gli TFs then translocate to the nucleus where they bind to target gene regulatory elements (reviewed in (Mar et al., 2011)) (Figure 1.7). In the zebrafish, hedgehog signalling was found to be critical for the emergence of *runx1*$^{+}$ hematopoietic

cells emerging from the ventral wall of the DA, as embryos treated with the Smo receptor inhibitor cyclopamine had severely reduced numbers of these cells compared to WT controls (Gering and Patient, 2005). Hedgehog is not thought to directly activate *runx1* expression however (Wilkinson et al., 2009).

VEGF signalling occurs via several receptors including the Flk1 trans-membrane receptor (Figure 1.7). This pathway has been shown to be critical for arterial specification in the zebrafish, with inhibition of VEGF leading to a loss of runx1^{+} cells forming in the DA (Gering and Patient, 2005). The ETS-family member TF, Tel/Etv6 has been shown to regulate the production of VEGF in developing HSCs in the zebrafish (Ciau-Uitz et al., 2010) and in the mouse, *Tel*$^{-/-}$ transgenic embryos die between E10.5 and E11.5 due to incorrect blood vessel formation (Wang et al., 1997).

BMP signalling molecules belong to the TGF-β superfamily. The intracellular mediators of BMP signalling are the SMAD proteins, with SMAD6 acting as an inhibitor of BMP signalling and SMAD1 functioning to transduce BMP signals to the nucleus (Ishida et al., 2000) (Figure 1.7). Work by the Gottgens group has shown that BMP4 and Smad6 transcripts localize to the ventral floor of the DA at E10.5 (Pimanda et al., 2007a). In the same study, it was found that Smad6 binds Runx1, targeting it for ubiquitination. In addition, a more recent investigation has demonstrated that *Runx1* expression is regulated via Runx1 binding to a novel *Smad6 cis*-reglatory element (Knezevic et al., 2011). In addition *in vitro* luciferase assays have shown that

Smad1 is able to trans-activate the *Runx1* P1 promoter, enhancing *Runx1*-mediated luciferase activity (Pimanda et al., 2007a). Finally, zebrafish studies have also demonstrated an importance for BMP signalling *in vivo*, where BMP4 signals act ventrally to induce the blood programme in ventral wall of the DA (Wilkinson et al., 2009).

1.6.6 The effect of blood flow on Runx1 expression

Recent data provide evidence that HSC emergence in the DA is also dependent on shear stress forces created by the flow of blood through the artery. Work in the Zon lab has shown that embryos of the mutant zebrafish line 'silent heart' (sih), which lack a heart beat and blood flow, generate significantly fewer runx1^{+} cells in the DA in comparison to WT embryos (North et al., 2009). Blood flow was shown to mediate its effect on the emergence of runx1^{+} cell emergence via the production of Nitric Oxide (NO). Indeed, in *sih* mutant embryos, the expression of *nos1* (and therefore production of NO) is severely down-regulated concomitant with the loss of *runx1*$^{+}$ cells. This phenotype could be rescued by the addition the NO donor S-nitroso-N-acetyl- penicillamine (SNAP), allowing runx1^{+} cells form once again in the *sih* mutants. These data provide an intriguing link between the effect of shear stress and the generation of runx1^{+} cells, and also strengthen the case for the DA being a site of de novo HSC generation in the absence of a blood produced in earlier waves in these *sih* mutants.

1.7 Chapter Summary

The process of developmental hematopoiesis is complex and sees the generation of different blood cells and progenitors in three major waves, most importantly the HSC during the adult definitive wave. Data strongly suggest that there are separate progenitors of the primitive/transient definitive blood populations and the HSC pool, with the AGM and VU arteries being the *de novo* sites of HSC generation. HSCs are formed via EHT and this process is critically dependent upon the TF Runx1. The expression of *Runx1* during development is affected by several upstream signalling pathways and TFs. The effect of these upstream proteins occurs by direct binding to *Runx1* (Scl and Gata2), indirectly (Hedgehog signalling) and in yet unclear ways (Notch signalling). Finally, we have shown that the +23 hematopoietic-specific enhancer mediates *LacZ* expression to the first emerging HSCs and to a sub-set of cells where endogenous *Runx1* is expressed, but not the entire complement.

1.8 Project Aim

The ultimate goal of this work is to understand what controls the emergence of blood during developmental hematopoiesis. Using *Runx1* as an entry point into this question, I will identify and characterize Runx1 *cis*-regulatory elements to understand what controls the specific spatio-temporal expression patterns of this critical gene. Data aimed at achieving this goal are presented in four results chapters, starting with chapter three. In this chapter, chromatin

assays (ChIP-chip, ChIP-seq and DNaseI-seq) are used to screen for putative enhancer elements within and surrounding the *Runx1* locus. *In vitro* luciferase assays are also used to test for enhancer activity of the selected regions in hematopoietic cell lines. Chapter four presents *in vivo* data, demonstrating the ability of seven previously unidentified enhancers to mediate *Runx1*-specific reporter gene expression in F0 TG embyos and chapter five uses a novel mouse line to characterize cells in which the +110 enhancer is active. Finally, chapter six uses the Chromosome Conformation Capture (3C) assay to characterize interactions between *cis*-regulatory elements and the *Runx1* P1 and P2 promoters.

Chapter Two:

Materials and Methods

2.1 Bioinformatics

2.1.1 Comparative Genomics

Alignments of genomic sequences from divergent species were performed using the Vista alignment program (www.gsd.lbl.gov/vista). The program uses a global algorithm with a 100bp moving window and a minimum conservation score of 70%. Conserved non-coding element (CNE) ranking in terms of likelihood of function was carried out by GUMBY (www.hazelton.lbl.gov/~sminovit/enhancers; (Prabhakar et al., 2006)).

2.1.2 ChIP-Chip, ChIP-Seq and DNaseI data

Data tracks generated by these methods were loaded on to the UCSC Genome Browser (http://genome.ucsc.edu/) as '.wig' files. Tracks were aligned against the *Runx1* syntenic region of mouse build NCBI37/mm9.

2.2 DNA preparation and manipulation

2.2.1 Polymerase Chain Reaction (PCR)

Primers were designed using the Primer3 web-tool http://frodo.wi.mit.edu. Optimal annealing temperatures were determined for each primer set using gradient PCR (50^{o}C to 70^{o}C) and an extension time of 1 minute (for fragments > than 1 kb, extension times were increased by 1 minute/ kb) on a Peltier-200 thermocycler (MJ research). For PCR reactions, including

genotyping, qPCR analysis and the generation of CNE fragments < 1 kb, HotStarTaq DNA Polymerase (Qiagen) was used. Amplification of CNEs > 1 kb in length was performed using Phusion® Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific). For primers used for generating elements for cloning, see Table 2.1.

2.2.2 Ligation of DNA into vectors

DNA was ligated into vectors using T4 DNA ligase (Fermentas). Reactions were carried out in both 2:1 and 3:1 insert to vector ratios, using 50 ng of vector DNA. Ligations were incubated at RT for 1 hour or at 4°C O/N.

2.2.3 Transformation of bacteria

Chemically competent DH5α *E.Coli* (Invitrogen) were used for routine transformations. Typically, 1 μl of ligation product was added to 50 μl of DH5α cells and incubated on ice for 30 minutes. Cells were then heat shocked at 42°C for 45 seconds and returned to ice for 2 minutes. After the addition of 1 ml of LB medium, the cells were incubated for 1 hour at 37°C with continuous shaking and then 100 μl plated onto agar plates containing the correct antibiotic. Plates were incubated O/N at 37°C.

2.2.4 Growth of bacterial colonies for DNA isolation

Six to twelve bacterial colonies were picked from agar plates and added separately to 5 ml volumes of LB containing the appropriate antibiotic. These

PCR Fragment	PCR Primers
-480	F: 5' GACTCGAGAGCACTTGGTGTTGCTGATG 3'
	R: 5' GAGTCGACTCTAGCAATGGTGCACAAGC 3'
-450	F: 5' GACTCGAGAGGAAAGGTTGTGACCAACG 3'
	R: 5' GAGTCGACTCTATGCCGGGCTTTACAGT 3'
-370	F: 5' GACTCGAGTTTGGAGACTCACGCATCTG 3'
	R: 5' GAGTCGACCGGCCCATTTATTATTGCAC 3'
+3	F: 5' GACTCGAGTTCACTGGTCTGCAGTCTCC 3'
	R: 5' GAGTCGACCCCAATGCGTCAGTCATCTT 3'
+110	F: 5' GACTCGAGTGCACACAAACACACACACC 3'
	R: 5' GAGTCGACCTTCTCTCAGCTTGCCCATC 3'
+205	F: 5' GACTCGAGGACACTGCAGAGACCAGACG 3'
	R: 5' GAGTCGACAAAACAACAAGCCGTTCACC 3'

Table 2.1. Primers used for generating sequences for cloning.

were then grown O/N at 37°C with continuous shaking to generate mini-cultures. Midi cultures were grown under the same conditions as minis by inoculating 25 ml of LB + antibiotic with 25 μl mini culture. DNA was isolated from the resulting cultures using either a mini-prep or midi-prep kit (Qiagen) according to the manufacturer's protocol. For mini-preps DNA was isolated using a silica column and eluted in 30 μl water. For midi-cultures, DNA was precipitated, washed in 70% EtOH, air-dried and resuspended in 120 μl water.

2.2.5 Digestion of DNA with restriction enzymes

DNA was digested with appropriate enzymes supplied either by New England Biolabs, Fermentas or Roche. Enzyme concentration was maintained at 10% or less of the total reaction volume (for optimal enzyme activity) and typically involved over-digestion concentrations prescribed by the manufacturers to ensure total digestion of the restriction sites. Reactions used the appropriate buffer for the enzyme and were performed at 37°C, from 5 minutes (Fermentas Fast Digest) to O/N.

2.2.6 Determination of DNA fragment size by agarose gel electrophoresis

Digested DNA was run on a 0.8% or 1% agarose in 0.5% TBE gel supplemented with 400 ng/ml ethidium bromide (lower concentration of agarose for larger fragment sizes) alongside undigested DNA as a control. DNA was visualized using a Gel Doc 2000 (Biorad), the sizes of the DNA

samples being assessed using either a 100 bp or 1 kb DNA ladder (Fermentas) run with the samples.

2.2.7 DNA extraction from agarose gels

DNA was cut from 0.8% agarose gels under long-wave UV light and extracted using the QIAquick Gel Extraction kit (Qiagen), as per the manufacturer's protocol. DNA was isolated using one or more silica column(s) (maximum 10 μg/column) and eluted in 30 μl water/column. Purity of the isolated DNA was confirmed by running a sample of the eluate on a 1% gel (to check for contamination from other fragments contained in the original sample e.g. vector backbone).

2.2.8 DNA sequencing

Sequencing of double-stranded DNA was done at the Weatherall Institute of Molecular Medicine sequencing service using Big Dye Chemistry (ABI). Samples were prepared to a concentration of < 300 ng/μl and amplified in house using 3.2 pmol of appropriate primers per reaction. Amplified products were then run on an ABI-3730 DNA analyser. Sequence chromatograms were aligned against original sequences and analyzed for homology using Macvector 10.0 software (Accelrys).

2.2.9 Cloning of DNA fragments into reporter constructs

To generate reporter constructs containing *Runx1* CNEs, sequences were first cloned into the pGEM®-T Easy vector (Promega) or the TOPO® vector (for

CNEs >1 kb in length) (Invitrogen) as described in 2.5. After sequencing and restriction digests to verify the insertion of the CNE, sequences were cut from these vectors using appropriate restriction enzymes and sub-cloned into one, or a combination of, the following reporter vectors: *hsp68LacZ*, *hsp68GFP* (for transgenesis) or *pGL3-Promoter* (for luciferase assays). Constructs used for transgenesis were purified from vector backbones by gel extraction and further ultra-purified using the Schleicher and Schuell Elutip-D-Column-Set according to manufacturer's protocol.

2.2.10 Genotyping using adult mouse and embryonic tissues

To verify the genotype of mice pups and embryos, a small piece of tissue was taken (e.g. ear notch or yolk sac) and incubated in 50 – 100 μl PBND (50 mM KCl, 10 mM Tris, 2.5 mM $MgCl_2$, 100 μg/ml gelatine, 0.45% NP40, 0.45% Tween) containing 250 μg/ml Proteinase K and incubated at 55°C (30 minutes for YS to O/N for pup tissue). Proteinase K was then heat inactivated by heating at 95°C for 10 minutes. 2 μl of the heat-inactivated sample were used to verify the genotype using sequence specific primers and HotStarTaq DNA Polymerase (Qiagen) (Table 2.2).

2.2.11 Isolation of Genomic DNA

Genomic DNA was typically obtained from mouse spleens. Spleens were first passed through a 70 μM cell strainer (BD Falcon). Single cell suspensions were then lysed in 10 ml cell lysis buffer (50 mM Tris pH8, 100 mM EDTA, 100 mM NaCl, 1% SDS, 150 μg/ml Proteinase K) by incubating at 55^{o}C

PCR Fragment	PCR Primers
Myogenin	F: 5' TTACGTCCATCGTGGACAGC 3'
	R: 5' TGGGCTGGGTGTTAGTCTTA 3'
LacZ	F: 5' GCAGATGCACGGTTACGATGC 3'
	R: 5' GTGGCAACATGGAAATCGCTG 3'
GFP	F: 5' GACGTGAACGGCCACAAGTTCA 3'
	R: 5' GTGGCGGATCTTGAAGTTCACC 3'

Table 2.2. Primers used in genotyping reactions. Myogenin was used as a positive control.

overnight. Samples were extracted twice from equal volumes of phenol: chloroform: isoamyl alcohol (25:24:1) and once from equal volume of chloroform. In each case the aqueous upper layer was retained after spinning at between 3000 and 13000 rpm for 30 minutes. Genomic DNA was precipitated by adding 1/10 volume of 3 M sodium Acetate (pH 7.4) and two volumes of 100% ethanol and gently mixed. DNA was either pelleted by centrifugation at 7000 rpm for 30 minutes or recovered using a Sheppard's crook (a Pasteur pipette whose tip has been sealed in the flame of a Bunsen burner). DNA was washed in 70% ethanol, air-dried and dissolved in desired volume of TE.

2.2.12 RNA isolation

RNA extractions were performed using equipment and bench space that had been cleaned using RNaseZap® (Ambion) to remove any contaminating RNases. Total RNA was extracted from cells using TRIzol® reagent (Invitrogen), according to manufacturer's instructions. Firstly, 5-10 x 10^6 cells were lysed in 1 ml of Trizol by repetitive pipetting. Homogenized samples were then incubated for 5 minutes at RT to dissociate nucleoprotein complexes. 200 μl of chloroform per 1 ml of Trizol were then added and tubes were shaken vigorously by hand for 15 seconds followed by a further incubation at RT for 3 minutes. Samples were then centrifuged at 11000 rpm for 15-20 minutes at 4°C and the resulting aqueous phase was transferred to a fresh tube. RNA was precipitated by adding 0.5 ml of isopropanol per 1 ml of Trizol used for homogenization. Samples were then incubated at -20°C for 15 minutes and centrifuged at 11000 rpm at 4°C for 10 minutes. The supernatent

was then removed and the RNA pellet was washed in at least 1 ml 75% EtOH per 1 ml Trizol used for homogenization. Samples were mixed by vortexing and then centrifuged at 12000 rpm for 10 minutes. Residual EtOH was removed and the pellet was air-dried for approximately 5 minutes before being re-suspended in an appropriate volume of RNase-free water. Re-suspended RNA was stored long-term at -70°C.

2.2.13 cDNA synthesis

cDNA was synthesized from total RNA using the SuperScript® III Reverse Transcriptase system (Invitrogen), following the manufacturer's instructions. Up to 5 μg RNA were added a total reaction volume of 10 μl, consisting of 1 μl 10 mM dNTP mix, 1 μl of 50 ng/μl random hexamers and the appropriate volume of DEPC-treated water. Samples were then incubated at 65°C for 5 minutes, followed by incubation on ice for 1 minute. Subsequently, the RNA/primer mix was added to a cDNA synthesis mix consisting of 2 μl of 10 x reverse transcriptase buffer, 4 μl of 25 mM $MgCl_2$, 2 μl of 0.1 M DTT, 1 μl of RNaseOUT Recombinant Ribonuclease Inhibitor (40 U/ μl) and 1 μl of Superscript III RT (200 U/ μl). After gentle mixing, the reaction mix was incubated at 25°C for 10 minutes and then 50°C for 50 minutes. The reaction was then stopped by heating to 85°C for 5 minutes and then chilling the sample on ice. After brief centrifugation, 1 μl of RNaseH was added and samples were incubated at 37°C for 20 minutes. cDNA was stored long-term at -20°C.

2.3 Cell Culture

2.3.1 Mouse 416B myeloid cell line culture

416B cells were originally derived from mouse bone marrow cultures infected with Friend leukaemia virus (FLV) (Dexter et al., 1978; Dexter et al., 1979). Cells were grown in Fischer's medium (Gibco) supplemented with 20 % horse serum, 2mM L-Glutamine (Gibco), 50 U/ml penicillin and 50 μg/ml streptomycin (Cambrex). The cells were maintained at a density of 2 – 8 x 10^5 cells/ml in a total volume of 40 ml in T175 flasks (Corning) at 37^{o}C and 5% CO_2.

2.3.2 Human K562 erythroleukemia cell line culture (by Jackie Sloane-Stanley-JSS)

K562 cells were originally derived from leukemic cells obtained in 1970 from a pleural effusion of a 53-year-old female suffering from chronic myelogenous leukaemia (CML) (Lozzio and Lozzio, 1975). Cells are grown in RPMI medium (Gibco) supplemented with 12 % FCS (Invitrogen), 2mM L-Glutamine (Gibco), 50 U/ml penicillin and 50 μg/ml streptomycin (Cambrex). The cells were maintained at varying densities as the cell concentration is not critical, up to a stationary phase concentration of 2-3 x 10^6 cells/ml. Cells were generally kept at a concentration on > 1 x 10^5 cells/ml and split twice a week.

2.3.3 Mouse HPC7 ES-derived progenitor cell line culture

HPC7 hematopoietic progenitor cells were originally derived from *LH2*-transformed mouse embryonic stem cells (Pinto et al., 1998). Cells were grown in IMDM medium (Gibco) supplemented with 8 % FCS, 8 % SCF-conditioned media, 50 μg/ml streptomycin (Cambrex) and 1.5 x 10^{-4} M MTG (Sigma). Cells were cultured at a starting density of 0.5 x 10^{6} cells/ml and split every 48 hours with a complete media change to ensure sufficient SCF concentration in the medium for good growth.

2.3.4 Mouse erythroleukemia (MEL) cell line culture (JSS)

MEL cells were originally derived from FLV-induced mouse leukemia cells (Friend et al., 1971 and references therein). MEL cells were grown in RPMI medium (Gibco) supplemented with 15 % FCS, 2mM L-Glutamine (Gibco), 50 U/ml penicillin and 50 μg/ml streptomycin (Cambrex). The cells were maintained at a density not exceeding 1 x 10^{6} cells/ml. To induce cells to differentiate into erythroid (I-MEL) cells, they were treated with 3mM Haemin (Sigma) plus 25 μM HMBA (Sigma). Typically, this was done using a culture containing 5 x 10^{5} cells/ml in 30 ml. The cells were then harvested 3 days later, erythroid induction being apparent by the red appearance of the cells.

2.3.5 J1 mESC cell line culture (JSS)

The J1 mESC line was originally derived from a male agouti 129S4/SvJae embryo (Li et al., 1992). J1 mESCs were cultured in DMEM medium (Gibco)

supplemented with 15% FCS, 2% Leukemia Inhibitory Factor (LIF) conditioned medium, 50 U/ml penicillin, 50 μg/ml streptomycin (Cambrex), 2mM L-glutamine and $1.5x10^{-4}$ M MTG (Sigma). Cells were passaged every 2-3 days.

2.3.6 Megakaryocyte differentiation from E14.5 mouse fetal livers

Megakaryocytes were differentiated from E14.5 fetal livers of *hsp68GFP*+23LF-Line 3 and *hsp68GFP*+110-Line 5 embryos, following a modified protocol from Yosef Dicken (Weizmann Institute of Science, Israel). Single cell suspensions of individual FLs were made by pipetting the FLs up and down in 0.5 ml of DMEM (Gibco) + 10 % FCS + 50 U/ml penicillin and 50 μg/ml streptomycin (Cambrex) in a 1.5 ml eppendorf tube. FL cell suspensions were then plated in a total of 2 ml of the above medium inoculated with a 1:100 dilution of TPO supernatent (kind gift of Dr. Catherine Porcher), in wells of a 6-well culture plate (Corning). The cultures were then placed in a clear plastic tissue culture box containing water at the bottom and incubated at $37^{o}C$ and 5% CO_2 for 4 days. At the end of the incubation period, cells were harvested and resuspended in PBS +10 % FCS + 50 U/ml penicillin and 50 μg/ml streptomycin (Cambrex) and processed for FACS anaysis by staining with CD41 ab (BD Pharmingen).

2.4 Luciferase Assay

2.4.1 Electroporation of cells

1 x 10^7 cells were electroporated per luciferase construct. Cells were centrifuged for 5 minutes at 1000 rpm to form a pellet and then washed in 5 ml PBS. Following a repeat centrifugation step, the PBS was removed and the pellet of cells re-suspended in 180 μl of cell-appropriate media. 10 μg of test plasmid + 1 μg Renilla plasmid were then added to a 4 mm electroporation cuvette (Flowgen). The cells were then transferred to the cuvette and mixed thoroughly with the plasmids by pipetting. Cells were electroporated using a gene pulser electroporator (Biorad) set at 960 μFD and 220 mV and were then placed immediately on ice before being added to 5 ml of media in a 15 ml falcon tube. The cells were then spun at 1000 rpm for 5 minutes and re-suspended in a total volume of 20 ml of media before being incubated O/N in a 75 ml flask at 37^{o}C and 5% CO_2.

2.4.2 Analysis of luciferase activity

Cells electroporated with test and renilla plasmids were harvested 24 hours after electroporation by centrifugation at 1000 rpm for 5 minutes followed by a wash step in PBS. Cell pellets were then re-suspended in 100 μl of 1 x passive lysis buffer (Promega) and incubated at RT for 5 minutes. Luminescence was determined using the Promega Dual luciferase kit protocol, and an automated Fluostar Optima luminometer (BMG Lab Tech). In this assay, 100 μl of LARII buffer (Promega) were added to 20 μl of the cell lysate, to activate the firefly luciferase activity. The activity of the firefly luciferase was subsequently quenched and the renilla luciferase activated by the addition of 100 μl of Stop and Glo buffer (Promega). Transfection efficiency between different samples was controlled for by normalising the luciferase activity to that of the renilla. Fold luciferase activity was then calculated by comparing test samples to the activity of the p-GL3-Promoter enhancer-less control vector.

2.5 Chromatin Techniques

2.5.1 Chromatin Immunoprecipitation (ChIP)

ChIP assays were performed according to a modified Millipore ChIP protocol as described in (Nottingham et al., 2007), using 1 x 10^7 cells per antibody. Cells were incubated in 0.5% formaldehyde in medium for 10 minutes at RT to cross-link protein:DNA. Glycine (0.125 M) was added to quench the reaction and cells were washed twice with cold PBS (Gibco) with proteinase

inhibitor (Complete, Roche). Nuclei were lysed using SDS lysis buffer (ChIP Assay Kit, Millipore). The lysate was then sonicated to produce fragments of chromatin between ~ 200-500 base pairs. ChIPs were performed, after overnight incubation with the appropriate antibody, using protein A beads (ChIP Assay Kit, Millipore). For antibodies used in ChIP, see Table 2.3. Relative enrichment of target sequences was measured by qPCR using probes and primers in Table 2.4 (*Runx1*) and Table 2.5 (*α-globin*). See 2.25.

2.5.2 Quantitative Real-Time PCR (qPCR)

qPCR was carried out using primers and 5' FAM - 3' TAMRA probes (Taqman) specific for sequences of interest, designed using Primer Express 3.0 software (Applied Biosystems). Before use in assays, the activity of the probe + primer mixes was tested using a dilution series of gDNA, to verify that the signal produced by the probe + primer set was in proportion to the amount of DNA in the reaction. In addition, products of the qPCR reaction were run on a 3 % agarose gel supplemented with 400 ng/ml ethidium bromide, to test for primer-specific single bands. See table 1.3. Each qPCR reaction was of 25 μl total volume and consisted of 400 nM primers: 200 nM probe in Taqman Universal PCR Mastermix (Applied Biosystems) plus the appropriate amount of DNA, depending on its origin. Reactions were then analysed using a Sequence Detection System thermocycler 7000 (Applied Biosystems). Cycling conditions were 50°C for 2 minutes and 95°C for 10 min followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 min. Subsequently, a Ct value was calculated for each reaction, indicating the cycle at which the reaction reaches the exponential phase. This value is

Antigen	Host/Ab Type	Company/Source	Cat. No./Reference
H3K4Me1	Rabbit Polyclonal	Abcam	ab8895
H3K4Me2	Rabbit Polyclonal	Millipore	07-030
H3K4Me3	Rabbit Polyclonal	Millipore	07-473
AcH3	Rabbit Polyclonal	Abcam	ab1791

.

Table 2.3. Antibodies used in ChIP experiments

Probe/Primer name	Sequence
+23 F	5' CGAAAAATAAACCGGCAGTTGA 3'
+23 R	5' CAAGCTGCCCACGTTATCAGT 3'
+23 Probe	5' CAAGAGCGAGAAAACCGCAGGCCT 3'
+24 F	5' CAGAGTCCAGCATGGAGATAGTTTT 3'
+24 R	5' CGCAGTTATTCACATCGTGGAA 3'
+24 Probe	5' CTTCTTGGAGTCATGTGCCCTGGCC 3'
+70 F	5' CCTGAAGCCCCTGAGTCTTG 3'
+70 R	5' TGACAAAATGACCTGGAGACAGA 3'
+70 Probe	5' TTGCTACTGGGAAAGGGACCCTCTCCT 3'
P1 Promoter F	5' CACAAATAACAACAGGATCTGAAAGC 3'
P1 Promoter R	5' TCCCCTGCCTCAGTCTTTTCT 3'
P1 Promoter Probe	5' ACCAAATCCGCACAGGACAGAGAGAGC 3'
P2 Promoter F	5' CCCAGGCTGATACCGGAAA 3'
P2 Promoter R	5' TCCCTCAGCTCGTTTGCATA 3'
P2 Promoter Probe	5' CCTGTGATTGGCCGTCGGTCTCC 3'
Gapdh F	5' CAAGGCTGTGGGCAAGGT 3'
Gapdh R	5' TCACCACCTTCTTGATGTCATCA 3'
Gapdh Probe	5' ACGGGAAGCTCACTGGCATGGC 3'

Table 2.4. Taqman probes and primers used in ChIP qPCR reactions.

Probe/Primer name	Sequence
HS-31 F	5' TTCTGACCTCACCTCAGCTAAGC 3'
HS-31 R	5' TGTGTGGGCAGAGGACACA 3'
HS-31 probe	5' TCTTCCTCCTCTGAGAATCCGCCATG 3'
3' F	5' TGGCTCCGTGCATTTTCC 3'
3' R	5' CACCCTTCCAGAGAGTGGGTAAG 3'
3' probe	5' TGCAGGTCCAAGACACTTCTGATTCTGACA 3'
Ex2 F	5' CACCAAGACCTACTTCCCTCACTT 3'
Ex2 R	5' AGAGCATCGGCGACCTTCT 3'
Ex2 probe	5' CCACGGCTCTGCCCAGGTCAAG 3'

Table 2.5. Mouse *α-globin* Taqman probes and primers used in ChIP experiments.

inversely proportional to the log of original template DNA ($2^{-\log Ct}$). Reactions were normalised to the Ct value of the *gapdh* gene. For ChIP material, the difference between the immuno-precipitated Ct value to that of the input material (ΔCt) was derived and fold enrichment of template material calculated ($2^{-\Delta Ct}$).

2.5.3 Design of custom Agilent ChIP-chip microarrays

ChIP-chip microarrays were designed and ordered from Agilent Technologies using the 'e-array' online design tool. Arrays were ordered in both a 2 x 105 K format which included 2 arrays per slide or a 4 x 44 K format, including 4 arrays per slide. 2 x 105 K arrays were designed to incorporate agilent probes spanning the murine *Runx1*, *Runx2*, *Runx3*, *CBFβ* and *α-globin* loci and the 4 x 44 K arrays were designed to cover the murine *Runx1* and the *α-globin* loci. Precise coordinates of the regions covered by the probe sets for each gene are shown in Tables 2.6 and 2.7.

2.5.4 Labelling and amplification of ChIP material

ChIP material was labelled and amplified using the Bioprime DNA labelling system (Invitrogen). 100-450 ngs of ChIP DNA were denatured at 100 ^{o}C in the presence of random primers, cooled on ice and then left to react O/N at 37 ^{o}C in a mix containing 1 x dNTPS, Cy3 (Input material) or Cy5 (ChIP material) and Klenow. The samples were then purified using illustra Microspin columns (GE Healthcare) and the amount of DNA and integration

Gene	Chromosome	Coordinates
Runx1	16	91700000 - 93800000
Runx2	17	43900000 - 46000000
Runx3	4	133800000-135900000
CBFβ	8	106700,000-108900000
α-globin	11	31903000 - 32300000

Table 2.6 Coordinates (UCSC build NCBI37/mm9) within which probes for each gene indicated were selected on 2 x 105K Agilent ChIP-chip custom arrays.

Gene	Chromosome	Coordinates
Runx1	16	91480000 - 93880000
α-globin	11	31439030 - 32439029

Table 2.7 Coordinates (UCSC build NCBI37/mm9) within which probes for each gene indicated were selected on 4 x 44K Agilent ChIP-chip custom arrays.

of the respective dyes checked using the Nanodrop Spectrophotometer (Thermo Fisher Scientific).

2.5.5 ChIP-chip Hybridization

Hybridization was performed according to the 'Agilent Mammalian ChIP-on-chip' protocol (Agilent Technologies), for both the 2 x 105K for the 4 x 44K array formats. For the 2 x 105K arrays, 4.5 μg of Cy3-labelled and 4.5 μg of Cy5-labelled material were used. For the 4 x 44K arrays, 2.5 μg of each labelled material were used. For both formats, the Cy3 and Cy5 labelled material were incubated together for 3' at 95 ^{o}C with 1mg/ml Mouse Cot-1 DNA (Invitrogen), 1x Agilent blocking agent and 1x Agilent hybridization buffer. Samples were then heated at 37^{o}C for 30'. Each array on the 2 x 105K array was then covered with 250 μl of the hybridization sample. In the case of the 4 x 44K slide, each array was covered with 100 μl of the hybridization sample. 2 x 105K arrays were left to hybridize at 65 ^{o}C at 20RPM in a rotisserie oven for 40 hours whilst 4 x 44K arrays were left for 24 hours. Following the O/N hybridization, the arrays was serially washed in Agilent wash buffers 1 and 2 and then scanned immediately using the Agilent DNA Microarray Scanner with Surescan high-resolution technology (Agilent Technologies). Data from the scanned array images were extracted using Agilent Feature Extraction software and converted to WIG files suitable for use on the UCSC Genome Browser (http://genome.ucsc.edu/) using a PERL script kindly created by Dr. Jim Hughes.

2.5.6 Chromosome Conformation Capture (3C)

3C was performed using 2 x 10^7 cells per assay. Cells were first fixed for 10 minutes in 10 ml of 2% fixative (2% formaldehyde in cell-appropriate media containing 10% FCS) to cross-link chromatin. The reaction was quenched by adding 1.5 ml 1M glycine. Cells were then pelleted by brief centrifugation, washed in PBS and lysed in 5 ml cold lysis buffer (10 mM Tris pH8, 10 mM NaCl, 0.2% NP-40, complete proteinase inhibitor, 1 tablet/50 ml buffer) for 30 minute to 1 hour at 4^{o}C. Lysed samples were then spun down and washed in PBS, before being taken up in 500 μl of lysis buffer. After a further PBS wash step, chromatin was re-suspended in 1.2 x HindIII restriction buffer (60 μl Roche buffer H + 440 μl H_2O). SDS was added to the chromatin-restriction buffer mix to a final concentration of 0.3% and left to shake horizontally at 37^{o}C in a hybridization oven (Amersham) for 1 hour. After re-suspension of the resulting nuclear aggregates using a P1000 pipette, TX-100 was added to a final concentration of 2% and the sample was left to shake horizontally at 37^{o}C for 1 hour. After further re-suspension, 400 units of HindIII restriction enzyme (Roche) were added and the reaction was left to shake O/N at 37^{o}C. The following morning, samples were spun down briefly and re-homogenized as described above. 20% SDS was added to the chromatin (final concentration 1.3%) before heating at 65^{o}C for 20-25 minutes. Samples were then transferred into 7 ml 1 x ligation buffer (10 x stock: 660 mM Tris 1M pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 10 mM ATP) including TX-100 (final 1%) and incubated at 37^{o}C, with occasional shaking. After cooling on ice for 2 minutes, 100 units of T4 DNA ligase (Roche) were added and the sample was incubated at 16^{o}C for 4 hours followed by a further

30 minutes at RT. 300 μg Proteinase K (Fermentas) were then added and the sample was left to de-crosslink at 65°C O/N. The following morning, the sample was incubated at 37°C with 30 μl RNase (Fermentas) for 30-45 minutes. Chromatin was finally purified by phenol chloroform extraction and precipitated in 100% EtOH + 3M Sodium acetate (pH 5.6). After cleaning in 70% EtOH, pellets were air-dried and re-suspended in 500-1000 μl water (Sigma), depending on pellet size.

2.6 Analysis of 3C products

2.6.1 TBE gel electrophoresis

3C qPCR products were analysed by electrophoresis using 0.8% TBE gels (Invitrogen) to check for the generation of single-band products. Briefly, 10-15 μl of qPCR product plus DNA loading dye (Fermentas) were loaded into a gel in an XCell Surelock tank (Invitrogen) filled 1/3 full with 1 x TBE buffer. The gel was then run at 80 V for 45 minutes to an hour and post-stained in 100 ml 1 x TBE containing 5-10 μl EtBr. After a rinse in 1 x TBE, the gels were finally imaged using a Gel Doc 2000 (Biorad).

2.6.2 3C qPCR

3C qPCR was performed using Taqman probes and primers as per section 2.25. For a full list of probes and primers used and their genomic locations in

UCSC mouse genome build mm9, see Appendix I1. Mouse *α-globin* and *xpb* 3C probes and primers are shown in Table 2.8.

2.7 Mice

2.7.1 Generation and analysis of LacZ*-reporter F0 transgenic embryos*

LacZ-reporter transient transgenic (F0) embryos were generated by injection of reporter constructs into 1-cell stage mouse embryos and harvested at different embryonic stages (usually E8.5, E10.5 and E14.5). Embryos were first fixed in X-gal fix (PBS supplemented with 1% formaldehyde, 0.2% glutaraldehyde, 2mM $MgCl_2$, 5mM EGTA and 0.02% Nonidet P-40) for 1 hour at 4°C for E8.5 and E10.5 embryos. E14.5 embryos were first fixed for 30 minutes and then cut in two above the region of the fetal liver, and fixed for a further 30 minutes, to allow better penetration of fixative into the core of the embryo. Embryos were then washed twice in X-gal wash (PBS supplemented with 0.02% Nonidet P-40) for 2 minutes each at RT. Embryos were then stained O/N with Xgal-Stain (PBS supplemented with 5mM $K_3Fe(CN)_6$, 5mM $K_4Fe(CN)_6.3H_20$, 2mM $MgCl_2$, 0.01% sodium deoxycholate, 0.02% Nonidet P-40 and 1mg/ml Xgal) in the dark at RT. The following morning, embryos were washed twice in PBS before being imaged using a Leica DFC 300F digital camera. Transgenic embryos (based on visible expression pattern and genotype) were then transferred to a solution of 15% sucrose and left at 4°C until they sank into the solution (0.5 hour to O/N depending on the developmental stage of the embryo). When saturated with

Probe/Primer name	Sequence
24 H3 F	5' ATGGGACAATGGAGGGCTAAG 3'
HS31 H3 F	5' AGCTGTGCCTGGCAGATGAC 3'
a2 H3 R	5' GTGGCCATGAGCCAAAGAAT 3'
a2 H3 Probe	5' CCTCCAGGTACCATAAATCTCAGACTTG 3'
xpb F	5' CTTCTGCTACAAAGCGCTGACTT 3'
xpb R	5' TGCCCTCCCTGAAAATAAGGA 3'
xpb Probe	5' TGCACCCTGCTTTAGTGGCCTTGTAGG 3'

Table 2.8 Mouse *α-globin* and *xpb* Taqman probes and primers used in 3C experiments.

sucrose solution, the embryos were finally embedded in O.C.T compound (Sakura Finetek) and snap frozen in moulds on a mixture of dry ice + EtOH, before being sectioned using a Leica CM 3050s cryostat. Sections were made at a thickness of 10 μM and collected on Superfrost microscope slides (VWR). Once dried, slides were submerged in PBS for 5 minutes to remove O.C.T and then cover-slipped using 22 x 64 mm cover glasses (VWR), using Kaiser's glycerol gelatine (Merck) to adhere the cover glasses to the slides. Sections were then analyzed for β-galactosidase using an Eclipse E600 camera (Nikon).

2.7.2 Generation and analysis of GFP-*reporter transgenic embryos*

GFP-reporter transient transgenic embryos (F0) were generated in the same way as for *LacZ*-reporter transient transgenics (see 2.32). Following harvest, embryos were fixed in 4 % PFA in PBS at 4°C for 0.5 hour to 2 hours (depending on the developmental stage of the embryo). Next, the embryos were washed twice in PBS at room temperature before being submerged in 15 % sucrose and prepared for sectioning as for *LacZ*-reporter transient transgenics (see 2.32).

2.7.3 Animal husbandry – by Gemma Swiers (GS), John O'Rourke (JOR)

All mice were housed with free access to food and water and handled in accordance with Home Office regulations. This study used three transgenic mouse lines; hsp68*GFP*+110-Line 5, *hsp68GFP*+23LF-Line 3 and the *Runx1*$^{Lz/+}$ knock in mouse line (kind gift of Prof. Nancy Speck, UPENN,

U.S.A.), which were all maintained on a mixed (CBAxC57BL/6) background. Weaning of transgenic offspring was done approximately 3 weeks after birth. For analyses of embryonic stages of each line, timed matings were set up between F1 transgenic males and (129S1xC57BL/6)/F1 wildtype females. The detection of a vaginal plug the morning following mating was designated as E0.5. At days E8.5, E10.5 and E14.5 pregnant females were sacrificed and embryos were harvested and dissected in PBS + 10 % FCS + 50 U/ml penicillin and 50 μg/ml streptomycin. The precise developmental stage of embryos between E8.5 and E10.5 was determined by counting the number of somite pairs of the embryo.

2.7.4 Preparation of embryonic tissues for cell sorting

Embryonic tissues were sub-dissected and treated with 0.12% w/v type I collagenase (Sigma) at 37^{o}C for 20 minutes. Tissues were mechanically separated into single-cell suspensions by passing up and down < 20 times using a P1000 pipette. Next, to dilute-out the collagenase a suitable volume of PBS + 10 % FCS + 50 U/ml penicillin and 50 μg/ml streptomycin was added to the single cell suspensions. Following brief centrifugation at 1000 rpm, cells were re-suspended in PBS + 10 % FCS + 50 U/ml penicillin and 50 μg/ml streptomycin and counted using a Neubauer hemocytometer following treatment with trypan blue (Invitrogen) to exclude dead cells.

2.7.5 Preparation of adult tissues for cell sorting

Spleens and thymi from adult mice were made into single cell suspensions by passing through a 70 μM cell strainer (BD Falcon) into PBS + 10 % FCS + 50 U/ml penicillin and 50 μg/ml streptomycin. Bone marrow was aspirated from femora using a 22-gauge needle, into a 70 μM cell strainer (BD Falcon) (to exlude bone fragments) and single cells collected in PBS + 10 % FCS + 50 U/ml penicillin and 50 μg/ml streptomycin. Cells were then washed and counted as in 2.35.

2.7.6 Flow Cytometry

GFP was detected in the FL1 channel of a MoFlo flow cytometer (DakoCytomation). Dead cells and debris were excluded from analyses and from sorts on the basis of Hoechst 33258 uptake and forward scatter characteristics. GFP^+ and GFP^- sorted cells were collected in 100% 0.45 μm filtered FCS, washed in PBS and viable cells counted using a Neubauer hemocytometer. Flow cytometry was performed as described in (Nottingham et al., 2007). For antibodies used in flow cytometry, see Table 2.9.

2.7.7 Methylcellulose colony assay (CFU-C)

CFU-C assays were performed using tissues dissected from E11.5 *hsp68GFP*+110-Line 5 embryos (typically AVU, FL and YS). Cells were sorted as described in 2.42, and plated in Methocult GF 3434 (Stem Cell Technologies) in a volume of 1 ml per 35 x 10 mm Cellstar tissue culture dish (Greiner Bio-One). This was usually performed in duplicate. Plates were then

Antibody	Clone	Working Conc.	Company	Cat. No.
B220-APC	RA3-6B2	2 ug/ml	BD	553092
cKit-APC	2B8	2 ug/ml	BD	553356
CD19-APC	1D3	2 ug/ml	BD	550992
CD45-APC-Cy7	30-F11	2 ug/ml	BD	559864
CD4-BIO	GK1.5	0.5 ug/ml	BD	553728
CD8-BIO	53-6.7	0.5 ug/ml	BD	553028
Ter119-BIO	TER119	0.5 ug/ml	BD	553672
Gr1-BIO	RB6-8C5	0.5 ug/ml	BD	553125
CD4-PE	L3T4	2 ug/ml	BD	553048
Mac1-PE	M1/70	2 ug/ml	BD	557397
Sca1-PE	E13-161.7	2 ug/ml	BD	553336
CD31-PE	MEC 13.3	2 ug/ml	BD	553373
CD41-PE	MWReg30	2 ug/ml	BD	558040
CD45-PE	30-F11	2 ug/ml	BD	553081
VE-CAD AF647	ebioBV13	2 ug/ml	ebioscience	51144182
LY6C	AL-21	0.5 ug/ml	BD	553104
SAV-APC			Caltag	SA1005
SAV-PE-Cy7			Invitrogen	SA1012
Hoechst 33258			Invitrogen	H3569

Table 2.9. Antibodies used in FACS experiments.

incubated at 37^{o}C and 5% CO_2 for 7 days, after which hematopoietic colonies were counted. This was repeated 7 days later, after a total period of 14 days.

2.7.8 Long-term reconstitution analysis

Adult male (C57BL/6 x CBA) F1 mice were conditioned with a split dose of 9 Gy from a ^{137}Cs source. GFP^{+} and GFP^{-} sorted cells were isolated from E11.5 AVU as outlined in 2.38, and embryonic equivalents (ee) calculated. Cells were injected via the tail vein of irradiated recipients, together with 2 x 10^5 (C57BL/6 x CBA) F1 spleen cells to promote survival. In order to analyse donor-derived hematopoietic reconstitution, recipient mice were bled by Dr Marella de Bruijn at 1 to 2 and 4 to 6 months post-transfer. Genomic DNA was isolated from blood samples and level of donor reconstitution determined by *GFP*-specific semi-quantitative PCR (see Table 2.2).

2.8 Statistics

The Student's two-tailed t-test was applied to evaluate statistical significance between data sets. A P value of <0.05 was considered statistically significant.

Chapter Three:

Identification and initial analysis of candidate *Runx1 cis*-regulatory elements

3.1 Chapter Aim

This aim of this chapter is to screen the *Runx1* syntenic region for novel *Runx1* enhancers. The syntenic region covers 1.3 Mb of sequence from *Morc3* upstream of *Runx1* to *Clic6* downstream of *Runx1* (see Figure 3.8). Genomic sequences will be cloned into the pGL3-promoter vector for *in vitro* enhancer assays if they show all, or a combination of, the following criteria: i) enrichment for an epigenetic mark of enhancers; ii) hypersensitivity to DNaseI treatment and iii) binding of upstream transcription factors.

3.2 The identification and characterization of *Runx1 cis*-regulatory elements

3.2.1 The +23 enhancer was found in a screen of a 270 Kb genomic region spanning the Runx1 locus

One of the aims of our lab is to understand the *cis*-regulatory control of *Runx1* during developmental hematopoiesis. To achieve this we have set out to identify all hematopoietic *Runx1 cis*-regulatory elements and characterize their enhancer activity during embryonic development. The first *Runx1 cis*-regulatory element identified, the +23 enhancer, was found in a screen for *cis*-regulatory elements within a 270 Kb region spanning *Runx1* plus 30 Kb of upstream and 15 Kb of downstream sequence. The screen was performed using two methods (Figure 3.1). Firstly, the Gumby algorithm was used to

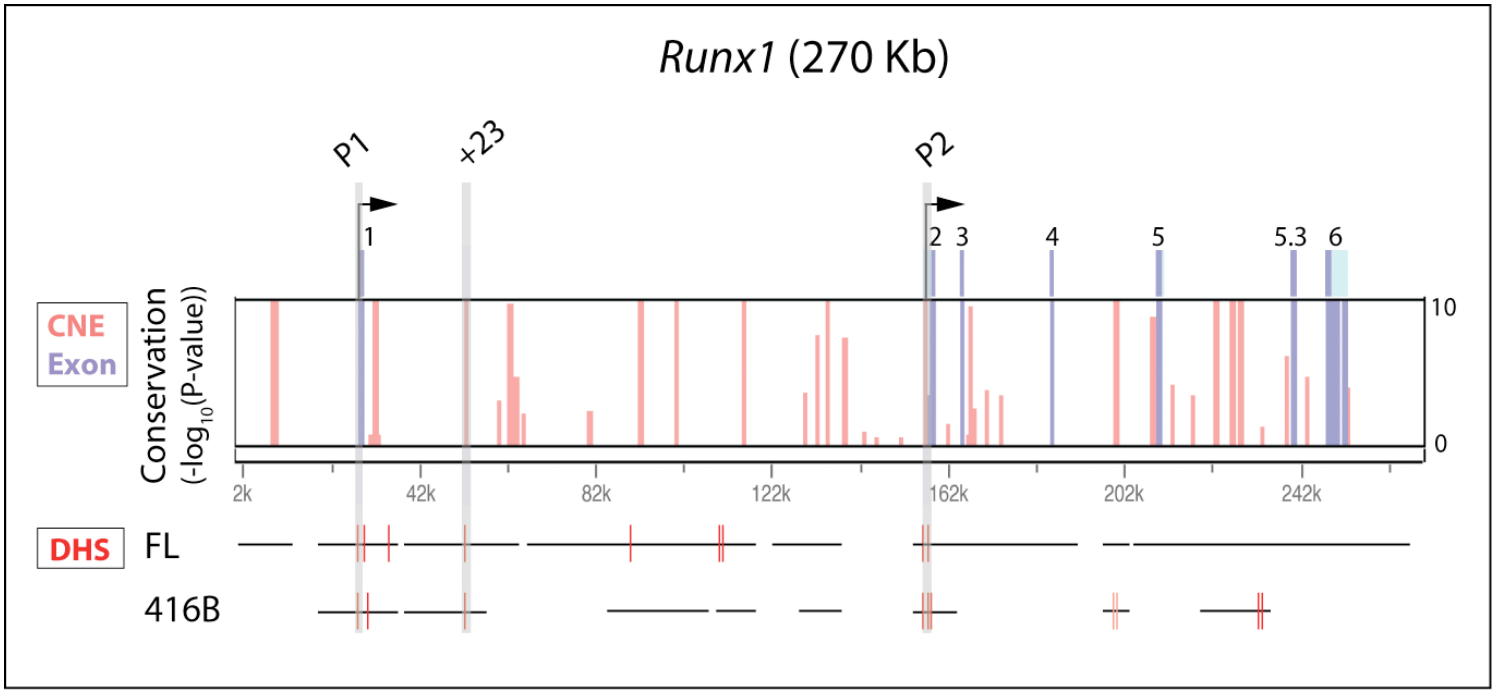


Figure 3.1. **Original identification of candidate *cis*-regulatory elements in the mouse *Runx1* locus**. Gumby output and DHS (Chr 16: 92473100–92743100 spanning *Runx1* plus 30 kb upstream and 15 kb downstream; NCBIM build 36). *Runx1* exons numbered 1 to 6 are indicated; coding sequence [dark blue], UTRs [light blue]). Gumby-identified CNEs (pink) and exons (blue) are shown as colored bars. Height of bars shows -log10(*P* value) (cutoff at 10). Below Gumby identified CNEs are hematopoietic DHS detected along the mouse *Runx1* locus (vertical red bars) in the area analyzed (horizontal black lines). Abbreviations: CNE, conserved non-coding element; DHS, DNaseI hypersensitive site; FL, E12.5 fetal liver; 416B, myeloid progenitor cell line. Adapted from (Nottingham et al, 2007).

score non-coding regions with a likelihood of *cis*-regulatory function, based on cross-species conservation. The algorithm predicted 42 such regions within the 270 Kb fragment including the +23 conserved non-coding element, or 'CNE'. Also included in the list of 42 Gumby-predicted CNEs were the P1 and P2 promoters. Secondly, in DNaseI hypersensitive site assays in E12.5 FL cells and 416B myeloid progenitor cells, the +23 CNE was also found to be hypersensitive to DNaseI treatment, indicative of an open chromatin state in these cell types, accessible to the binding of upstream regulatory proteins. The P1 and P2 promoters were the only other regions from the 42 Gumby CNEs shown in Figure 3.1 to map to corresponding hypersensitive sites in both the E12.5 FL and 416B cells (grey bars).

The +23 enhancer was shown to be active *in vitro* in 416B myeloid progenitor cells using the dual-luciferase reporter assay (Nottingham et al., 2007). This enhancer activity was confirmed *in vivo* where it mediates *Runx1*-specific expression to HSCs and hematopoietic progenitor cells in hsp68*LacZ*+23 and hsp68*GFP*+23 transgenic embryos. Expression was also seen in a subset of endothelial cells and a few mesenchymal cells where *Runx1* is expressed (Bee et al., 2009b; Nottingham et al., 2007). However, the +23 enhancer is unable to mediate expression to all mesenchymal cells where endogenous *Runx1* is expressed in the DA at E10.5 (compare Figure 1.5 B v with 1.5 i). Therefore, are there other site or temporal-specific *Runx1* enhnacers that mediate sub-sets of *Runx1* expression during developmental hematopoiesis? Such information would be useful to understand how the process of developmental hematopoiesis develops.

3.2.2 Identifying novel Runx1 enhancers: a 1.3 Mb screen in the Runx1 syntenic region

The identification of the +23 enhancer within the 270 Kb fragment demonstrates that *cis*-regulatory elements can be identified using a combination of bioinformatics and biochemical assays. We therefore decided to extend the search for novel *Runx1* enhancers to a 1.3 Mb region covering the *Runx1* syntenic region. Within this syntenic region, the order of the Runx genes is conserved between multiple species, which suggests that *cis*-regulatory elements would fall within this region. In addition, work by the Bickmore group proposes that the most mobile gene loci in human cells have an average range of movement of 0.5 μm within the nucleus (approximately 1 MB of sequence) (Chubb and Bickmore, 2003). This therefore raises the possibility that enhancers up to 1 Mb away from either of the *Runx1* P1 or P1 promoters can freely interact with a promoter in cells expressing *Runx1*.

To search effectively for *Runx1* enhancers over the proposed 1.3 Mb distance, we based the design of our screening approach on data being published at the start of this study. Whole genome ChIP-chip and ChIP-seq data from mouse and human showed that *cis*-regulatory elements could be identified by enrichment for element-specific epigenetic marks (Barski et al., 2007; Bernstein et al., 2006; Birney et al., 2007; Heintzman et al., 2009; Heintzman et al., 2007; Mikkelsen et al., 2007; Roh et al., 2007) and by the regulatory proteins that bind to them (Visel et al., 2009; Wilson et al., 2009).

Active promoters are enriched for tri-methylation of the Lys4 residue on the tail of histone 3 (H3K4Me3) and a depletion of monomethylation at the same residue (H3K4Me1) at the transcription start site (TSS). Conversely, active enhancers are enriched for H3K4Me1 and a lack of H4K4Me3. H3K4Me2 and acetylation of Histone3 (AcH3) are known to mark *both* active enhancers and promoters (Heintzman et al., 2007). ChIP-chip for these markers was therefore performed as the method for finding additional *Runx1* enhancers.

3.2.3 Histone modifications at known Runx1 CNEs

In advance of a ChIP-chip screen, I investigated the enrichment of AcH3, H3K4Me1, H3K4Me2 and H3K4Me3 at four *Runx1* CNEs using ChIP. This was done to verify that these markers could distinguish between promoters and enhancers in the *Runx1* locus. The elements tested by qPCR were: the +23 enhancer, the +24 booster fragment (see Chapter 1, Section 1.5.4) and the P1 and P2 promoters. The ChIP assays were performed using material from four different cell types: i) 416B myeloid progenitor cells; ii) E11.5 WT FL cells; iii) E12.5 WT FL cells and iv) E11.5 WT AGM cells. As 416B and E12.5 FL cells were used in the DNaseI screens for these CNEs in the 270 Kb fragment (Figure 3.1) (Nottingham et al., 2007) it seemed likely that histone markers should be enriched in the expected pattern in these cells. E11.5 AGM cells were used because the AGM is the major site of *de novo* HSC generation in the embryo. The future identification of active enhancers in this cell pool in the ChIP-chip screen will be of great interest.

In 416B cells, AcH3, H3K4Me2 and H3K4Me3 enrichment was tested at the P2 and +24 CNEs. In addition, the '+70' CNE was used as a negative control in this cell type. The +70 CNE is located within intron 1 of *Runx1* and is deeply conserved to frog. However, it is not a DNAseI hypersentitive site in 416B cells or E12 FL cells and binds little to no upstream TFs (Nottingham et al., 2007). In these ChIP assays, the P2 promoter was enriched for AcH3 and H3K4Me2/3 (Figure 3.2). In comparison, the +70 CNE showed very little enrichment for these markers. The +24 element also showed enrichment for AcH3 and H3K4Me2 indicating its active conformation in this cell type, and a small enrichment for H3K4Me3, usually associated with promoter function.

In E11.5 wild-type FL cells, a comprehensive analysis of the histone modifications present at the +23, +24 and the P1 and P2 promoters was carried out (Figure 3.3). The +24 showed greater enrichment for H3K4Me1 and H3K4Me2 than the +23. Both promoters were enriched for H3K4Me1, with the P2 showing less enrichment for the marker than the P1. The P2 showed the greatest enrichment for H3K4Me2 and H3K4Me3 in comparison to the other elements. In E12.5 FL cells, both promoters were enriched for H3K4Me3 and all elements were enriched for the active markers H3K4Me1 and H3K4Me2 (Figure 3.4).

In E11.5 wild-type AGM cells a different pattern of enrichments was observed (Figure 3.5). The P2 promoter showed greater H3K4Me1 enrichment in comparison to E11.5 FL cells and was the only promoter to have high enrichment for H3K4Me2/3. The P1 does not show H3K4Me3

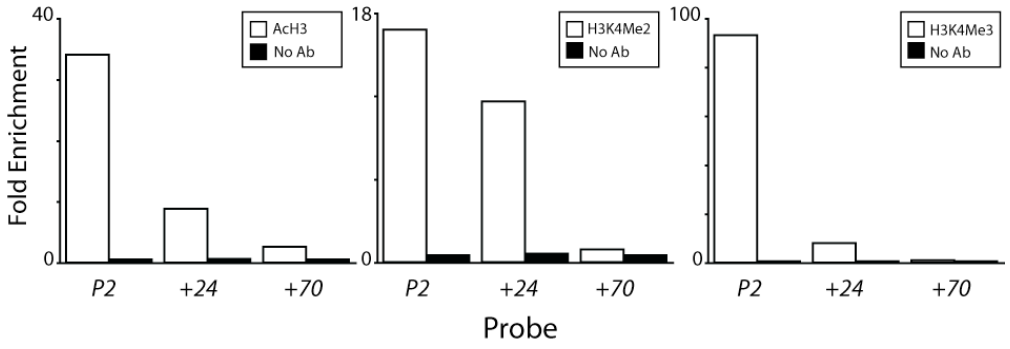


Figure 3.2 Histone modifications at the P2 promoter and +24, +70 CNEs in 416B cells. Real-time PCR analysis of ChIP with antibodies directed against AcH3, H3K4Me2 and H3K4Me3 (1 x 10^7 cells/antibody). Data are the mean of one ChIP experiment with qPCR performed in duplicate.

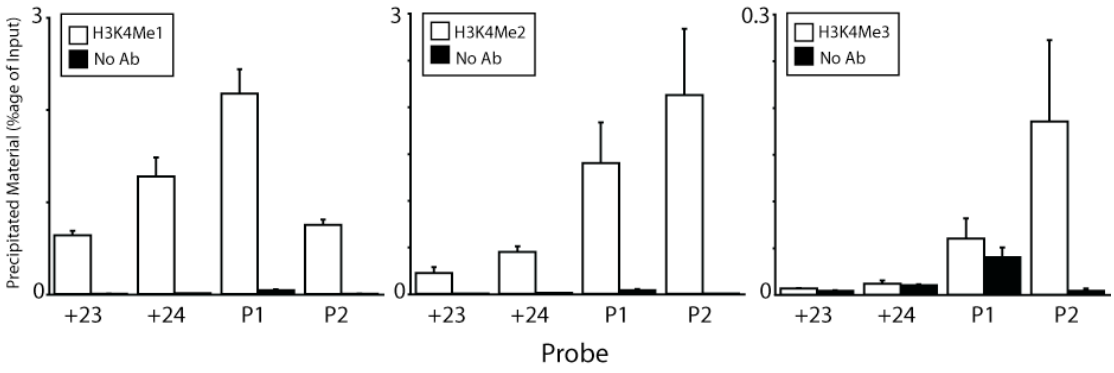


Figure 3.3 *In vivo* histone modifications at the +23 and +24 CNEs and P1, P2 promoters in E11.5 FL cells. Real-time PCR analysis of ChIP with antibodies directed against H3K4Me1, H3K4Me2, H3K4Me3 and total Pol II. Data are the mean (+SD) from two independent ChIP experiments with two real-time PCR assays per ChIP. Average number of cells per antibody = 6 x 10^6.

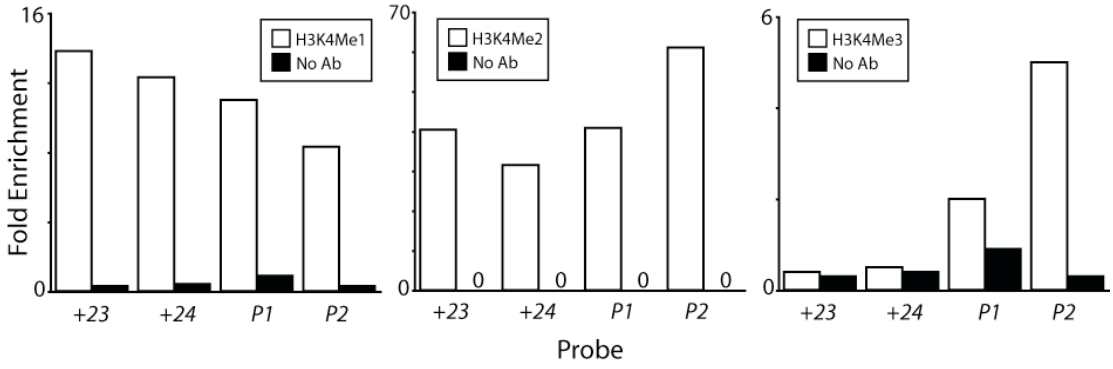


Figure 3.4 *In vivo* histone modifications at the +23, +24 CNEs and P1, P2 promoters in E12 fetal liver cells. Real-time PCR analysis of ChIP with antibodies directed against H3K4Me1, H3K4Me2, H3K4Me3 and total Pol II (1 x 10^7 cells/antibody). Data are the mean of one ChIP experiment with qPCR performed in duplicate.

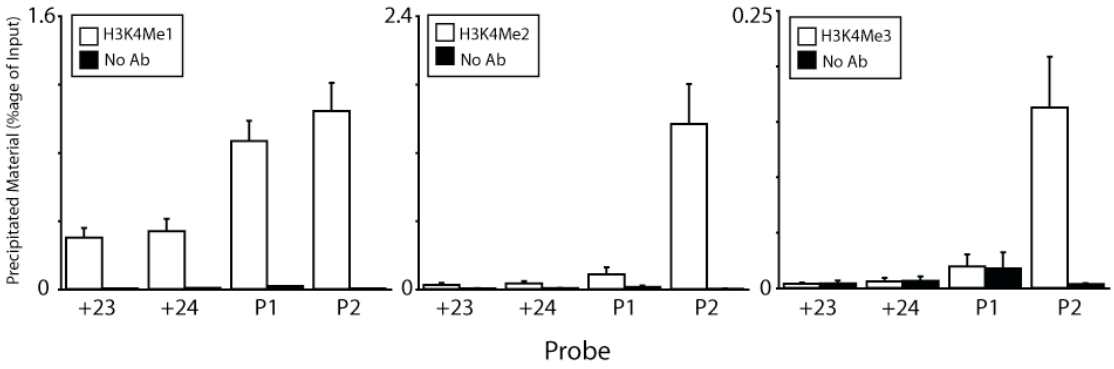


Figure 3.5 *In vivo* histone modifications at the +23, +24 CNEs and P1, P2 promoters in E11 AGM cells. Real-time PCR analysis of ChIP with antibodies directed against H3K4Me1, H3K4Me2, H3K4Me3 and total Pol II. Data are the mean (+SD) from two independent ChIP experiments with qPCR performed in dupliate per ChIP. Average number of cells per antibody = 4 x 10^6.

enrichment over the no-antibody control. Both the +23 and +24 showed enrichment for H3K4Me1.

3.2.4 Interim Conclusions

In each cell type tested in section 3.2.3, the P2 promoter was enriched for all active markers, including H3K4Me1. The major difference between cell types was in E11.5 AGM cells, where, compared to E11.5 FL cells, the P2 showed greater enrichment for H3K4Me1 than the P1 and the +23 and +24. Implicating the P2 with enhancer activity in E11 AGM cells from these data would be false, as both promoter fragments have been shown not to confer any *Runx1*-specific staining in P1*LacZ* and P2*LacZ* F0 transgenic embryos (Bee et al., 2009a). Rather, the enrichment represents the known pattern of H3K4Me1 levels at active promoters, where enrichment for the modification peaks more than 1kb downstream and upstream of the TSS (Heintzman et al., 2007) and could reflect a more active state of the P2 in AGM cells compared to the P1. This has been confirmed by qPCR analysis of transcripts coming from both promoters in these tissues, where, at E11.5, more transcripts are generated by the P2 in the AGM, whereas, at the same stage of development in the FL, the transcripts derived from the P1 promoter are more prevalent than those from the P2 (Bee et al., 2009b). Low enrichment for H3K4Me3 at the P1 could be accounted for by the positioning of the real-time probe being upstream of the core promoter.

In embryonic tissues, the +23 and +24 were consistently enriched for H3K4Me1 and H3K4Me2 apart from in E11.5 AGM cells, where elements show less H3K4Me2 enrichment. The small enrichment for H3K4Me3 seen at the +24 in 416B cells is also interesting in the light of unpublished data where the combined +23, +24 (+23+24) element has weak promoter activity *in vitro* in transiently transfected 416B cells and *in vivo* in F0 transgenics (Thomas Bee, unpublished data). In addition, in each cell type tested, the +24 shows enrichment for at least one marker that would predict an active enhancer. However, the +24 has been shown to have no classical tissue-specific enhancer activity of its own in 416B cell luciferase assays and in F0 transgenic embryos. The +24 does however act to boost the activity of the +23 (as the +23+24 'LF' regulatory unit) resulting in a higher frequency of F0 TG embryos expressing *LacZ* at the sites of hematopoiesis (Wade Nottingham, unpublished data). The +24 could therefore be classed as a booster enhancer, acting to maintain the +23 in a chromatin conformation more permissive to either upstream regulatory proteins and/or the general transcriptional machinery in these cell types. Such a booster enhancer exists in the *Scl* locus, known as the '+18' (Smith et al., 2008). It functions to boost the activity of the *Scl* +19 enhancer, which directs *Scl*-specific expression to HSCs, hematopoietic progenitors and the endothelial cells.

In conclusion, known *Runx1 cis*-regulatory elements follow the general pattern of histone marks expected, based on their regulatory function. As such, using these markers when looking for novel *cis*-regulatory elements located within the *Runx1* syntenic region in ChIP-chip experiments was

deemed to be reliable. At the same time, the +24 data highlight the need for functional validation of elements in *in vivo* assays after their initial identification.

3.3 ChIP-chip: a screen for novel *cis*-regulatory elements

3.3.1 Establishing the ChIP-chip method

ChIP-chip technology allows the investigator to hybridize ChIP material to microarrays tiled with DNA probes of choice. Probes can be ordered on the arrays in such a way that the sequence of a gene/genes and flanking 5' and 3' regions are represented to the user's specifications (Ren et al., 2000), reviewed in (Schones and Zhao, 2008). There are several advantages to using ChIP-chip over the traditional ChIP-qPCR method. The first is seeing enrichment of ChIP material over large genomic distances, using relatively little starting material. Covering such distances by traditional qPCR would be expensive, time consuming and require far more ChIP material. ChIP-chip also allows the investigator to identify regions of interest (in this case putative enhancers) in an unbiased way. There are no pre-conceived ideas based on conservation for example, which is important as not all enhancers are deeply conserved throughout evolution (see (Bulger and Groudine., 2011)).

To establish ChIP-chip in the lab, I first performed this method on a well-characterized locus, the *α-globin* locus, studied by the Higgs group in our department. A human *α-globin* ChIP-chip array had been used previously to identify putative *cis*-regulatory elements at the locus. This was done using ChIP material from primary erythroid cells in comparison to non-erythroid cells (De Gobbi et al., 2007). H3K4Me2-enriched material from human K562 erythroid cells is known to mark *α-globin* regulatory elements on the human *α-globin* ChIP-chip array (De Gobbi et al., 2007). I therefore generated H3K4Me2 ChIP material using the K562 cell line and hybridized it to the same array. My material hybridized in a pattern that mirrored the original experiment (Figure 3.6), successfully establishing ChIP-chip in our lab.

3.3.2 Additional control experiments prior to Runx1 ChIP-chip

Prior to performing *Runx1* ChIP-chip, further controls were done to verify that the quality of ChIP material was suitable for hybridization onto arrays. Real-time PCR experiments using Taqman primers/probes for defined regions of the mouse *α-globin* locus were carried out. These included the enhancer element HS-31, Exon2 (which contains the promoter minus coding sequences) and the non-regulatory element '3'' (kind gift of Dr. Doug Vernimmen). ChIP material from E12.5 FL cells was used as the FL consists of mainly erythroid cells. As a negative control, H3K4Me2 material from 416B myeloid cells was also tested (*α-globin* is not expressed in 416B cells). Low H3K4Me1 enrichment was seen at all three *α-globin* elements in E12.5

FL cells. H3K4Me1 and H3K4Me3 were enriched at HS-31 and Ex2 but not at the 3'region as expected (Figure 3.7 A, left and centre panels). Also, H3K4Me3 only marked the Ex2 region containing the gene's promoters, (Figure 3.7 A, right-hand panel). Finally, in 416B cells, no H3K4Me2 enrichment was seen at any of the three regions, which fits with *α-globin* not being expressed in this myeloid-progenitor cell type (Figure 3.7 B). These results confirmed that this material was suitable for hybridization to arrays, with one caveat. The enrichment of H3K4Me1 (enhancer marker) was lower than H3K4Me2 (enhancer and promoter marker) at HS-31 in the E12.5 FL cells. Therefore, it was decided that H3K4Me2 material should be used in *Runx1* ChIP-chip assays, as it was evident that the H3K4Me1 antibody was less effective at pulling down chromatin.

3.3.3 Runx1 ChIP-chip, ChIP-seq and DNaseI-seq

Custom arrays were designed and ordered from Agilent (Agilent Technologies Inc, USA) using the Agilent e-Array web-based design tool (for more details about the arrays, see Appendix I). Prior to hybridization to these arrays, qPCR at *Runx1* regulatory elements was always performed using the ChIP material to verify its quality, as described in section 3.2.3. ChIP was performed with antibodies against H3K4Me2 and H3K4Me3 using E11.5 AGM cells, E12.5 FL cells and 416B myeloid progenitor cells. The ChIP-chip results using these materials are shown as peaks of enrichment across the *Runx1* syntenic region in Figure 3.8. As can be seen in this figure, there is widespread enrichment for H3K4Me2 and H3K4Me3 at the gene-containing

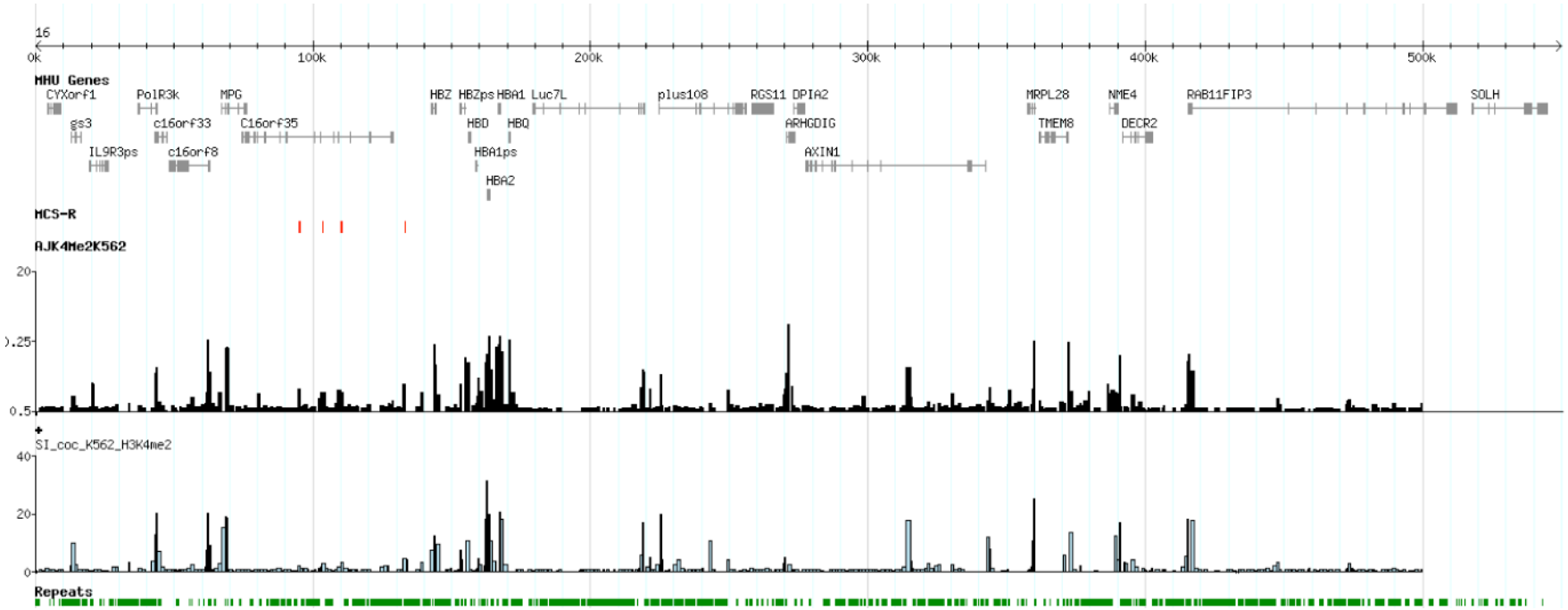


Figure 3.6 ChIP-chip analysis of H3K4Me2 enrichment across the *α-globin* syntenic region in the human K562 erythroid cell line. The x-axis shows the alpha-globin syntenic region and the y-axis represents enrichment of ChIP DNA over input DNA. The top track (black) shows data generated in my hands (1×10^7 cells/antibody, n=1). The bottom track shows data generated by the ENCODE project. Red boxes represent the alpha-globin *cis*-regulatory elements.

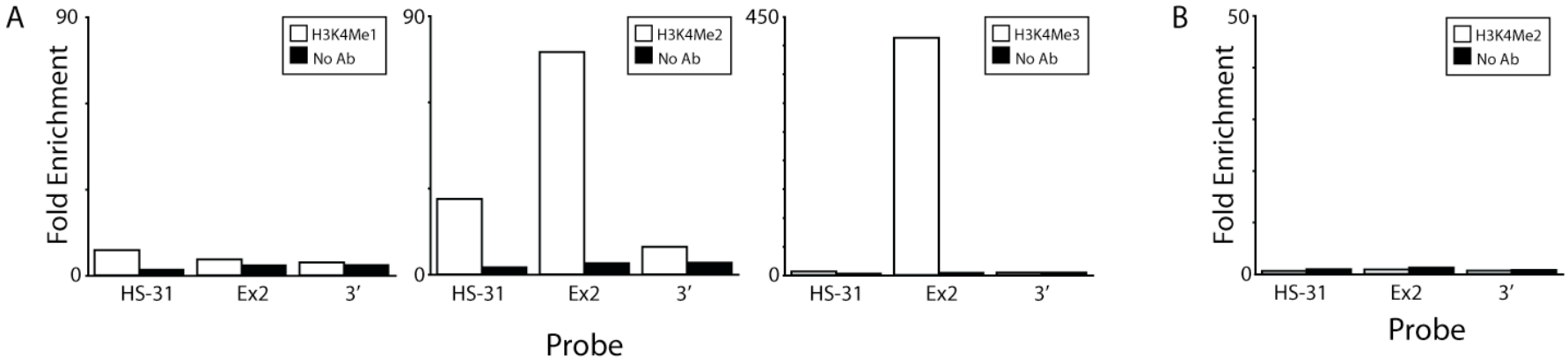


Figure 3.7 Pre-hybridization control of ChIP material: enrichment at known regulatory elements in the *α-globin* locus. (A) H3K4Me1, H3K4Me2 and H3K4Me3 enrichment at the HS-31 (enhancer) and Ex2 (promoter-containing) regulatory regions and the 3' non-regulatory region in E12.5 FL cells (1 x 10^7 cells/antibody). (B) H3K4Me2 enrichment at the same regions in 416B myeloid progenitor cells, where *α-globin* is not expressed (1 x 10^7 cells/antibody). Data are the mean of one ChIP experiment with qPCR performed in duplicate.

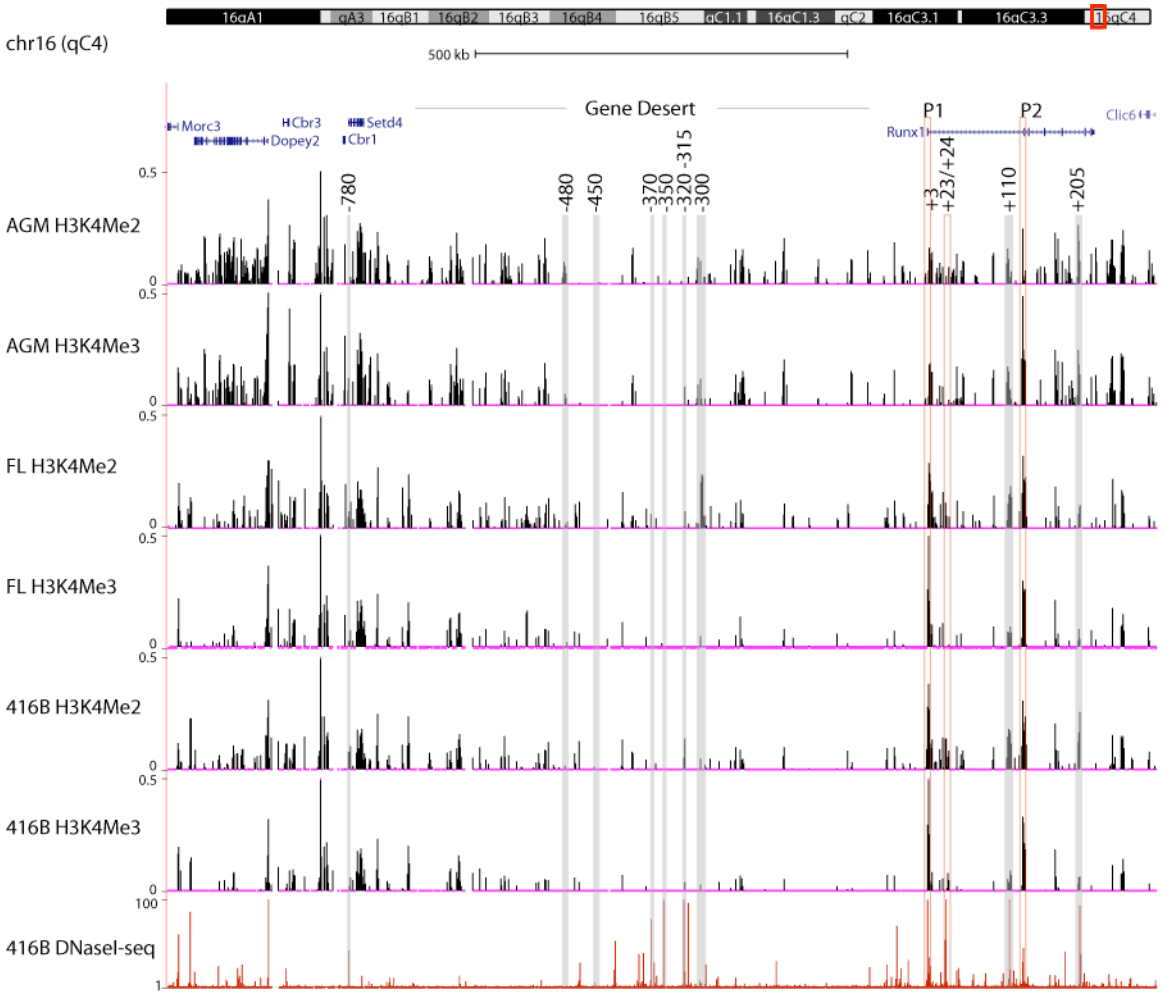


Figure 3.8. For legend, see overleaf.

Figure 3.8 ChIP-chip analysis of H3K4Me2 and H3K4Me3 enrichment and DNaseI-sequencing reads across the *Runx1* syntenic region. Above the track is a chromosome ideogram indicating the region represented below (red box) The x-axis shows the *Runx1* syntenic region, which covers 1.28 Mb from *Morc3* upstream of the gene to the beginning of *Clic6* downstream of *Runx1*. For ChIP-chip data, the y-axis represents enrichment of ChIP DNA over input DNA from one experiment per cell type. For DNase1-seq data, the y-axis represents enrichment for sequence reads from one experiment. The region upstream of *Runx1* is gene poor and is classified as a 'gene desert' as it spans an area greater than 500bp {Nobrega, 2003 #210}. ChIP material is from E11.5 AGM cells (3.86 x 10^6 cells/antibody), E12.5 FL cells (1 x 10^7 cells/antibody) and 416B myeloid cells (1 x 10^7 cells/antibody). *Runx1* elements selected for *in vitro* enhancer assays are shown as numbered grey regions in Kb from the *Runx1* ATG (- = upstream, + = downstream). Known *cis*-regulatory elements, the +23, +24, P1 and P2 are highlighted in clear boxes with red borders.

regions of the syntenic region, with less in the gene desert (between the end of *SetD4* and the start of *Runx1*). These results are also aligned with DNaseI-seq data from 416B cells, generated in collaboration with John Stamatoyannopoulos (University of Washington, U.S.A.). DNaseI-seq is a high-throughput version of the DNaseI hypersensitivity assays used to uncover the +23 enhancer. These assays were previously done in our lab by southern blot analysis (Nottingham et al., 2007), which was a slow and laborious process. DNaseI-seq however has the advantage of generating genome-wide data in far less time and so is a far more efficient technique (much like ChIP-chip Vs. qPCR). 416B DNaseI-seq peaks (red) overlap with many of the ChIP-chip peaks, particularly in the *Runx1* gene body, although in the gene desert the DNaseI-seq data reveal regions of hypersensitivity that aren't always enriched for H3K4Me2/3, even in 416B cells.

Owing to the large number of peaks in both the ChIP-chip and DNaseI-seq assays, a third data set was analyzed to help short-list the regions shown. Data were taken from a recent publication from the Gottgens' group, in which they performed genome-wide ChIP-seq assays for ten hematopoietic transcription factors: Runx1, Gata2, Scl, Lmo2, Fli1, Gfi1b, PU.1, Erg, Lyl1 and Meis1, using HPC7 mouse ES cell-derived hematopoietic progenitor cells (Wilson et al., 2010) (Figure 3.9). ChIP-seq results show distinct peaks of enrichment for TF binding across the syntenic region. Some regions are bound by all ten TFs e.g. the +23+24 enhancer, whereas others show binding of some TFs but not others. The results from the ChIP-chip, DNaseI-seq and ChIP-seq assays are shown in Tables 3.1 and 3.2. The peaks highlighted were selected as they

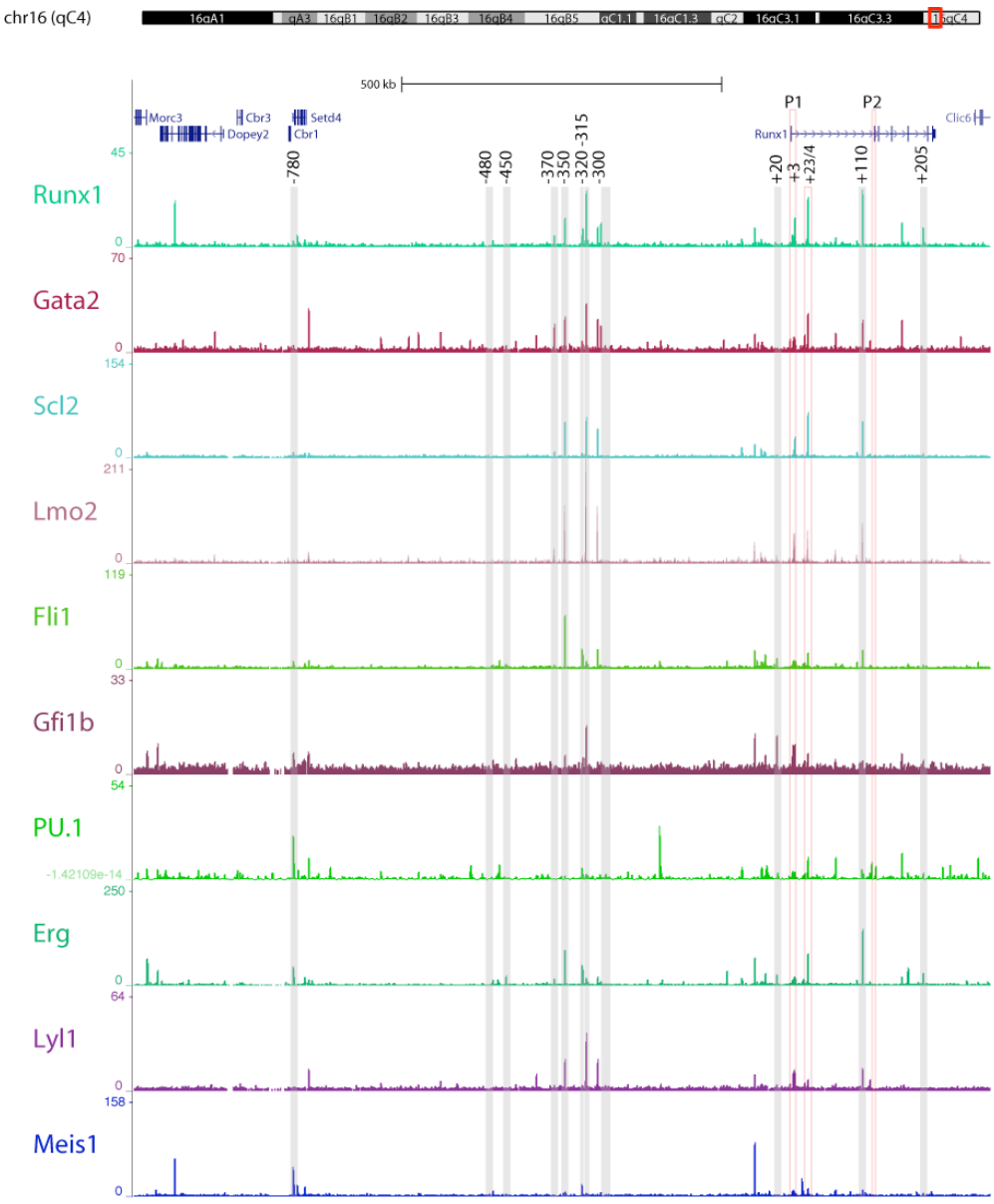


Figure 3.9 ChIP-Seq for ten hematopoietic TFs across the *Runx1* syntenic region in the mouse HPC7 cell line. Above the tracks is a chromosome ideogram indicating the region represented below (red box) The x-axis shows the *Runx1* syntenic region, which covers 1.28 Mb from *Morc3* upstream of the gene to the beginning of *Clic 6* downstream of *Runx1*. The y-axis shows enrichment for sequencing reads for the TFs listed. Genomic regions to be tested in *in vitro* enhancer assays are shown as numbered grey regions in Kb from the *Runx1* ATG (- = upstream, + = downstream). Data are from (Wilson et al, 2010).

	E11.5 AGM		E12.5 FL		416B Myeloid Progenitors		
	H3K4Me2	H3K4Me3	H3K4Me2	H3K4Me3	H3K4Me2	H3K4Me3	DNaseI HS
-780							
-480							
-450							
-370							
-350							
-320							
-315							
-300							
P1 Promoter							
+3							
+23							
+24							
P2 Promoter							
+110							
+205							

	HPC7, ES-derived hematopoietic progenitor cells									
	Runx1	Gata2	Scl	Lmo2	Fli1	Gfi1b	PU.1	Erg	Lyl1	Meis1
-780										
-480										
-450										
-370										
-350										
-320										
-315										
-300										
P1 Promoter										
+3										
+23										
+24										
+110										
P2 Promoter										
+205										

Tables 3.1 (top) and 3.2 (bottom). For figure legend, see overleaf.

Table 3.1 (top): Summary of enrichment of H3K4Me2/3 peaks in E11.5 AGM, E12.5 FL and 416B cells and DNaseI-seq peaks in 416B cells at 15 selected regions thoughout the *Runx1* syntenic region. The fifteen regions in the syntenic region are numbered in relation to their distance in Kb upstream (-) or downstream (+) of the *Runx1* ATG, except for the P1 and P2 promoters. Blue boxes = peaks for either H3K4Me2 or H3K4Me3 at the corresponding region/cell type (the blue box for H3K4Me2 enrichment in E12.5 FL cells is highlighted as the peak was particularly strong in this cell type). Red boxes = DNaseI-seq peaks at regions in 416B myeloid progenitor cells. White boxes = no enrichment.

Table 3.2 (bottom): Summary of binding of 10 hematopoietic transcription factors to 15 selected regions thoughout the *Runx1* syntenic region in HPC7 cells. Genomic regions are numbered as as in 3.1. Grey boxes = peaks of enrichment for the transcription factor at the corresponding region. White boxes = no enrichment. Data are from {Wilson, #128}.

represent a variety of putative enhancers. Those regions that are enriched for H3K4Me3/4, DNaseI hypersensitivity and the binding of hematopoietic TFs were the first to be selected as they are the most likely to have enhancer function. These include the +3 and +110 regions. In total, eleven new regions have been selected for further analysis in *in vitro* enhancer assays (as indicated as grey bars in Figure 3.8 and summarized in Table 3.1). Some elements are marked by varying combinations of enrichments for TFs, H3K4Me2/3 and DNaseI hypersensitivity, for example the -370 peak. Such regions may represent enhancer elements with more restricted activity in hematopoiesis, dependent on cell type/TF binding. Some regions are also included, for example the -450 region, despite being devoid of any peaks in any of the assays. However, previous analysis showed the -450 (and the -480 and -370) are enriched for GATA, Runx and CSL binding sites, and so has been included to see if these sites indicate enhancer activity, even in the absence of the peaks that would indicate enhancer potential. In conclusion eleven regions in the *Runx1* syntenic region with likelihood of enhancer activity have been selected (based on our selection criteria for active enhancers). It will be interesting to see if elements with fewer peaks of enrichment, or indeed no peaks like the -370, have enhancer activity.

3.4 *In vitro* enhancer activity of novel elements within the *Runx1* sytenic region

To test the enhancer activity of the eleven elements selected in section 3.3.3, each genomic region was cloned separately downstream of the firefly

Luciferase gene under the control of the SV40 minimal promoter to test in the dual-luciferase reporter assay (Promega) (Figure 3.10). In addition, the +3 and +110 were cloned together with the +23+24 sequence to make pGL3-Promoter+110+23+24 and pGL3-Promoter+3+110+23+24 composite constructs. This was also done for the -480, -450 and -370 elements, to see if composite elements in either the *Runx1* gene locus or in the gene desert could act cooperatively to increase enhancer activity in comparison to individual elements alone.

Luciferase assays were performed by transiently transfecting 416B myeloid progenitor cells and K562 human erythroid cells by electroporation. K562 cells were used, despite being of human origin, as there is extensive sequence conservation in the *Runx1* locus, particularly between putative *cis*-regulatory elements. Therefore we thought mouse enhancer sequences would function in human cells. It is also possible to efficiently electroporate K562 cells which is not the case for some mouse cell lines described in this study, for example the HPC7 hematopoietic progenitor line.

Luciferase assay results (Figure 3.11) show that the -350 and the +205 elements have significant enhancer activity in 416B cells ($p<0.05$). The same was true for the +23 enhancer ($p<0.01$), which has been described previously in (Nottingham et al., 2007) and acts as a positive control for this data set. In K562 cells, the -370, -480-370-350, -315, +23, and +110 were all shown to have significant enhancer activity (all with a p-value of <0.01, except for -480-370-350 which was <0.05). Therefore, five out of the eleven selected

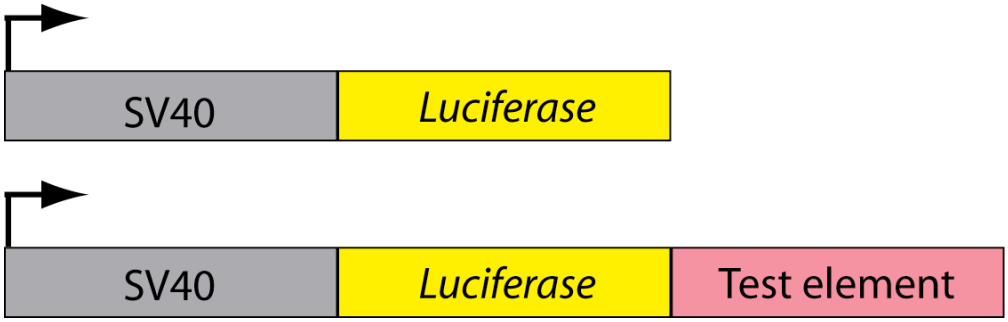


Figure 3.10 Schematics of *Luciferase* reporter-gene constructs used in transient transfection assays to test for *in vitro* enhancer activity of selected elements. The top schematic represents the pGL3-Promoter enhancer-less control (Promega), where *luciferase* reporter gene transcription is under the control of the SV40 minimal promoter. The bottom schematic represents a pGL3-promoter-test element construct. Test elements were cloned downstream of the *Luciferase* reporter gene in the pGL3-Promoter control vector. The ability of a test element to mediate *luciferase* activity was assessed by comparing the output of test constructs to control constructs (after controlling for transfection efficiency).

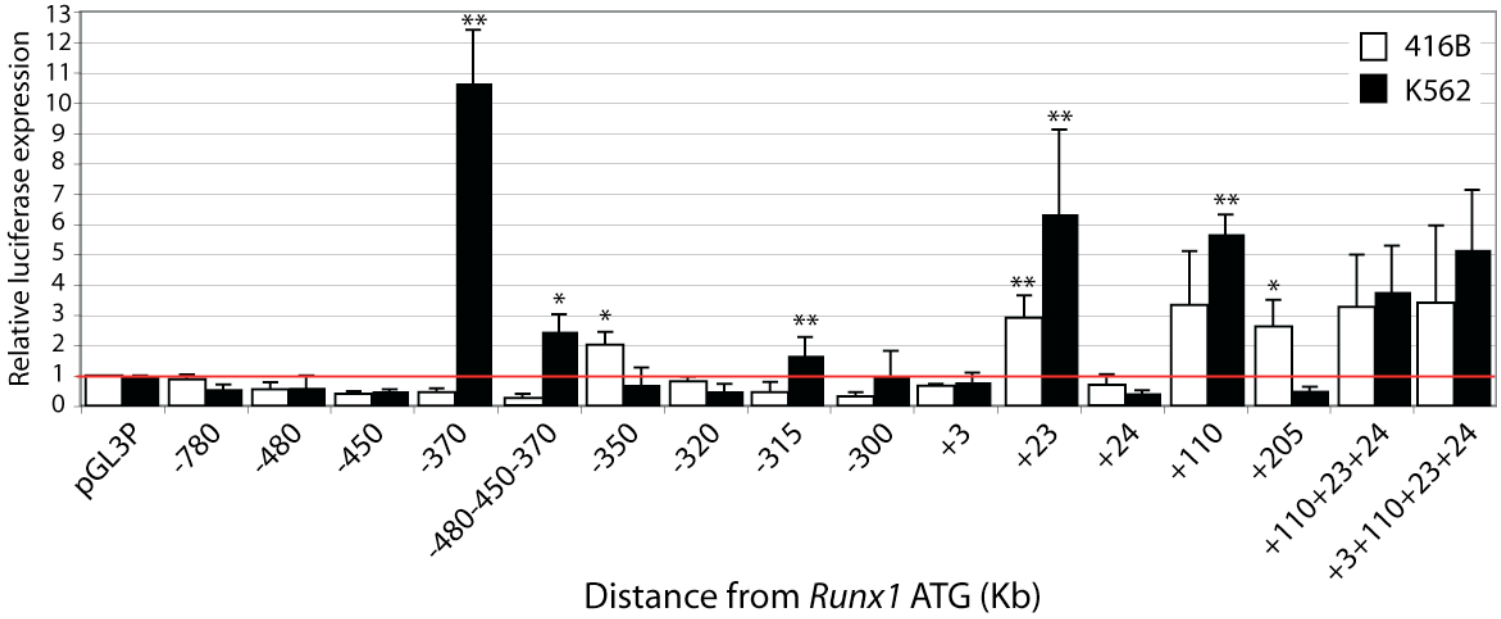


Figure 3.11. *In vitro* enhancer activity of putative and known *Runx1 cis*-regulatory elements. The x-axis represents the position of elements in Kb from the *Runx1* ATG that were cloned into the pGL3-Promoter vector. The y-axis represents the mean fold luciferase activity (corrected for transfection efficiency in comparison the *Renilla* transfection control) from at least 3 experiments, normalised to the pGL3P enhancer-less control vector (+ SD). The red horizontal line represents the fold activity of the pGL3P control vector (1) to which everything else is compared. White bars = results obtained using the mouse 416B myeloid progenitor cell line. Black bars = human K562 erythroid cell line results. CNEs with statistically significant enhancer activity over the pGL3P enhancer-less control are indicated: '**' = $p<0.01$, '*', $p<0.05$ (2-tailed, Student t-Test).

elements have enhancer activity in these cell types. The +23 is active in both cell types and this also looks to be the case for the +110 element, although its activity in 416B cells was just below statistical significance, due to large standard deviations. Further repeats would likely remedy this however. Interestingly, the +110+23+24 and +3+110+23+24 composite constructs do not appear to more or less active than the +23 enhancer alone, suggesting that these elements do not co-operate to mediate greater enhancer activity, at least in these cell types.

3.5 Chapter Discussion

In this chapter a combination of techniques (ChIP-chip and DNaseI-seq) in comparison to published data (ChIP-seq, (Wilson et al., 2010)) have been used to successfully identify five new genomic elements within the *Runx1* syntenic region that have enhancer activity in *in vitro*. Two of these elements, the +110 and +205, are located within the *Runx1* locus, whereas the -370, -350 and the -315 are located in the gene desert.

Interestingly, not all elements that were enriched for H3K4Me2/3, DNaseI peaks and TF binding were found to be enhancers *in vitro*. The +3 element for example, which was enriched for all four, was inactive as an enhancer in both 416B and K562 cells. This is surprising considering the +23 and +110, with similar enrichment patterns, were found to be active as enhancers. Also the -300 element, which was strongly enriched for H3K4Me2 in E12.5 FL cells, was inactive as an enhancer in both cell types tested. In contrast, the -

370 element, which has some H3K4Me2 enrichment in E12.5 FL cells only, with no DNaseI peak in 416B cells but the binding of some TFs in HPC7 cells, proved to have the strongest enhancer activity in K562 cells.

These results demonstrate the importance of testing predictions with a range of likelihoods, as even those that seem the most likely to yield expected results may not. In addition, luciferase assays do not represent *in vivo* enhancer activity of putative enhancers, just their activity in a prescribed cell type. Therefore, as a screen these assays serve an initial 'look see' purpose, but must be tested further. This philosophy is continued in chapter four where the *in vivo* potential of the -480, -370, -350, -480-370-350, +3, +110, and +205 to mediate *Runx1*-specific expression is tested though the generation of transient F0 transgenic enhancer-reporter embryos.

Chapter Four:

In vivo analysis of candidate *Runx1 cis*-regulatory elements

4.1 Chapter Aim

In Chapter three, eleven novel genomic elements within the *Runx1* syntenic region were selected as potential enhancers. These were: the -780 (the 5' end of *SetD4*), -480, -450, -370, -350, -320, -315, -300, +3, +110 and +205. Five of these individual elements and the composite -480-450-370 element were active enhancers in *in vitro* luciferase assays in either mouse 416B myeloid progenitor cells or human K562 erythoid cells. In this chapter, the *in vivo* potential of three of these elements, the -370, +110 and +205 have been tested through the generation of transient transgenic (F0) embryos. The -350 and -315 are yet to be tested. In addition, the composite -480-450-370 is tested, as it is active as an enhancer in K562 cells. Three elements that lacked *in vitro* enhancer activity, the -480, -450 and +3, are also tested *in vivo*. The -480 and -450 are included as they are part of the composite -480-450-370 enhancer. It is therefore of interest to see what role each element plays in the total -480-450-370-mediated expression pattern. The +3 element is tested as it showed characteristics that would strongly suggest putative enhancer activity, which may not have been measurable in the luciferase assays (the elements may be active in different cell types for example). Indeed, luciferase assays do not mimic development. Therefore, seven genomic elements in total are assessed for *in vivo* enhancer activity. These are the -480, -450, -370, -480-450-370, +3, +110, and the +205. The genomic position of these seven elements is illustrated in Figure 4.1.

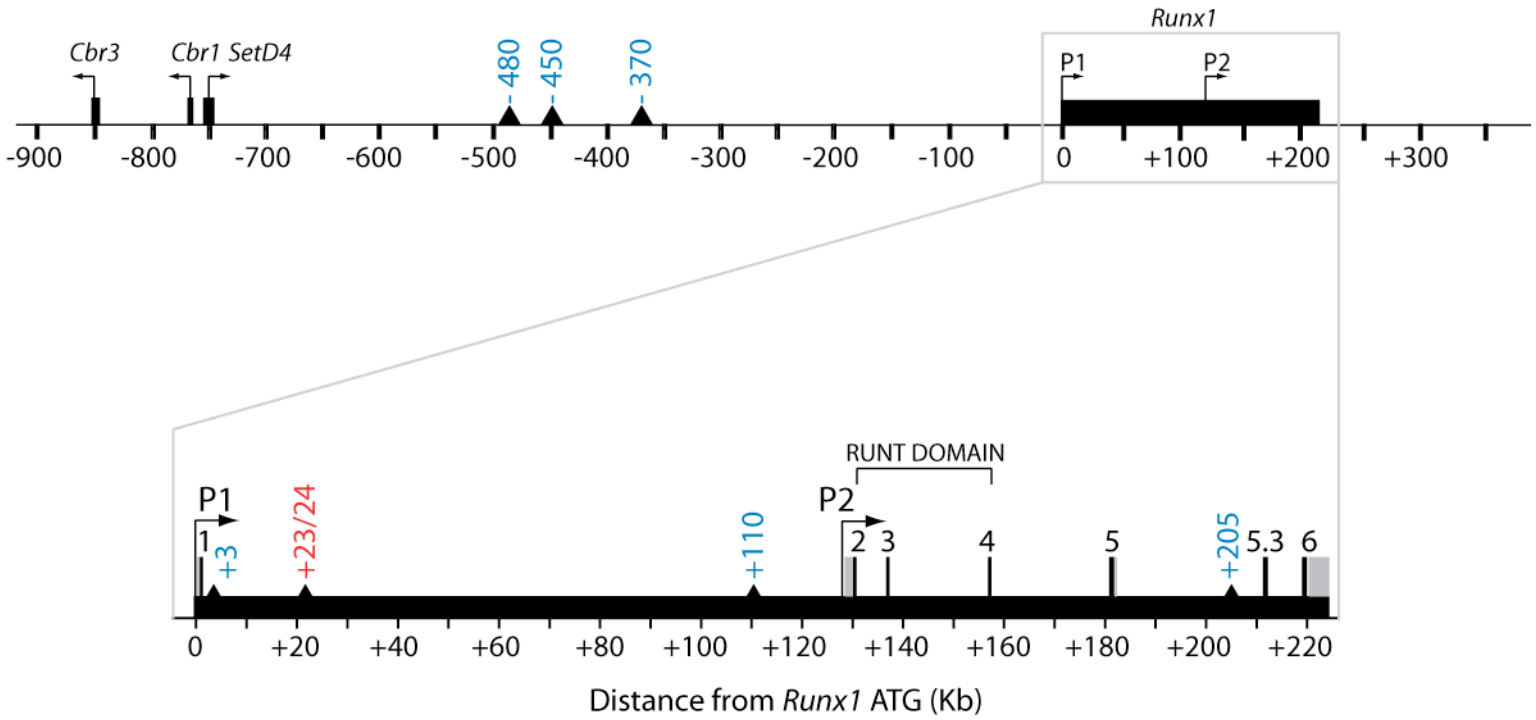


Figure 4.1. Schematic representation of the *Runx1* locus, upstream gene desert, and the *Cbr3*, *Cbr1* and *SetD4* loci. The top image shows the position of the *Runx1* locus in relation to the upstream gene desert that spans from the 3' end of SetD4 to the 5' end of *Runx1*. Genes are represented as black rectangles with arrows representing promoters and the direction of transcription. Highlighted in the grey box is the *Runx1* locus. Represented are the P1 and P2 promoters, exons (numbered black points) and UTRs (grey). The exons coding for the DNA-binding Runt domain are indicated. Black triangles and blue numbers represent the location of genomic elements being tested for *in vivo* enhancer activity. The +23/24 enhancer is indicated in red.

4.2. Generation of transient transgenic (F0) embryos

To test for *in vivo* enhancer activity, transient transgenic (F0) embryos were created using reporter gene constructs. These constructs included the hsp68*LacZ* and hsp68*GFP* promoter-reporter genes. Elements were cloned downstream of the *LacZ* or *GFP* genes under the control of the minimal, heterologous hsp68 promoter in these constructs (Figure 4.2 A and B). Subsequently, transgenic embryos were generated by pro-nuclear injection of the constructs into 1-cell stage mouse embryos and harvested at three major time points: E8.5, E10.5/E11.5 and E14.5. Transgenic embryos were analyzed for reporter gene expression at the sites of hematopoiesis by X-gal staining (in the case of hsp68*LacZ* constructs,) or by expression of *GFP*. These methods have been employed successfully in the lab in the past (Bee et al., 2009a; Nottingham et al., 2007). Also documented was any expression mediated to non-hematopoietic sites of *Runx1* expression, as *Runx1* is also expressed in other developing organ systems (Levanon et al., 2001).

Expression mediated by the seven reporter gene constructs was compared to endogenous *Runx1* expression, visualized by *in situ* hybridization for *Runx1* mRNA on sections of WT embryos (see Figure 4.3, done by Chris Vink, an undergraduate student under my supervision). In addition, endogenous *Runx1* expression was visualized using the *Runx1-LacZ* KI mouse model (Figure 4.4 Av-viii and Bv-viii), in which the last two exons of *Runx1* are replaced with the coding region of the *LacZ* gene (North et al., 1999). In this model, *LacZ* expression is seen in transgenic embryos at all sites where full-length *Runx1*

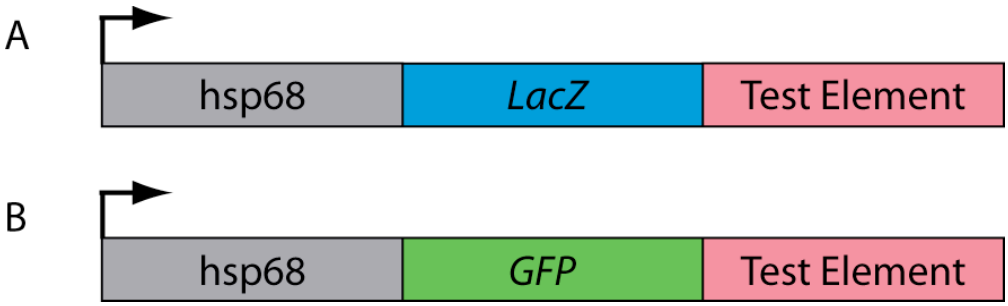


Figure 4.2. Schematic of reporter-gene constructs used to generate F0 transgenic mouse embryos. (A) The -480, -450, -370, -480-370-350 +3, +110 and +205 test elements were cloned downstream of the *LacZ* reporter gene under the transcriptional control of the hsp68 promoter. (B) The +110 test element was also cloned downstream of the *GFP* gene under the transcriptional control of the hsp68 promoter.

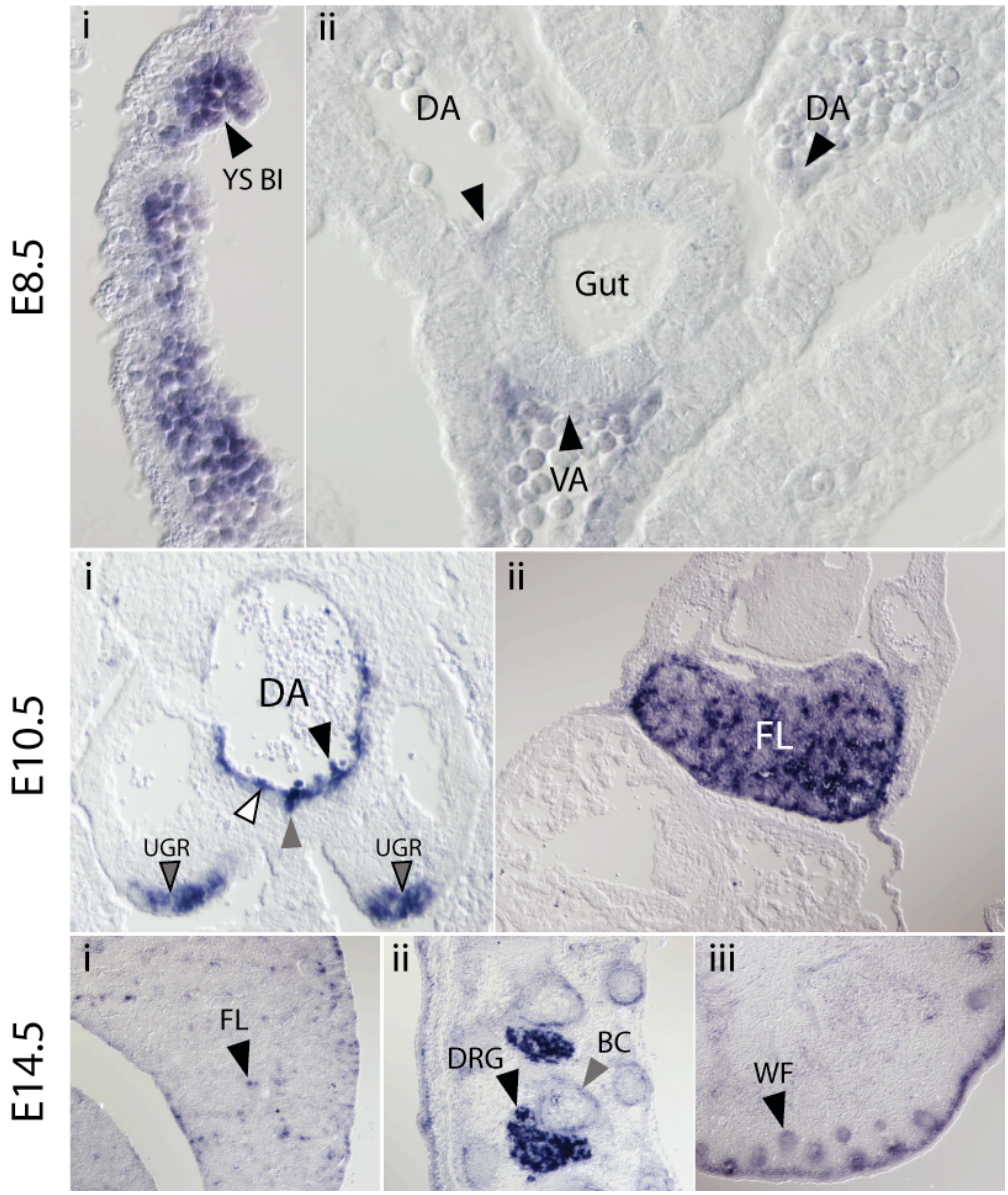


Figure 4.3. Analysis of endogenous *Runx1* expression using *in situ* hybridization (ISH). *In situ* hybridization for *Runx1* full-length mRNA transcripts on sections through E8.5, E10.5 and E14.5 WT embryos. Sections shown at **E8.5** (i) YS (original magnification x 400). Arrow head points to expression within cells of a YS BI. (ii) Posterior section of an ~ 8 sp embryo, (original magnification x 400). Arrowheads point to endothelial expression in the paired DA and the VA (dorsal is up). At **E10.5**, expression is seen in (i) the DA, (original magnification x 100). Black arrowhead = expression in an emerging cluster of hematopoietic cells (which include HSCs). White arrowhead = expression in cells of the endothelium, Grey arrowhead = expression in the mesenchyme underlying the endothelium (dorsal is up), Grey arrowheads with black borders = expression in the UGRs (ii) *Runx1* expression in the E10.5 FL (original magnification x 100). At **E14.5** (all original magnification x 400) (i) black arrow

head points to expression in cells of the FL. (ii) Black arrow head = expression in the dorsal root ganglia, grey arrow head = expression in the immature cartilage of the rib. (iii) Black arrow head = expression in the whisker follicle of the snout. ISH done by Chris Vink, undergraduate supervisee. Abbreviations: YS BI, yolk sack blood islands; DA, dorsal aorta; UGR, urogenital ridges; FL, fetal liver; DRG, dorsal root ganglia; BC, bone cartilage; WF, whisker follicle.

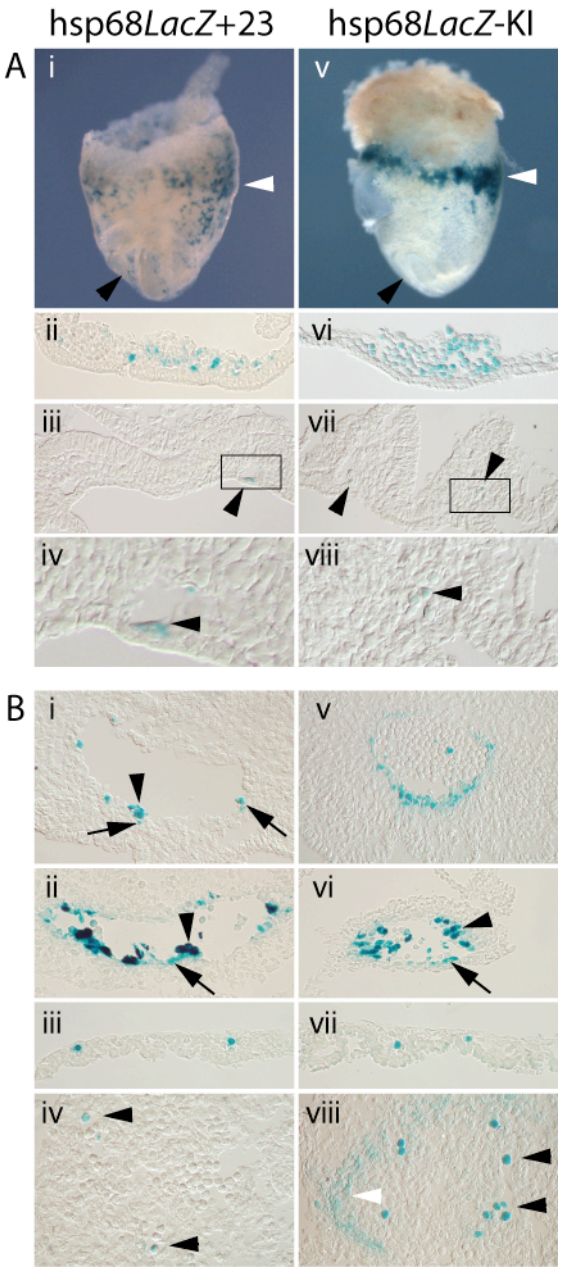


Figure 4.4. Comparison of +23-mediated *LacZ* expression to *Runx1-LacZ*KI expression at hematopoietic tissues of the developing mouse embryo. (A) *LacZ* expression in early E8 (5-6 sp) *hsp68LacZ+23* F0 (i-iv) and *Runx1-LacZ*KI (v-viii) embryos. (i,v) *LacZ* expression in whole mounts. White arrowheads point at the YS BI: black arrowheads point at the anterior paired DA, outside of the hematopoietic PAS region. (ii,vi) *LacZ* expression in the YS BI. (iii,vii) Expression in the posterior paired DA in the PAS region. Arrowheads indicate expression in the wall of the DA. (iv and viii) Enlargement of boxed areas in panes iii and vii, respectively. Images in panels i and v were made using a 1.6x objective lens; panels ii-iv and vi-viii: 20x Normarski objective. (B) *LacZ* expression in E10.5 *hsp68LacZ+23* F0 transgenic (i-iv) and *Runx1-LacZ*KI (v-viii) embryos. (i,v) Expression in the DA in the AGM region (dorsal = up). Arrowhead in panel Bi indicates +23 enhancer activity in a cluster of

hematopoietic cells. *LacZ* expression is also seen in a few individual cells in the endothelial layer and the mesenchyme (arrows). *LacZ* expression in the *Runx1-LacZ*KI (v) also includes a wider population of endothelial and mesenchymal cells. (ii,vi), Expression in hematopoietic cells clusters of the umbilical artery (arrowheads) and cells in the endothleial wall (arrows). (iii, vii) Expression in YS blood cells. (iv,viii) *LacZ* expression in FL. Black arrowheads indicate expression in individual hematopoietic cells. The white arrowhead points at the FL capsule. Original magnification, x200. Figure from (Nottingham et al, 2007).

transcripts (i.e those containing both exons 5.3 and 6, see (Bee et al., 2009b)) are transcribed. Expression patterns mediated by test elements were also compared to those of the +23 hematopoietic-specific enhancer (Bee et al., 2009a; Bee et al.; Nottingham et al., 2007) (Figure 4.4 Ai-iv and Bi-iv) to establish any overlap in activities.

Results for F0 TG experiments are shown in sections 4.3-4.8 starting with elements within the *Runx1* locus (+3, +110 and +205) and followed by the gene desert elements (the composite -480-370-350 and individual -480, -370 and -350 elements). When looking at these results, please compare to the expression patterns for endogenous *Runx1* expression and +23 enhancer-mediated expression patterns shown in figures 4.3 and 4.4 respectively. The numbers of expressing transgenics and the tissues in which reporter-gene expression is observed in these embryos is summarized in Table 4.1.

4.3. Analysis of reporter gene expression in hsp68*LacZ*+3 F0 transgenic embryos

Analysis of *LacZ* expression patterns in hsp68*LacZ*+3 F0 transgenic embryos showed that the +3 element is an enhancer *in vivo*, despite not being so in *in vitro* luciferase enhancer assays, described in chapter three, section 3.4. The +3 enhancer mediates *Runx1*-specific *LacZ* expression to sites of hematopoiesis at all three time points analyzed (E8.5, E10.5 and E14.5). It also mediated *LacZ* expression to bone cartilage and whisker follicles in

E14.5 transgenic F0 embryos. The latter are both sites of non-hematopoietic *Runx1* expression.

At E8.5, 3/6 hsp68*LacZ*+3 F0 transgenic embryos were positive for X-gal staining (i.e. expressing transgenics). The +3 enhancer mediated *LacZ* expression to the blood islands (BI) of the extra-embryonic yolk sac (YS) in 3/3 of these embryos in a *Runx1*-specific pattern (Figure 4.5 Ai and Aii). In addition, faint expression was seen in the paired dorsal aortae (DA) and angiogenic cords (AC) (which precede the formation of aortae) in 1/3 of these embryos (Figure 4.5 Aiii). At E10.5, 4/9 hsp68*LacZ*+3 F0 transgenic embryos were positive for X-gal staining. *Runx1*-specific hematopoietic expression was seen in intra-aortic clusters and distinct endothelial cells of the DA in 3/4 of these embryos (Figure 4.5 B), in a pattern similar to that seen in hsp68*LacZ*+23 E10.5 embryos (Figure 4.4 Bi). At E14.5 *Runx1*-specific hematopoietic expression was seen in 5/5 of the hsp68*LacZ*+3 F0 transgenic embryos analyzed, in large cells within the FL (Figure 4.5 Ci). In addition, *Runx1*-specific non-hematopoietic expression was seen in 3/5 of these hsp68*LacZ*+3 F0 transgenic embryos, in the cartilage of developing ribs (Figure 4.5 Cii) and in the whisker follicles (Figure 4.5 Ciii).

4.4. Analysis of +110 activity in hsp68*LacZ*+110 and hsp68*GFP*+110 F0 transgenic embryos

Analysis of *LacZ* expression patterns in hsp68*LacZ*+110 F0 transgenic embryos showed that the +110 element is an enhancer *in vivo*, but its activity

A.

		Positive / Total E7.5/8.5 β-gal expressing embryos		
Construct	βgal+/Tg	YS	Other Runx1 Specific	Other ectopic
hsp68LacZ-480	ND			
hsp68LacZ-450	ND			
hsp68LacZ-370	ND			
hsp68LacZ-480-450-370	ND			
hsp68LacZ+3	3/6	3/3	1/3	0/3
hsp68LacZ+23	5/5	4/5	5/5	5/5
hsp68LacZ+23+24	9/10	6/9	6/9	9/9
hsp68LacZ+110	5/7	5/5	0/5	0/5
hsp68LacZ+205	8/21	4/8	1/8	1/5

B.

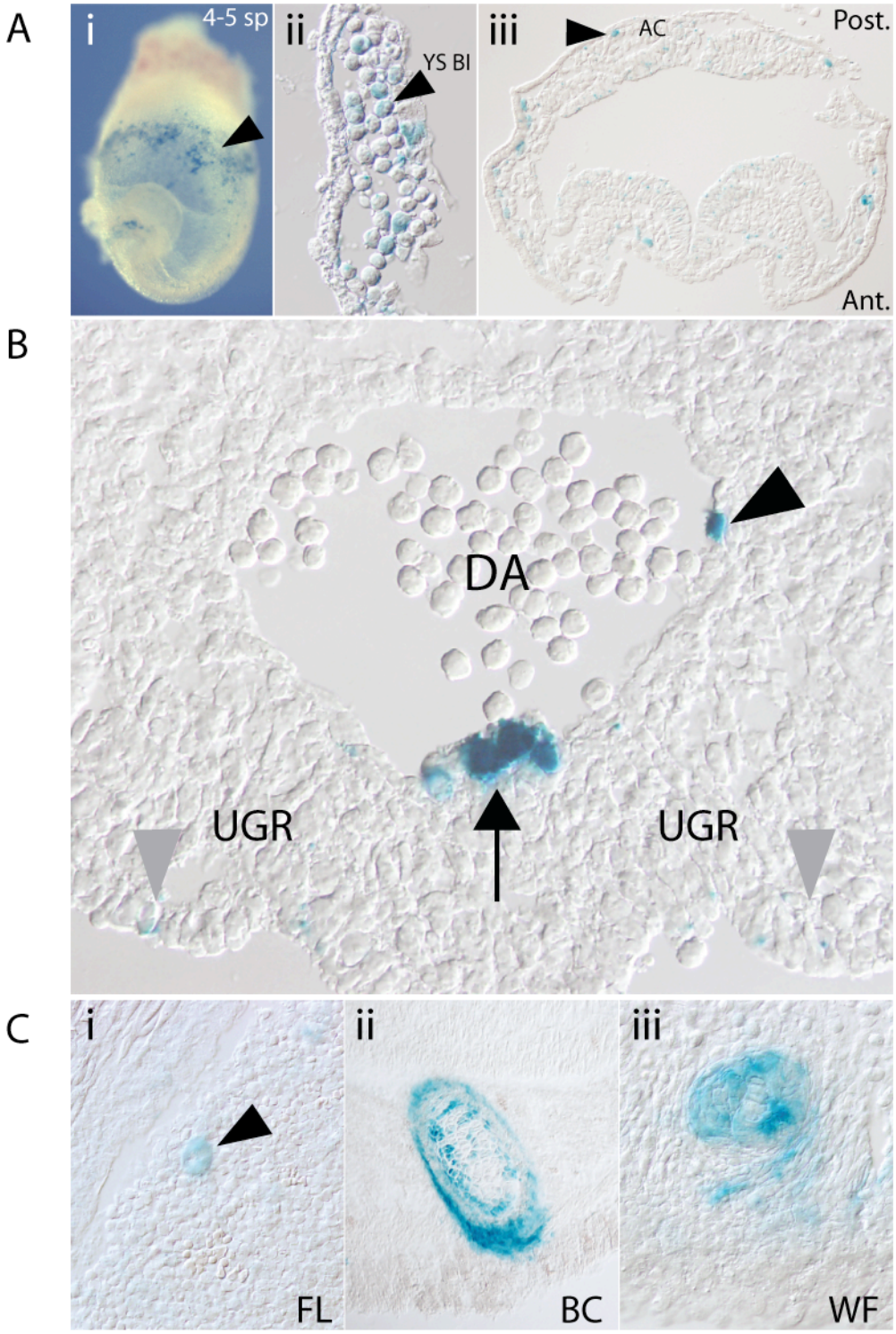
		Positive / Total E10.5/11.5 β-gal expressing embryos						
Construct	βgal+/Tg	AGM endo. + clusters	VU	PB	YS	FL	Other Runx1 Specific	Other ectopic
hsp68LacZ-480	9/12	3/5	0/5	1/5	2/5	0/5	1/5	5/5
hsp68LacZ-450	5/6	1/5	1/5	2/5	4/5	0/5	3/5	3/5
hsp68LacZ-370	5/9	0/4	0/4	0/4	2/4	0/4	0/4	2/4
hsp68LacZ-480-450-370	4/8	2/4	2/4	1/4	1/4	1/4	3/4	1/4
hsp68LacZ+3	4/9	3/4	0/4	1/4	3/4	0/4	0/4	1/4
hsp68LacZ+23	11/14	5/11	4/11	3/11	9/11	4/11	0/11	6/11
hsp68LacZ+23+24	6/8	5/6	5/6	4/6	5/6	4/6	0/6	6/6
hsp68LacZ+110	4/11	1/4	0/4	1/4	0/4	1/4	1/4	4/4
hsp68LacZ+205	3/6	1/3	0/3	0/3	0/3	0/3	0/3	3/3

C.

		Positive / Total E14.5 β-gal expressing embryos		
Construct	βgal+/Tg	FL	Other Runx1 Specific	Other ectopic
hsp68LacZ-480	ND			
hsp68LacZ-450	ND			
hsp68LacZ-370	ND			
hsp68LacZ-480-450-370	ND			
hsp68LacZ+3	5/5	5/5	3/5	2/5
hsp68LacZ+23	9/12	6/9	3/9	7/9
hsp68LacZ+23+24	2/2	2/2	2/2	2/2
hsp68LacZ+110	2/5	2/2	2/2	0/2
hsp68LacZ+205	4/6	0/4	4/4	0/4

Table 4.1. For legend, see overleaf.

Table 4.1. Summary of expression seen in sections of transient transgenic (F0) hsp68*LacZ*-enhancer embryos at (A) E7.5/8.5, (B) E10.5 or *E11.5 (C) E14.5. Shown in each table are the different types of hsp68*LacZ*-enhancer F0s, the number of expressing F0 transgenics (βgal+/Tg) and the hematopoietic tissues to which the enhancers mediate expression of *LacZ*. In (A) 'other *Runx1* specific' refers to expression seen in the allantois. In (B) 'other *Runx1* specific' refers to expression seen in the limb buds and in (C) 'other *Runx1* specific' refers to expression seen in the bone cartilage, DRG or whisker follicles. ND = not done. *hsp68*LacZ*-480-450-370 and hsp68*LacZ*-480, hsp68LacZ-450 and hsp68LacZ-370 F0 transgenics were analysed at E11.5.



For figure legend, please see overleaf.

Figure 4.5 *LacZ* expression mediated by the +3 CNE in E8.5, E10.5 and E14.5 hsp68*LacZ*+3 F0 embryos. (Ai) *LacZ* expression seen in the YS of a WM 4-5 sp expressing F0. (Aii) Transverse section through the YS of the embryo in Ai (original magnification x 200). Black arrow heads point to a LacZ+ cell within the YS BI (Aiii) Transverse section through the PAS of the embryo in Ai (original magnification x 100) showing *LacZ* expression in the angiogenic cords (black arrow head) of the posterior region of the embryo, as well as expression in the anterior aortae. (B) *LacZ* expression in the DA of an E10.5 expressing F0 transgenic (transverse section, original magnification x 200). The black arrow points to *LacZ* expression in an emerging cluster of hematopoietic cells in the ventral wall of the the DA. The black arrow head points to a LacZ+ endothelial cell in one of the dorso-lateral walls of the DA. In this embryo, punctate expression was also seen in cells at the tips of the urogenital ridges (UGR), indicated by grey arrowheads. (Ci) *LacZ* expression in a large cell of the FL, indicated by black arrowhead (original magnification x 400). (Cii) *LacZ* expression is seen in the bone cartilage of the rib (original magnification x 400). (Ciii) Black arrowhead indicates *LacZ* expression in a whisker follicle of the snout (original magnification x 400).

appears to vary depending on the developmental stage analyzed. At E8.5, 5/7 hsp68*LacZ*+110 F0 transgenic embryos were positive for X-gal staining. Strong X-gal was seen in a hematopoietic-specific *Runx1* pattern in the YS of all five of the E8.5 expressing F0 transgenics. This was particularly evident in earlier stage embryos (e.g. 1-2 sp), the time at which *Runx1* expression is strongest in the YS BIs (Figure 4.6 Ai and ii). This expression pattern is similar to that seen in *Runx1* ISH experiments at the same time point (Figure 4.3 E.8.5i) However, no expression could be seen in the paired DA of these embryos suggesting that the +110 enhancer may be inactive in terms of driving *Runx1* expression at this site in the embryo proper (Figure 4.6 B).

To further examine the +110 enhancer activity in the YS at E8.5, hsp68*GFP*+110 F0 transgenic embryos were generated, to analyze the cells of the YS by FACS. GFP allows for direct visualisation of the cells in which the +110 is active. F0 transgenic embryos expressing *GFP* were identified by fluorescence microscopy and single cell suspensions were made. Cells were then incubated with combinations of Ter119, CD41 or VE-Cadherin antibodies to distinguish endothelial cells (VE-CAD^{+}CD41^{-}) and EryPs (Ter119^{+}CD41^{lo} or VE-CAD^{-}CD41^{lo}) from definitive hematopoietic cells (CD41^{hi}). An example of a 6 sp hsp68*GFP*+110 F0 TG embryo with *GFP* expression in the YS is shown in Figure 4.7. Analysis of antibody binding using the CYAN flow cytometer revealed that the majority of Ter119^{+}CD41^{lo} EryPs were in the GFP^{+} cell population (93%), indicating that the +110 mediates *GFP* expression to the majority of EryPs in this F0 embryo (Figure 4.8 A). In addition, almost half of the Ter119^{-}CD41^{hi} definitive hematopoietic

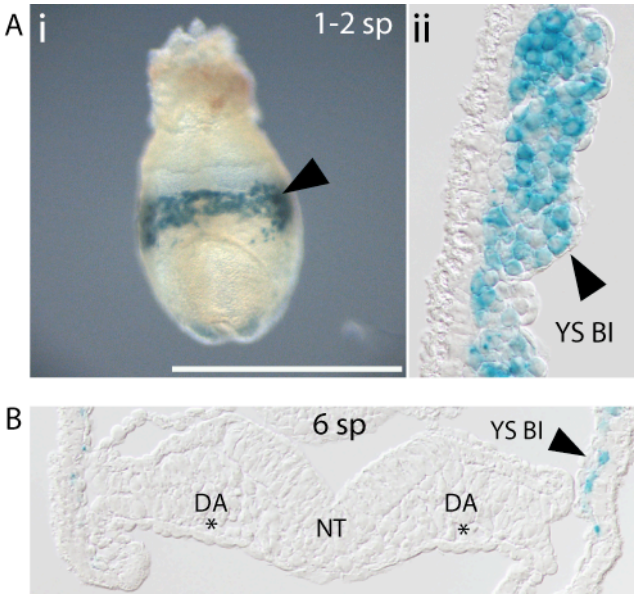


Figure 4.6 *LacZ* expression mediated by the +110 element in representative E8-8.5 hsp68*LacZ*+110 expressing F0 transgenic embryos. (A) *LacZ* expression in the YS BIs of a 1-2 sp F0 seen by (i) whole mount (white scale bar = 1000 μm) and (ii) transverse section though the YS (original magnification x 400). (B) Transverse section through the PAS region of a 6 sp hsp68*LacZ*+110 F0 transgenic embryo. Asterisks are positioned within the lumens of the paired DA where X-gal staining is seen. Black arrowheads point to the YS BIs in both A (i) and (ii) and B.

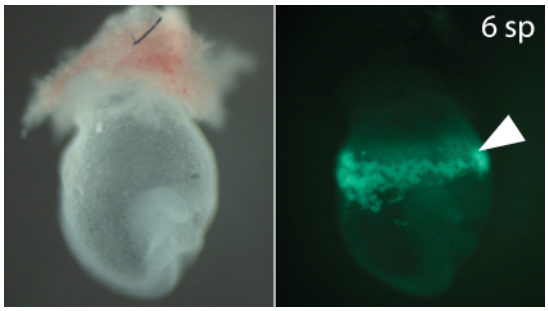


Figure 4.7. *GFP* expression in an E8.5 (~ 6sp) hsp68*GFP*+110 F0 transgenic embryo used for FACS analysis of cells in the YS. The image on the left is illuminated with bright light and the image on right is illuminated with fluorescent light. The +110 clearly mediated *GFP* expression to a band in the YS where the blood islands are located, indicated by white arrowhead.

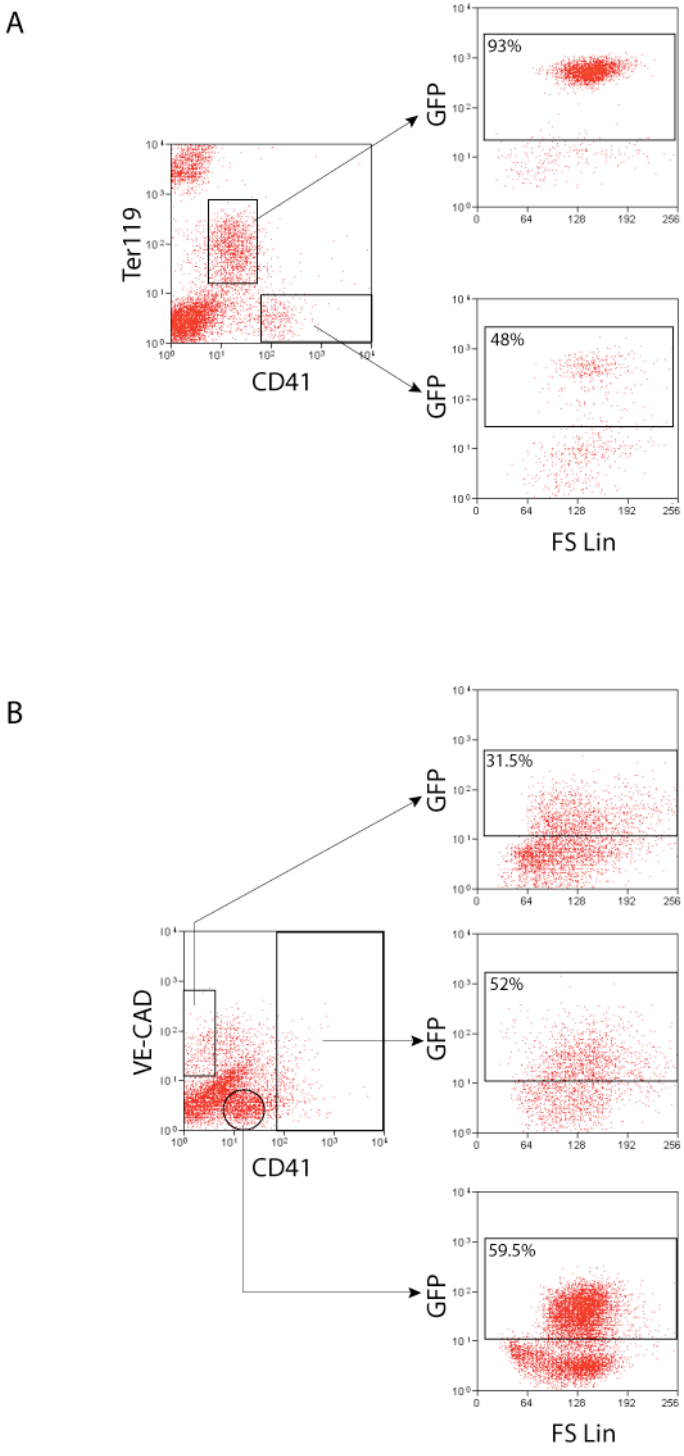


Figure 4.8. For figure legend, see overleaf.

Figure 4.8. The +110 enhancer mediates *GFP* expression to endothelial cells and both primitive and definitive hematopoietic cells in the YS of E8 *hsp68GFP*+110 F0 transgenic embryos. (A) Dot plots of +110 enhancer activity in EryPs ($Ter119^{+}CD41^{lo}GFP^{+}$) and hematopoietic cells ($Ter119^{-}CD41^{hi}GFP^{+}$) in an E8 F0 transgenic embryo. (B) Dot plots of +110 enhancer activity in endothelial cells ($VE\text{-}CAD^{+}CD41^{-}GFP^{+}$), EryPs ($VE\text{-}CAD^{-}CD41^{lo}GFP^{+}$) and hematopoietic cells ($CD41^{hi}GFP^{+}$) in two pooled E8 F0 transgenic embryos.

cells from the embryo (48%) were also in the GFP^+ cell population. This indicates that the +110 enhancer mediates *GFP* expression to the majority of EryPs and to a significant proportion of hematopoietic cells of the transient definitive wave in the YS of this F0.

To further characterize the +110 enhancer at E8.5, a subsequent experiment was performed to assess if the +110 was active in endothelial cells of the YS, from where the YS definitive hematopoietic cells are thought to emerge (reviewed in McGrath (McGrath and Palis, 2005). Two expressing E8.5 *hsp68GFP+110* F0 TG embryos were pooled and incubated with antibodies against VE-Cadherin (endothelial marker) and CD41 (hematopoietic marker). These embryos showed more punctate *GFP* expression than the embryo described above (data not shown). Analysis of the cell populations showed that 31.5% of the VE-$Cadherin^+CD41^-$ endothelial cells were GFP^+ indicating that the +110 mediates expression to the YS endothelium in these F0s (Figure 4.8 B). In the VE-$Cadherin^-CD41^{lo}$ EryP population, the +110 mediated *GFP* expression to 59.5% of the cells. 52% of the $CD41^{hi}$ hematopoietic cell population were also GFP^+ (Figure 4.8 B).

At E10.5, 4/11 F0 transgenics were positive for X-gal staining. However, of these four expressing F0 transgenic embryos, only one showed X-gal staining in the DA and the FL (Figure 4.9 Ai-ii). In the same embryo, ectopic expression was seen in the intersomitic vessels and throughout the CNS. The latter is frequently seen in F0 TG embryos generated using the *hsp68LacZ* vector alone and hence not *Runx1*-specific. The three remaining embryos

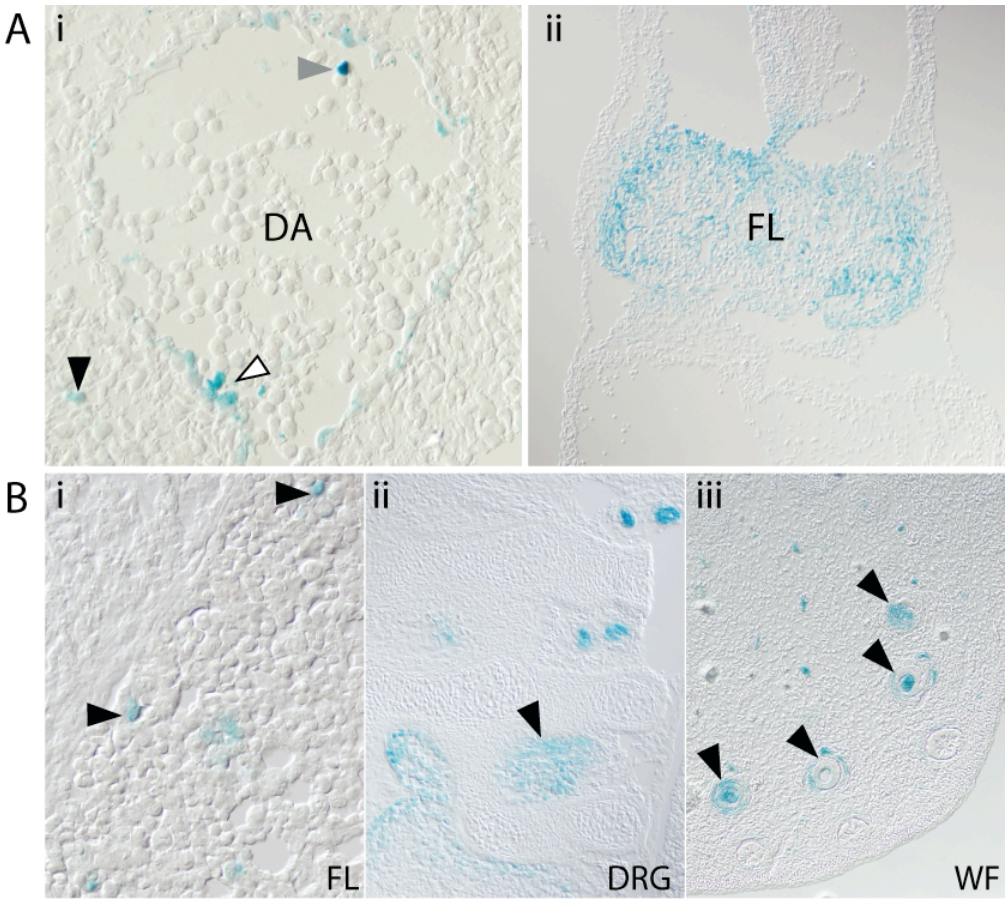


Figure 4.9. *LacZ* expression mediated by the +110 element in representative E10.5 and E14.5 hsp68*LacZ*+110 transgenic F0 embryos. (Ai) Transverse section through the DA of a representative E10.5 transgenic embryo (original magnification x 400). The white arrowhead (with black border) indicates a cluster of cells emerging from the hemogenic endothelium. The grey arrowhead points to a circulating cell in the lumen of the DA and the black arrowhead indicates a mesenchymal cell lateral to the ventral wall of the DA (dorsal = up). (Aii) Transverse section through the FL of a representative E10.5 transgenic embryo (original magnification x 100) where widespread X-gal staining is seen. (Bi) Longitudinal section through a representative E14.5 hsp68*LacZ*+110 transgenic F0, showing expression in large hematopoietic cells of the FL (black arrowheads, original magnification x 400). *LacZ* expression is also seen in the DRG (black arrow head, original magnification x100). (Biii) shows a longitudinal section through the snout of a representative E14.5 hsp68*LacZ*+110 transgenic F0 embryo. X-gal staining is seen in the whisker follicles (black arrow heads, original magnification x 100). Abbreviations: DA, dorsal aorta; FL, fetal liver; DRG, dorsal root ganglia; WF, whisker follicle.

showed ectopic expression at non-hematopoietic sites only (not shown). At E14.5, 2/5 hsp68*LacZ*+110 F0 transgenics were positive for X-gal staining. In both transgenic embryos, hematopoietic-specific *Runx1* expression was seen in cells of the FL (Figure 4.9 Bi). In addition, X-gal staining was seen at non-hematopoietic sites of *Runx1* expression in the dorsal root ganglia (DRG) (Figure 4.9 Bii) and the whisker follicles (Figure 4.9 Biii) in both expressing F0s.

4.5. Analysis of reporter gene expression in hsp68*LacZ*+205 F0 transgenic embryos

Analysis of *LacZ* expression patterns in hsp68*LacZ*+205 F0 transgenic embryos showed that the +205 element is an enhancer *in vivo*, but with a more limited activity when compared to the other enhancers located within the gene body (+3, +23 and +110).

At E8.5, only 8/21 hsp68*LacZ*+205 F0 transgenic embryos were positive for X-gal staining. Of these expressing embryos, the majority harvested were < 4 sp, including two pre-somitic (i.e. E7.5) embryos. In both of these pre-somitic embryos, X-gal staining was seen in the developing YS as well as in another extra-embryonic tissue, the allantois, which later goes on to fuse with the chorion to form the placenta (Figure 4.10 Ai and Aii). In later stage (e.g. 7-8 sp) F0 transgenic embryos, X-gal staining was seen in the embryo proper in a streak extending from the posterior of the embryo to the head (Figure 4.10 Bi). Transverse sections though the YS (Figure 4.10 Bii) show hematopoietic

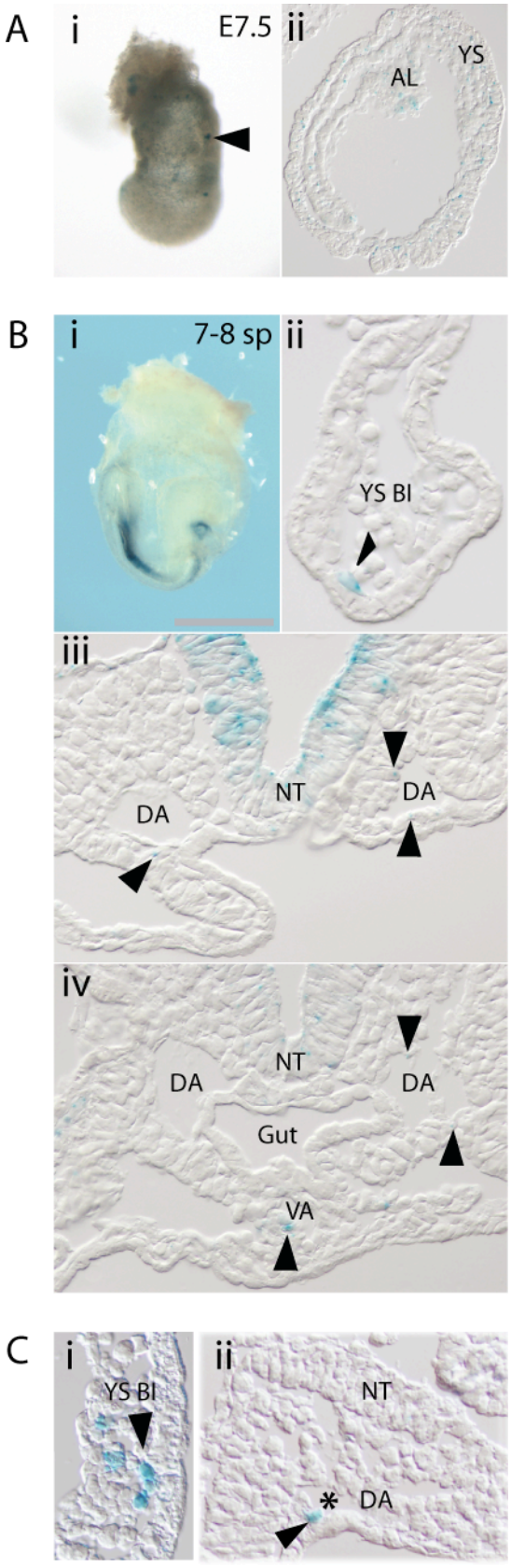


Figure 4.10. For legend, please see overleaf.

Figure 4.10 *LacZ* expression mediated by the +205 element in representative E7.5 and E8.5 hsp68*LacZ*+205 F0 transgenic embryos and the +194-205 element in E8.5 hsp68*LacZ*+194-205 F0 transgenic embryos. (Ai) *LacZ* expression in the YS BI of an E7.5 hsp68*LacZ*+205 transgenic F0. The scale is not shown, but embryos at this stage are approximately 400 μm in diameter. Black arrowhead points to a focus of *LacZ* expression in the YS, although punctate expression is seen throughout the YS. (Aii) Transverse section through the embryo in Ai (original magnification x 100). *LacZ* expression is seen in both the AL and in the YS BI. (Bi) *LacZ* expression in a representative 7-8 sp whole mount hsp68*LacZ*+205 F0 transgenic embryo (Scale bar = 1000 μm). (Bii) Transverse section though a YS BI of the embryo in Bi (original magnification x 400). Black arrowhead points to single X-gal+ cell, which appears to be part of the YS endothelium. (Biii) Transverse section through the PAS region of the embryo proper (original magnification x 400). Black arrowheads point to punctate X-gal staining in the dorsal and ventral aspects of the endothelium of the paired DA (dorsal = up). (Biv) Transverse section through a more posterior section of the same embryo (original magnification x 400). Expression is seen as for B iii (black arrow heads), but is seen now in the VA also. Expression seen in the neural tube is ectopic. (Ci) Transverse section through the YS of an E8.5 hsp68*LacZ*+194-205 transgenic F0 embryo (original magnification x 400). X-gal staining is seen in cells of the YS BI (black arrow head). (Cii) Transverse section through the PAS of a representative hsp68*LacZ*+194-205 transgenic F0 embryo (original magnification x 400). Expression in the endothelium of one of the paired DA of this embryo is indicated by the black arrowhead (lumen indicated by asterisk). Abbreviations: AL, allantois; YS BI, yolk sac blood island; NT, neural tube; DA, dorsal aorta; VA, vitelline artery.

Runx1-specific *LacZ* expression restricted to what appear to be endothelial cells of the YS and not to the EryPs (based on morphology alone). In the embryo proper of this F0, punctate X-gal staining was seen in endothelial cells of the paired DA in both ventral and dorsal positions (Figure 4.10 Biii-iv) as well as in the wall of the vitteline artery (VA) (Figure 4.10 Biv). Shown in Figure 4.10 Ci-ii is the YS and DA of an E8.5 hsp68*LacZ*+194-205 expressing F0 for comparison (Wade Nottingham, unpublished). This construct contains an element that spans 11 Kb from the +194 position in intron 5 up to and including the +205 position. Slightly stronger *LacZ* expression is seen in both tissues in comparison to those shown for the hsp68*LacZ*+205 transgenic F0 embryos but expression remains restricted nonetheless.

At E10.5, 3/6 hsp68*LacZ*+205 F0 transgenic embryos were positive for X-gal staining. Of these three expressing transgenic F0 embryos, *LacZ* expression was seen in the AGM of one embryo only. This was in a mesenchymal cell in the AGM region above the dorsal wall of the DA (Figure 4.11 ii). Such expression is similar to expression seen at the same time point in hsp68*LacZ*+194-205 F0s (Compare Figure 4.11 Ai with Aii). The nature of these cells is not clear. It seems unlikely however that this expression is *Runx1* hematopoietic-specific, as similar X-gal positive cells are not present at this location in the mesenchyme in E10.5 *Runx1-LacZ* KI embryos (see Figure 4.4 Bv).

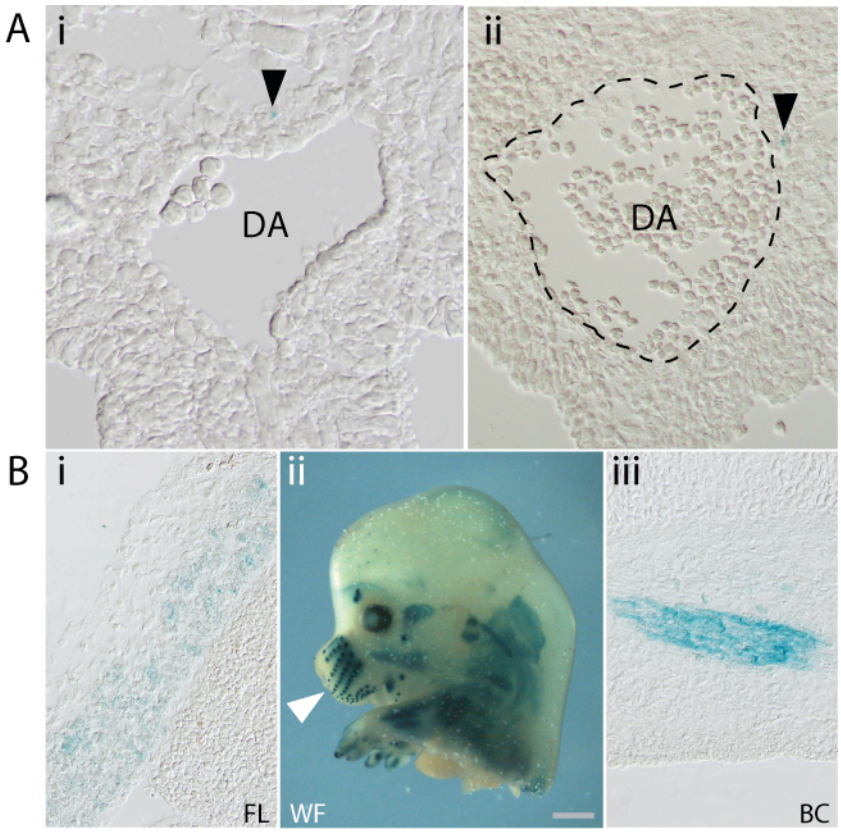


Figure 4.11. *LacZ* expression in representative E10.5 and E14.5 hsp68*LacZ*+205 F0 transgenic embryos and in the DA of a representative hsp68*LacZ*+194-205 F0 transgenic embryo. (Ai) Transverse section through the DA of an E10.5 hsp68*LacZ*+205 F0 transgenic embryo. Black arrowhead points to a single LacZ+ cell in the mesenchyme above the dorsal wall of the DA (dorsal = up). (Aii) Transverse section of an E10.5 hsp68*LacZ*+194-205 F0. Black arrowhead points to expression in a single LacZ+ cell in the mesenchyme dorso-lateral to the DA. (Bi) *LacZ* expression is seen in cells surrounding the FL, but not in the hematopoietic cells of the FL itself (original magnification x 200). (Bii) Whole mount expression in the anterior half of a representative hsp68*LacZ*+205 F0 transgenic at E14.5. White arrowhead indicates expression in the whisker follicles of the snout. Expression can also be seen in the CNS and in the bone cartilage of the limb. (Biii) *LacZ* expression in the BC of the limb of the embryo in Bii (original magnification x 200). Abbreviations: DA, dorsal aorta; FL, fetal liver; BC; bone cartilage.

At E14.5, 4/6 hsp68*LacZ*+205 F0 transgenic embryos were positive for X-gal staining. *LacZ* expression was only seen in 1/4 of these expressing embryos in cells on the periphery of the FL, but not in cells within the FL where *Runx1* is expressed (Figure 4.11 Bi). In the other three hsp68*LacZ*+205 F0 transgenic embryos, no X-gal staining was seen in the FL. *Runx1*-specific non-hematopoietic expression was seen however in the whisker follicles of all four hsp68*LacZ*+205 F0 transgenic embryos (Figure 4.11 Bii), in a pattern similar to that seen mediated by the +3 and +110 enhancers. In addition, *Runx1*-specific non-hematopoietic expression was seen in 3/4 of the hsp68*LacZ*+205 F0 transgenic embryos, in the bone cartilage of the fore and hind limbs and in the digits (Figure 4.11 Biii) (Levanon et al., 2001).

4.6 Analysis of reporter gene expression in hsp68*LacZ*-480-450-370 F0 transgenic embryos

The -480, -450 and -370 sequences each contain binding motifs for GATA, Runx and CSL. They were initially tested together at one time point to see if they were worth testing separately. E11.5 F0 TGs were analyzed to see if the elements were active at sites of adult definitive hematopoiesis, before looking for expression in E8.5 or E14.5 F0 transgenics. At E11.5, 4/8 hsp68*LacZ*-480-450-370 F0 transgenic embryos were positive for X-gal staining. In 2/4 of the expressing F0 transgenics, *Runx1*-specific hematopoietic X-gal staining was seen in the DA in cells of the emerging clusters, the endothelium and in mesenchymal cells underneath the ventral wall of the DA. These expression patterns are similar to those seen in the DA of hsp68*LacZ*+23 F0 transgenics

(compare Figure 4.12 Ai with Figure 4.4 Bi). In 2/4 expressing F0 transgenics, expression was also seen in the VA in both endothelial cells of the vessel and cells budding from the endothelium, into the lumen (Figure 4.12 Aii).

4.7 Analysis of reporter gene expression in hsp68*LacZ*-480 F0 transgenic embryos

At E11.5, 9/12 hsp68*LacZ*-480 F0 transgenic embryos were positive for X-gal staining. Of these nine expressing embryos, five have so far been analyzed for expression. In 2/5 of these embryos, faint *Runx1*-specific X-gal staining is seen in cells of the YS (Figure 4.12 Bi). In 3/5 of the hsp68*LacZ*-480 F0 transgenic embryos, X-gal staining is also seen in the mesenchyme surrounding the DA, but not the endothelium or any emerging clusters (Figure 4.12 Bii). This mesenchymal expression pattern appears more distant to the endothelial layer than that seen in the AGM of *Runx1-LacZ*-KI embryos (see Figure 4.4 Bv); and is therefore unclear whether this is *Runx1*-specific.

4.8 Analysis of reporter gene expression in hsp68*LacZ*-450 F0 transgenic embryos

At E11.5, 5/6 hsp68*LacZ*-450 F0 transgenic embryos were positive for X-gal staining. In 5/5 of these embryos, *Runx1*-specific hematopoietic expression

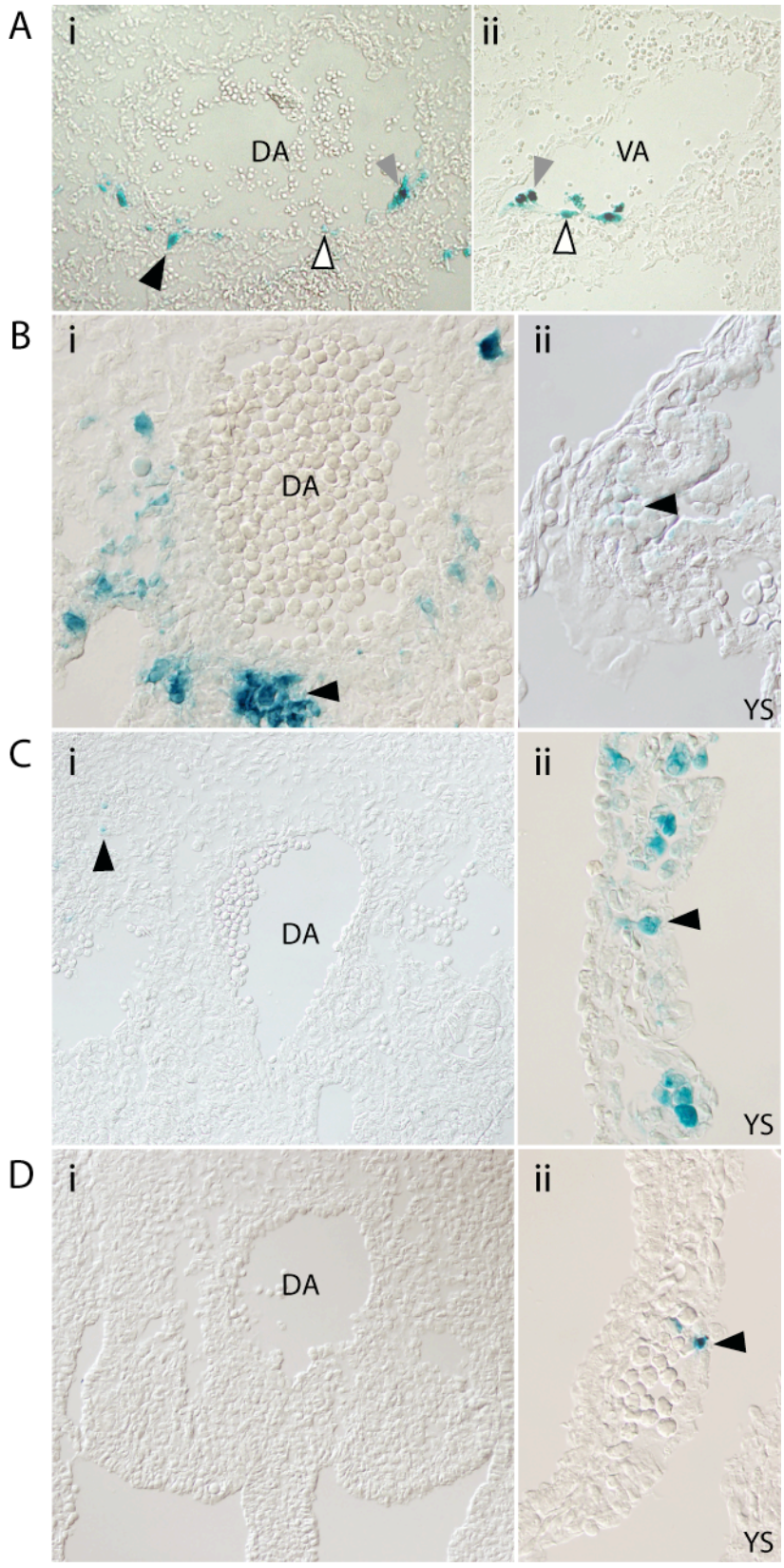


Figure 4.12. For legend, see overleaf.

Figure 4.12 *LacZ* expression in representative hsp68*LacZ*-480-450-370 and hsp68*LacZ*-480, hsp68*LacZ*-450, hsp68*LacZ*-370 transgenic F0 embryos. (Ai) Transverse section through the DA of an E11.5 hsp68*LacZ*-480-450-370 transgenic F0 embryo (original magnification x 400). Black arrowhead indicates expression in the mesenchyme underlying the aortic endothelium. The white arrowhead with black border points to expression in the aortic endothelium lining the DA. The grey arrowhead indicates *LacZ* expression in an emerging cluster of hematopoietic cells from the endothelium. (Aii) *LacZ* expression in the VA of the same hsp68*LacZ*-480-450-370 transgenic F0 embryo (original magnification x 400). The grey arrowhead indicates expression in a cluster emerging from the endothelium of the vessel and the white arrowhead with black border is pointing to a LacZ+ cell within the endothelium. (Bi) Transverse section through the DA of a representative E11.5 hsp68*LacZ*-480 transgenic embryo (dorsal = up). Expression is seen in the mesenchyme surrounding the DA, as indicated by the black arrowhead, which is not in keeping with endogenous *Runx1* expression. (Bii) Transverse section through the YS of a representative E11.5 hsp68*LacZ*-480 transgenic embryo (original magnification x 400). Faint, punctate *LacZ* expression is seen in hematopoietic cells (as indicated by black arrow head) (Ci) Transverse section through the DA of an expressing E11.5 hsp68*LacZ*-450 transgenic F0 embryo (dorsal = up). No *LacZ* expression is visible in the immediate area surrounding the DA, but some faint X-gal staining is seen dorso-laterally (black arrow head). (Cii) Transverse section through the YS of a representative E11.5 hsp68*LacZ*-450 transgenic embryo (original magnification x 400). Strong *LacZ* expression is seen in hematopoietic cells (black arrow head). (Di) Transverse section through the AGM of a representative E11.5 hsp68*LacZ*-370 transgenic F0 embryo (dorsal = up, original magnification x 200). No *LacZ* expression is evident in any cell type in or surrounding the DA. (Dii) Transverse section through the YS of a representative E11.5 hsp68*LacZ*-370 transgenic embryo (original magnification x 400). Limited but strong expression is seen in some hematopoietic cells. Abbreviations: DA, dorsal aorta; VA, vitteline artery; YS, yolk sac.

was seen in cells of the YS (Figure 4.12 Ci). Interestingly, no *Runx1*-specific hematopoietic expression was seen in the DA (Figure 4.12 Cii) or the FL (not shown) of any of the five expressing F0 transgenic embryos.

4.9 Analysis of reporter gene expression in hsp68*LacZ*-370 F0 transgenic embryos

At E11.5, 5/9 hsp68*LacZ*-370 F0 transgenic embryos were positive for X-gal staining. This staining was more restricted than staining seen in the hsp68*LacZ*-480 and hsp68*LacZ*-450 expressing F0 transgenics. Indeed, *Runx1*-specific hematopoietic expression was only seen in 2/5 of the expressing F0 transgenic embryos in the YS (Figure 4.12 Di) and not in the DA (Figure 4.12 Dii). Analysis of whole mount embryos suggested that X-gal staining may have been present in the VA, but this was not observed on sections through this artery.

4.10 Chapter Discussion

In this chapter, seven new genomic elements have been characterized as *Runx1* enhancers *in vivo*. All seven enhancers are able to mediate hematopoietic-specific *Runx1* expression during one or more of the developmental stages analyzed (E8.5, E10.5/11.5 and E14.5). The activity of the enhancers is not the same during developmental hematopoiesis and highlights the complexity of *Runx1 cis*-regulation during this process. A

summary of the activity of the seven enhancers in relation to each other is shown in figure 4.13.

4.10.1 Enhancers within the Runx1 locus: the +3, +23, +110 and +205

As illustrated in Figure 4.13, of the enhancers located within the *Runx1* locus, the +3, +23 and +110 enhancers all are active during the primitive wave of YS hematopoiesis. The +110 enhancer was shown to be highly active in the YS at this time point (based on X-gal staining and high frequency of expressing transgenics) with FACS experiments demonstrating that it mediates *Runx1*-specific hematopoietic *GFP* expression in EryPs, as well as endothelial cells and definitive hematopoietic cells of the YS. Conversely, the +205 enhancer seemed less active and did not show widespread expression in the EryPs. The restricted pattern of X-gal staining mediated by the +205 enhancer indicates that it may be active in the less-abundant, transient definitive hematopoietic cells of the YS. This requires further testing which will be done using an hsp68*GFP*+205 mouse line that is being established in the lab (Christina Rode). This will allow us to further characterize the activity of the +205 enhancer by both FACS experiments and by analysis of *GFP* expression on sections of TG embryos.

The activity of the +3, +23, +110 and +205 enhancers during definitive hematopoiesis also varied. In Figure 4.13, the ability of the enhancers to mediate *LacZ* reporter gene expression to hemogenic endothelial cells (HE), hematopoietic progenitors (HPs) or committed hematopoietic cells

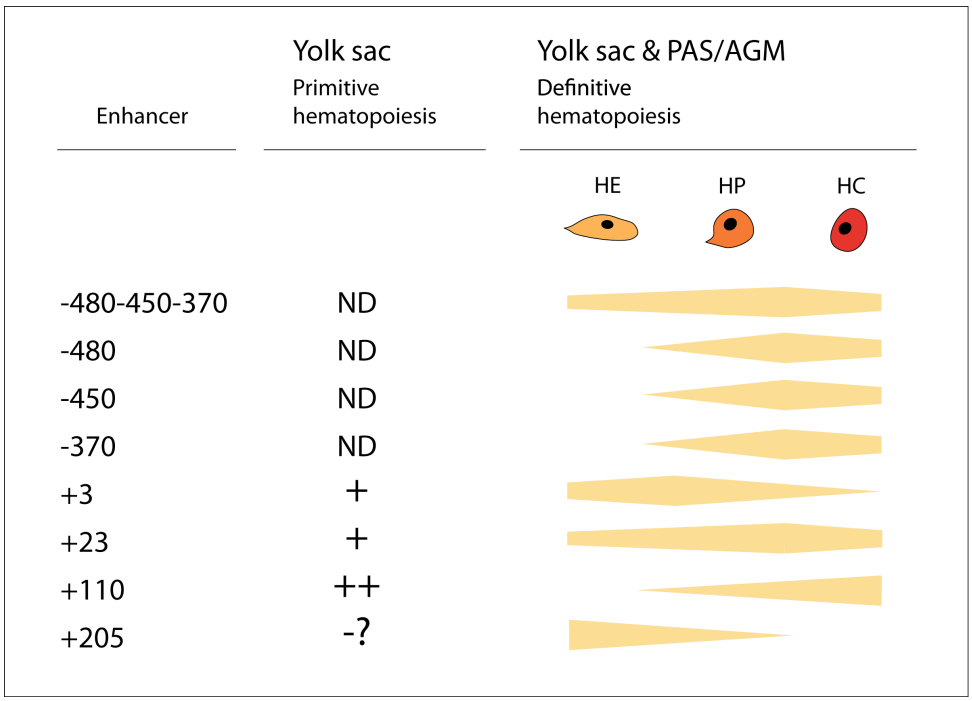


Figure 4.13. Summary of enhancer activities in expressing F0 transgenics during developmental hematopoiesis. Shown in the left-hand column are the enhancer elements tested, numbered based on their position in Kb from the *Runx1* ATG. In the middle column, the activity of enhancers in primitive erythrocytes is summarized. ND = Not Done, '+' = active at this developmental stage based on reporter-gene expression, '++' = highly active, '-?' = seems inactive but requires further testing. In the right-hand column the activity of enhancers during definitive hematopoiesis is summarized. Schematics of hemogenic endothelial (HE), hematopoietic progenitor (HP) and committed hematopoietic cells (HC) are depicted above yellow bars, which indicate in which of these cell types the different enhancers were active.

(HCs) is summarized in schematic form. This summary is based on observed reporter gene expression in these three cells types (based on morphology) in the embryo at known sites of *Runx1* expression. HE in this context refers to expression seen in the hemogenic endothelium of the paired DA at E8.5. HPs and HCs refer to cell types observed later in develoment, for example in the E10.5 DA or the E14.5 FL. As previously published, the +23 enhancer mediates *LacZ* expression to all these cell types ((Nottingham et al., 2007) and Gemma Swiers, unpublished data). F0 transgenic data suggest that the +3 enhancer is similarly active, but its activity decreases as cells take on the HC fate. In contrast, the +110 enhancer appears to be active later in hematopoiesis (expression was not seen in the DA of the E8.5 F0 TGs and the expression in the E10.5 DA is not convincing, for example). Conversely, the +205 is most active early in development in the HE cells of the E8.5 TGs, but becomes inactive in later-stage embryos, demonstrated by the lack of expression seen in cells of the DA at E10.5 or the FL at E14.5

4.10.2 Enhancers in the Runx1 gene desert: the -480-450-370 composite and the -480, -450 and -370 individual enhancers.

As summarized in Figure 4.13, the activity of the -480-450-370 composite and -480, -450 and -370 individual enhancers during definitive hematopoiesis was interesting. When combined, the three enhancers mediated *Runx1*-specific hematopoietic expression to the DA and the VA in a pattern similar to that mediated by the +23 enhancer in these cell-types. In contrast, the -480, -450 and -370 enhancers were unable to mediate *Runx1*-specific expression to

cells in the DA and reproducible expression was not seen in the VAs of the individual enhancer TGs. The -480, -450 and -370 enhancers did mediate *Runx1*-specific *LacZ* expression to cells in the YS, however.

One possible reason for the difference in expression patterns seen in the DA of hsp68*LacZ*-480, hsp68*LacZ*-450 and hsp68*LacZ*-370 F0 TGs vs. hsp68*LacZ*-480-450-370 composite F0 TGs, could relate to the binding of TFs to the enhancers *in vivo*. It is possible that the combined enhancer sequences in hsp68*LacZ*-480-450-370 F0 TGs, was able to integrate a combination of TFs sufficient to mediate *Runx1*-specific hematopoietic expression in the DA, that the -480, -450 and -370 enhancers could not do alone. This remains to be further explored.

4.10.3 Concluding remarks

Seven novel *Runx1* enhancers that mediate *Runx1* expression to sites of hematopoiesis in the developing embryo have been characterized. To try and understand what causes the enhancers to function at different times and in different cell types in the embryo, it will be important to assess the differences between the elements at the sequence level. This may reveal differences in the putative consensus sequences for TF binding between different enhancers, allowing alternative binding of TFs between the enhancers and therefore different methods of *Runx1 cis*-regulation. In addition, investigating which upstream signalling pathways act via these enhancers will be important. Finally, although not a priority for work in our

lab, it would be interesting to understand why the +3, +110 and +205 are able to mediate *Runx1*-specific non-hematopoietic expression to (some or all of) the developing cartilage of the bones, the whisker follicles and also the DRG of the nervous system in E14.5 embryos.

Chapter Five:

Functional analysis of the cells in which the +110 enhancer is active

5.1 Chapter Aim

In the previous chapter, *Runx1*-specific expression patterns mediated by seven novel *Runx1* enhancers were characterized *in vivo*. To establish the function of these enhancers in developmental hematopoiesis, we need to characterize the phenotype of the cells in which the enhancers are active and the hematopoietic potential of these cells. This chapter begins to functionally characterize one of these enhancers, the +110 enhancer, using a novel hsp68*GFP*+110 mouse line (hsp68*GFP*+110 ln5). The +110 is chosen as it has deeply conserved phylogenetic footprints interspersed with less conserved regions. Such factors are thought to be characteristic of enhancers, as was the case for the +23 enhancer (Nottingham et al., 2007). The +110 enhancer-targeted cells will be characterized by: i) Flow cytometry using cells from the AVU, FL and YS of E11.5 TG embryos to characterize the phenotype of cells in which the +110 is active; ii) CFU-C assays to see if the +110 enhancer targets clonogenic progenitors in the E11.5 AVU, FL and YS; iii) transplant assays to assess targeting of HSCs; iv) megakaryocyte differentiation assays to test the activity of the +110 enhancer in a *Runx1*-expressing mature hematopoietic cell type, and finally, v) FACS experiments using hematopoietic cells from the bone marrow and spleen of an adult hsp68*GFP*+110 ln5 mouse to see if the +110 is active in hematopoietic cells in the adult.

5.2 Generating an hsp68*GFP*+110 mouse line

To generate a stable transgenic line, the hsp68*GFP*+110 reporter construct was injected into 1-cell stage mouse embryos by pro-nuclear injection. Embryos were allowed to develop to term, resulting in five TG founder mice, (founders 1-5). Of these founders, four transmitted the hsp68*GFP*+110 transgene through the germline (founders 1, 2, 3 and 5) to generate transgenic embyros. However only founder 5 was able to generate *expressing* transgenic embryos (hsp68*GFP*+110 ln 5). Therefore, timed matings were set-up using hsp68*GFP*+110 ln 5 mice, to visualize *GFP* expression in E8.5, E11.5 and E14.5 TG embryos to verify that expression patterns in hsp68*GFP*+110 ln 5 embryos were similar to those in hsp68*LacZ*+110 F0 and hsp68*GFP*+110 F0 TG embryos. In E8.5 hsp68*GFP*+110 ln 5 TG embryos, *GFP* was seen in a tight band of expression in the YS BIs (Figure 5.1), as was the case in the hsp68*LacZ*+110 and hsp68*GFP*+110 F0 TGs at this time point (chapter four, Figures 4.6 and 4.7). In E11.5 hsp68*GFP*+110 ln 5 TG embryos, *GFP* was expressed in cells of the FL and the YS. In addition, ectopic expression was seen in the CNS and somites (Figure 5.2). In E14.5 hsp68*GFP*+110 ln 5 TG embryos expression was seen in in the FL, DRG and whisker follicles (Figure 5.3) as was seen in hsp68*LacZ*+110 F0 TGs (Figure 4.9 Bi-iii). In the case of FL, clearer expression was seen in hsp68*GFP*+110 ln 5 TG embryos but this may be because generally X-gal stain does not efficiently penetrate the FL thus failing to provide sufficient substrate for cells expressing β-galactosidase.

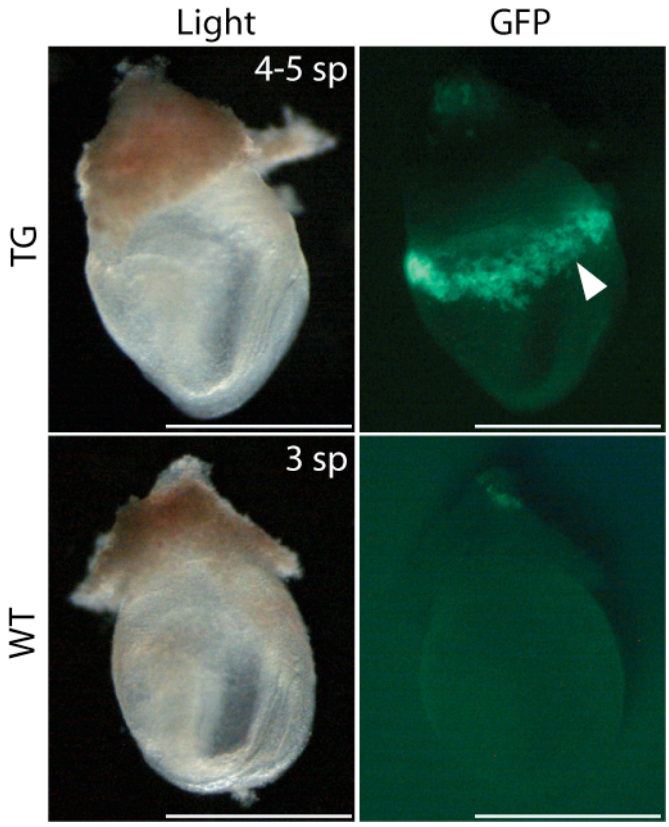


Figure 5.1. *GFP* expression in E8-8.5 whole mount TG and WT hsp68*GFP*+110 line 5 embryos. Images in the left-hand column show a 4-5 sp TG embryo (top) and a 3 sp WT littermate (bottom) shown in the light field. Images in the right-hand column show the same embryos illuminated to see *GFP* expression. In the TG embryo, *GFP* expression is seen in a tight band in the YS blood islands (white arrow head). No expression is seen in the YS of the WT embryo, but ectopic expression is seen in the ectoplacental cone (as is also the case in the TG). Scale bar = 1000 μm.

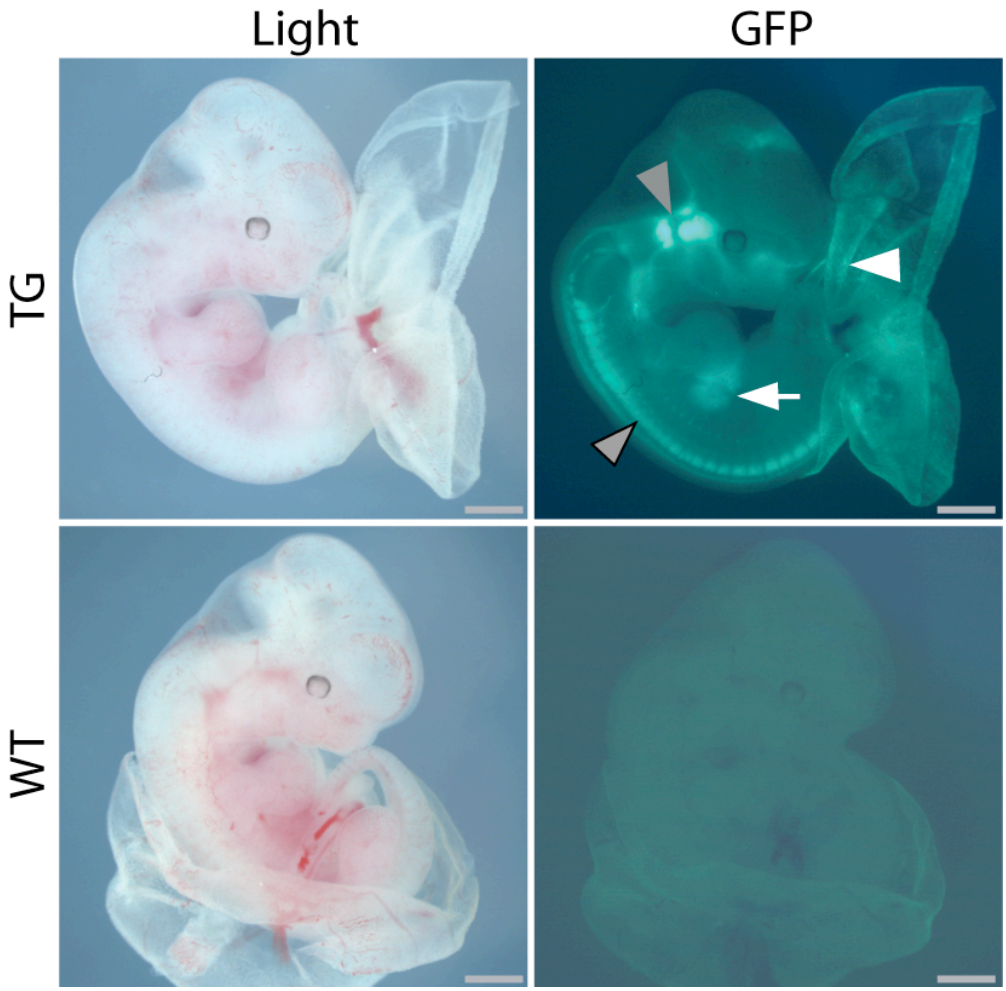


Figure 5.2. *GFP* expression in E11.5 whole mount TG and WT hsp68*GFP*+110 line 5 embryos. Images in the left-hand column show a TG embryo (top) and a WT littermate (bottom) shown in the light field. Images in the right-hand column show the same embryos illuminated to see *GFP* expression. In the TG embryo, *Runx1*-specific *GFP* expression is seen in cells of the FL (white arrow) and the YS (white arrow head). *GFP* expression seen in the CNS (grey arrow head) may be ectopic (hsp68-reporter gene TGs give CNS expression) and expression in the somites is known to be ectopic (grey arrow head with black border). No expression is seen in the WT embryo. Grey scale bar represents 1000 μm.

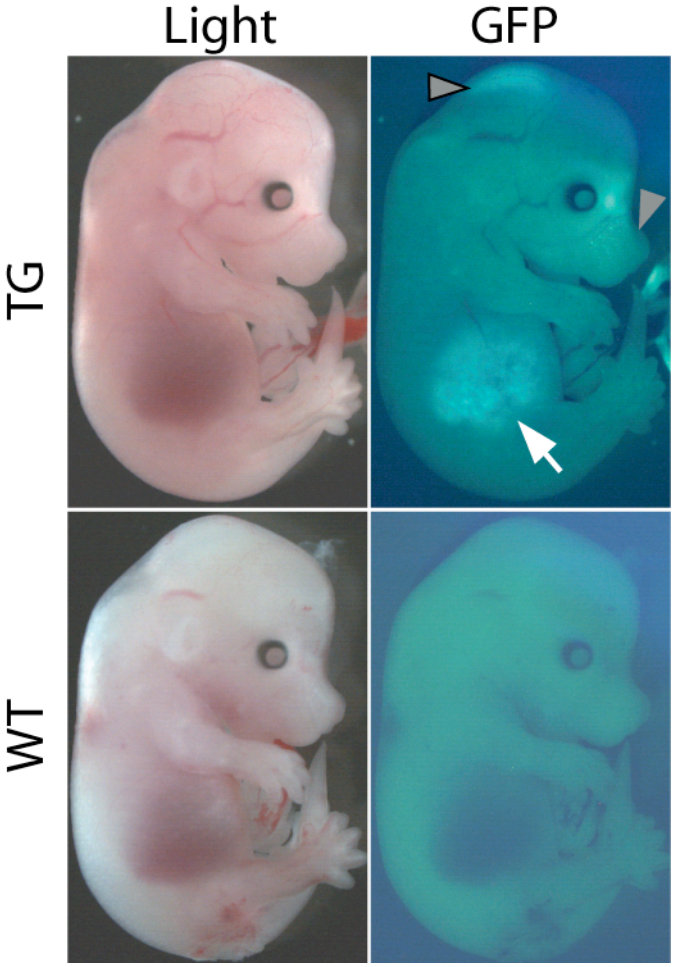


Figure 5.3. *GFP* expression in E14.5 whole mount TG and WT hsp68*GFP*+110 line 5 embryos. Images in the left-hand column show a TG embryo (top) and a WT littermate (bottom) shown in the light field. Images in the right-hand column show the same embryos illuminated to see *GFP* expression. In the TG embryo, *GFP* expression is seen in cells of the FL (white arrow), the whisker follicles (grey arrow head), and the CNS (grey arrow head with black border). No expression is seen in the WT embryo.

5.3 Characterization of cells to which the +110 enhancer mediates GFP expression during developmental hematopoiesis

To identify the cells to which the +110 enhancer mediates *GFP* expression during developmental hematopoiesis, single cell suspensions were prepared from embryonic tissues of E8.5 and E11.5 hsp68*GFP*+110 ln 5 TG embryos. To determine the nature of cells to which the +110 enhancer mediates *GFP* expression during the first waves of hematopoiesis (primitive and transient definitive), whole E8.5 ln 5 embryos were dissociated into single cell suspensions and incubated with antibodies against TER119 (erythroid marker) and CD41 (marker of first definitive-type hematopoietic cells). The percentage of EryPs ($TER119^{+}CD41^{lo}$) and definitive hematopoietic cells ($TER119^{-}CD41^{hi}$) to which the +110 mediates *GFP* expression was then quantified by FACS (Figure 5.4), (Ferkowicz et al., 2003; Mikkola et al., 2003; Nishikawa et al., 1998). 78% of EryPs and 88% of the definitive hematopoietic cells expressed *GFP*, confirming findings from chapter 4 where the +110 enhancer was also found to mediate expression to both the EryPs and definitive hematopoietic cells in an E8.5 hsp68*GFP*+110 F0 TG embryo (Figure 4.8 A and B).

Runx1 is expressed in endothelial and hematopoietic cell types, and is critical for the emergence of the EHT, as described in chapter one (Bertrand et al., 2010; Boisset et al., 2010; Kissa and Herbomel., 2010; North et al., 1999). Therefore, to determine if the +110 enhancer is active in endothelial or

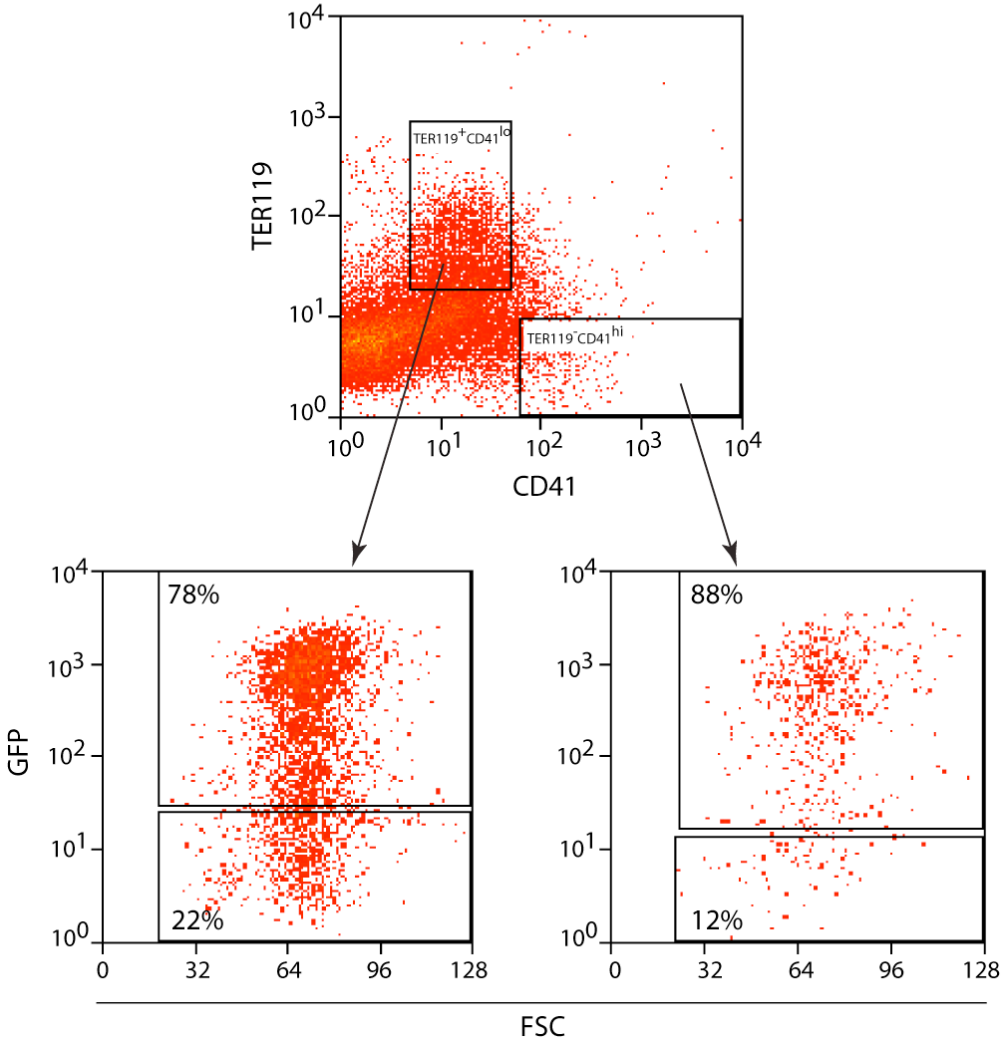


Figure 5.4. Phenotypic characterization of GFP$^+$ and GFP$^-$ cells from a whole hsp68*GFP*+110 line 5 E8.5 TG embryo. Cells were incubated with antibodies against Ter119 and CD41 and analysed by FACS (top panel) to distinguish Ter119$^+$CD41^{lo} (EryPs) from Ter119$^-$CD41^{hi} (definitive hematopoietic) cells. These populations were then gated on to a FSC vs. GFP plot to assess the percentage of cells within these populations in which the +110 enhancer is active (bottom two panels, black arrows direct each population onto the the FSC vs. GFP plots). The top gate for each of the bottom panels indicates GFP$^+$ cells and the bottom gate GFP$^-$ cells (n = 1).

hematopoietic cells, cells from hsp68*GFP*+110 ln 5 TG embryos were analysed by FACS using cell surface markers that distinguish endothelial from hematopoietic cells.

At E11.5, FACS experiments to identify the cell types in which the +110 enhancer is active in cells of the AVU, FL and YS, were carried out. For all tissue types, single cell suspensions were incubated with antibodies against VE-Cadherin (vascular endothelial antigen), TER119 (erythroid marker), CD41 (marker of first definitive-type hematopoietic cells) and CD45 (pan-hematopoietic marker). VE-Cadherin^{+}TER119^{-} (endothelial, non-erythroid) cells were analyzed for their expression of CD41 and CD45, resulting in four populations of interest to analyze for the expression of *GFP*: 1) VE-Cadherin^{+}TER119^{-}CD41^{-}CD45^{-} endothelial cells 2) VE-Cadherin^{+}TER119^{-} CD41^{+}CD45^{-} early hematopoietic cells, 3) VE-Cadherin^{+}TER119^{-} CD41^{+}CD45^{+} hematopoietic cells and 4) VE-Cadherin^{+}TER119^{-}CD41^{-}CD45^{+} further-differentiated hematopoietic cells.

In AVU cells, only 0.5% of endothelial cells were GFP^{+} (Figure 5.5, Gate/Plot 1). Moving clockwise through Figure 5.5 to Gate/Plot 2, 37.1% of early hematopoietic cells (i.e cells that express CD41 but not yet CD45) were GFP^{+}. In population 3, in which hematopoietic cells that express both CD41 and CD45 are found, the majority of cells were GFP^{+} (79.3% of the population). In the further-differentiated hematopoietic cell population, which has lost CD41 expression but maintains CD45 expression, (Gate/Plot 4), the number of GFP+ cells was less than in Gate/Plot 3, at 48.4% of the

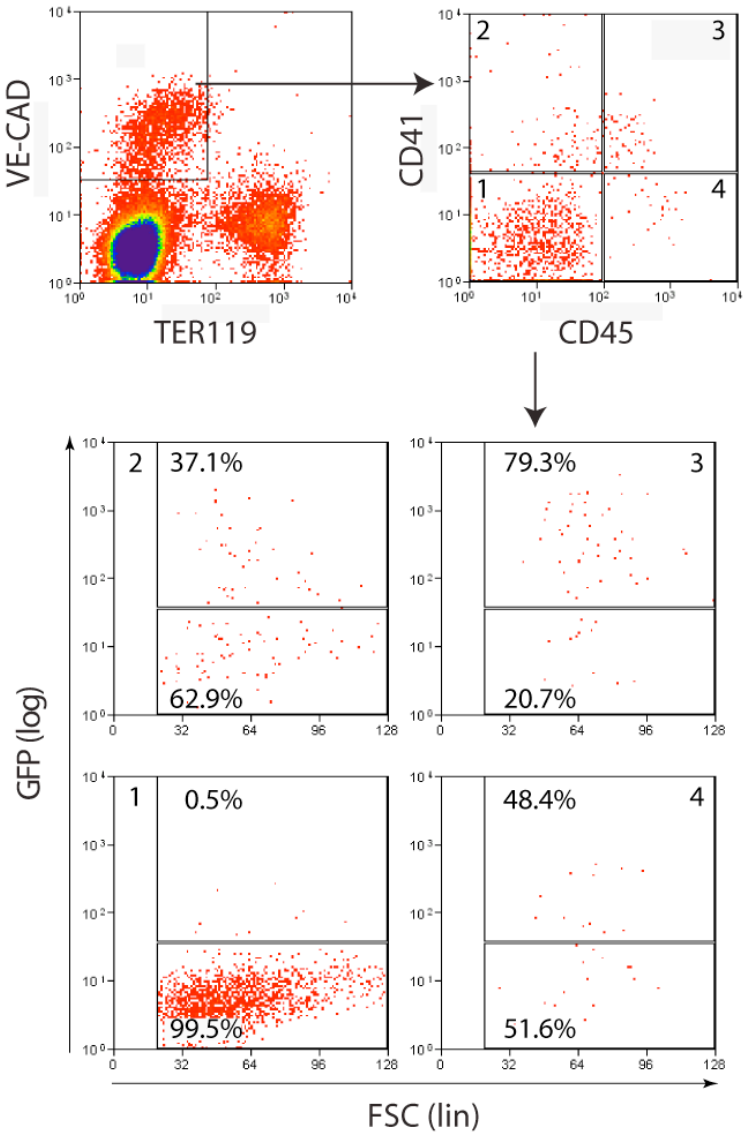


Figure 5.5. Phenotypic characterization of GFP$^+$ and GFP$^-$ cells from the AVU of E10.5 hsp68*GFP*+110 line 5 TG embryos. Cells were incubated with antibodies against VE-Cadherin, Ter119, CD41 and CD45 and analysed for the presence of these markers by flow cytometry. In the top right panel, VE-Cadherin$^+$Ter119$^-$ (non-erythroid) cells are gated onto a CD45 (pan-hematopoietic cell) vs. CD41 (hematopoietic progenitor) plot (horizontal arrow). The vertical arrow from the CD45 vs. CD41 plot gates populations 1- 4 onto FSC vs. GFP plots to assess the percentage of cells from these different populations that are GFP$^+$ and GFP$^-$. Gate+Plot 1 = VE-Cadherin$^+$Ter119$^-$CD41$^-$CD45$^-$ endothelial cells, Gate+Plot 2 = VE-Cadherin$^+$Ter119$^-$CD41$^+$CD45$^-$ early hematopoietic cells, Gate+Plot 3 = VE-Cadherin$^+$Ter119$^-$CD41$^+$CD45$^+$ hematopoietic cells and Gate+Plot 4 = VE-Cadherin$^+$Ter119$^-$CD41$^-$CD45$^+$ further matured hematopoietic cells. Percentages within the top gates represent GFP$^+$ cells within these populations and percentages within the bottom gates represent GFP$^-$ cells.

population. This latter population is enriched for HSC-precursors and the first HSC at this time point in the DA (Taoudi et al., 2008; Taoudi et al., 2005).

Interestingly, when looking at the same markers in cells of the FL and YS of the same TG embryos, the percentages of GFP^+ cells in the four different populations were very similar to those seen in the same populations in the AVU cells (Figures 5.6 for FL and Figure 5.7 for YS). A very low percentage of GFP^+ cells in the endothelial population was observed, with the percentage of GFP^+ cells being highest in the intermediate VE-$Cadherin^+TER119^-$ $CD41^+CD45^+$ population. These data are summarized in Figure 5.8, where the percentage of GFP^+ cells in a population can be seen to increase as cells in the AVU, FL and YS begin to take on a hematopoietic fate (expressing CD41). *GFP* expression peaks when cells express both CD41 and CD45 and then as cells mature to become committed hematopoietic cells (by losing CD41) the percentage of cells expressing GFP then decreases again slightly.

5.4 Clonogenic potential of +110 targeted cells: CFU-C assays

To assess if the +110 enhancer is active in clonogenic progenitor cells in the embryo, the AVU, FL and YS were dissected from E11.5 hsp68*GFP*+110 ln 5 TG embryos and made into single cell suspensions. Cells were then sorted into GFP^+ and GFP^- populations for each tissue type, using a MoFlo cell sorter (Kevin Clark). GFP^+ and GFP^- sorted cells were then plated in MethoCult (Stemcell Technologies) supplemented with hematopoietic growth factors. Unsorted cells from the AVU, FL and YS of hsp68*GFP*+110 ln 5 TG

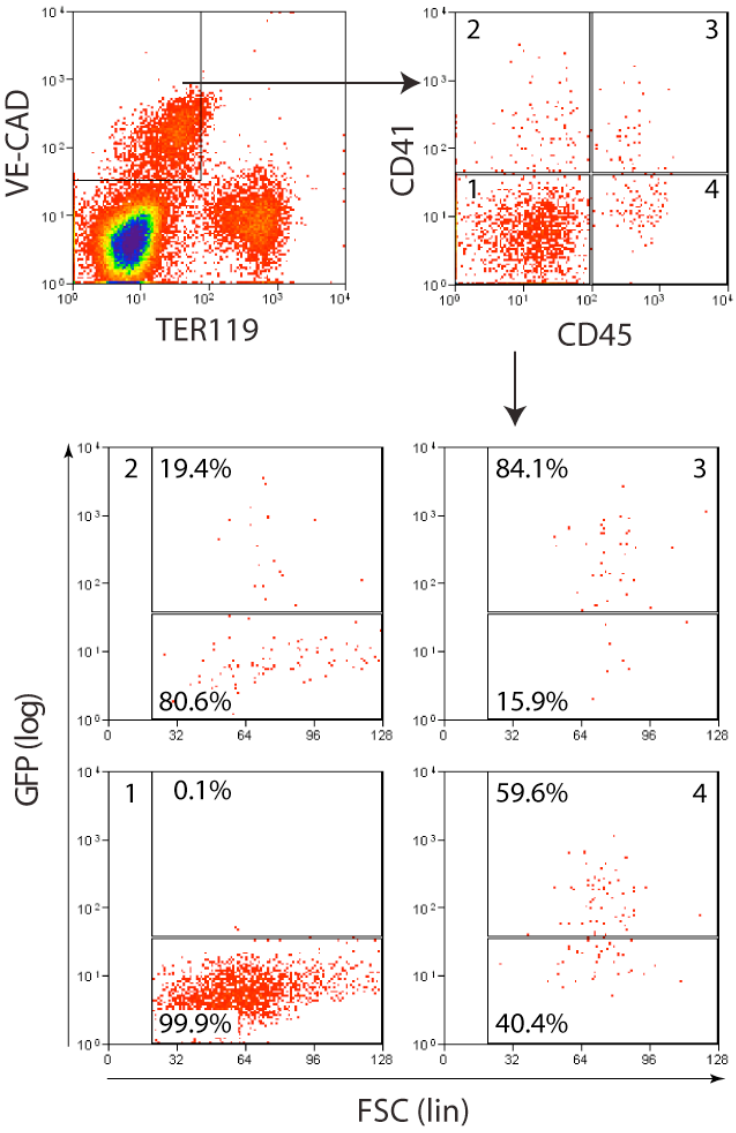


Figure 5.6. Phenotypic characterization of GFP$^+$ and GFP$^-$ cells from the FL of E10.5 hsp68*GFP*+110 line 5 TG embryos. Cells were incubated with antibodies against VE-Cadherin, Ter119, CD41 and CD45 and analysed for the presence of these markers by flow cytometry. In the top right panel, VE-Cadherin$^+$Ter119$^-$ (non-erythroid) cells are gated onto a CD45 (pan-hematopoietic cell) vs. CD41 (hematopoietic progenitor) plot (horizontal arrow). The vertical arrow from the CD45 vs. CD41 plot gates populations 1- 4 onto FSC vs. GFP plots to assess the percentage of cells from these different populations that are GFP$^+$ and GFP$^-$. Gate+Plot 1 = VE-Cadherin$^+$Ter119$^-$CD41$^-$CD45$^-$ endothelial cells (there is no hemogenic endothelium in the FL), Gate+Plot 2 = VE-Cadherin$^+$Ter119$^-$CD41$^+$CD45$^-$ early hematopoietic cells, Gate+Plot 3 = VE-Cadherin$^+$Ter119$^-$CD41$^+$CD45$^+$ hematopoietic cells and Gate+Plot 4 = VE-Cadherin$^+$Ter119$^-$CD41$^-$CD45$^+$ further matured hematopoietic cells. Percentages within the top gates represent GFP$^+$ cells within these populations and percentages within the bottom gates represent GFP$^-$ cells.

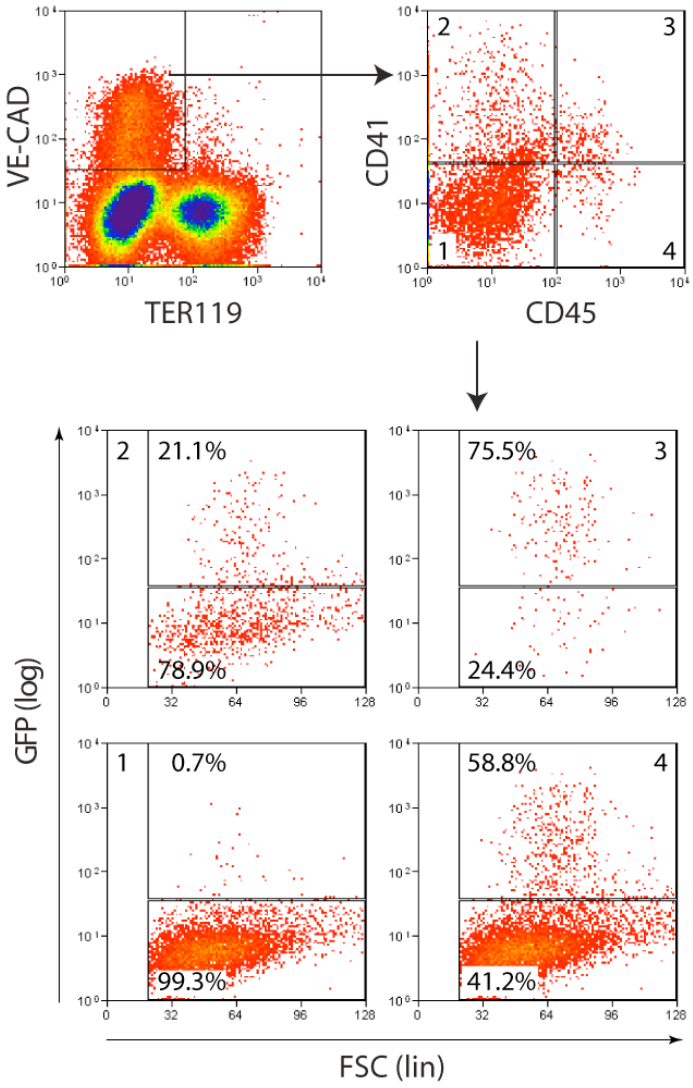


Figure 5.7. Phenotypic characterization of GFP$^+$ and GFP$^-$ cells from the YS of E10.5 hsp68*GFP*+110 line 5 TG embryos. Cells were incubated with antibodies against VE-Cadherin, Ter119, CD41 and CD45 and analysed for the presence of these markers by flow cytometry. In the top right panel, VE-Cadherin$^+$Ter119$^-$ (non-erythroid) cells are gated onto a CD45 (pan-hematopoietic cell) vs. CD41 (hematopoietic progenitor) plot (horizontal arrow). The vertical arrow from the CD45 vs. CD41 plot gates populations 1- 4 onto FSC vs. GFP plots to assess the percentage of cells from these different populations that are GFP$^+$ and GFP$^-$. Gate+Plot 1 = VE-Cadherin$^+$Ter119$^-$CD41$^-$CD45$^-$ endothelial cells, Gate+Plot 2 = VE-Cadherin$^+$Ter119$^-$CD41$^+$CD45$^-$ early hematopoietic cells, Gate+Plot 3 = VE-Cadherin$^+$Ter119$^-$CD41$^+$CD45$^+$ hematopoietic cells and Gate+Plot 4 = VE-Cadherin$^+$Ter119$^-$ CD41$^-$CD45$^+$ further matured hematopoietic cells. Percentages within the top gates represent GFP$^+$ cells within these populations and percentages within the bottom gates represent GFP$^-$ cells.

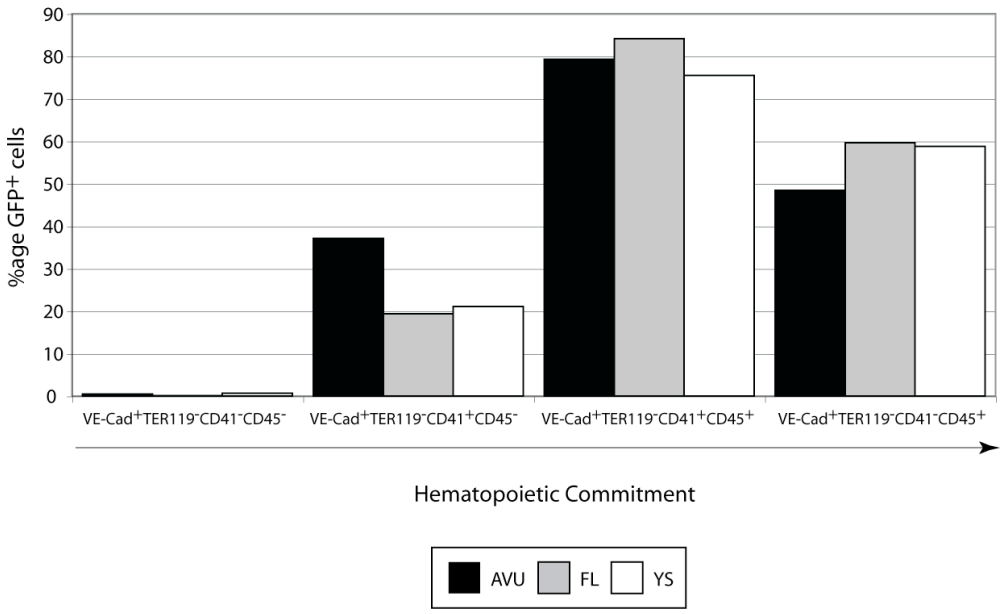


Figure 5.8. Summary of *GFP* expression mediated by the +110 enhancer to endothelial and hematopoietic cells of the AVU arteries, FL and YS of E11.5 hsp68*GFP*+110 ln5 TG embryos. Data represent the %age of *GFP* positive cells in populations of increasing hematopoietic commitment from endothelial cells (VE-Cadherin^{+}TER119^{-}CD41^{-}CD45^{-}) through early hematopoietic cells (VE-Cadherin^{+}TER119^{-}CD41^{+}CD45^{-}) and hematopoietic cells (VE-Cadherin^{+}TER119^{-}CD41^{+}CD45^{+}) to further matured hematopoietic cells (VE-Cadherin^{+}TER119^{-}CD41^{-}CD45^{-}). Perecentages from AVU cells are represented by black bars, FL = grey bars, YS = white bars (n = 1).

embryos and from WT embryos were plated as controls. Cells were then incubated and the number of hematopoietic colonies formed was scored after 7 days. Results form three independent experiments show that CFU-Cs are enriched in the GFP+ population in all three tissues analyzed from TG E11.5 hsp68*GFP*+110 ln 5 embryos in comparison to the GFP- populations (Figure 5.9). This indicates that the +110 enhancer is active in CFU-Cs in the AVU, YS and FL.

5.5 Long-term reconstitution ability of cells in which the +110 is active

The gold standard for cells to be classed as HSCs dictates that cells must be able to provide long-term, multi-lineage reconstitution of the hematopoietic system of a lethally irradiated adult mouse recipient. To assess whether the +110 enhancer targets HSCs in E11.5 hsp68*GFP*+110 ln 5 embryos, single cell suspensions from TG and WT AVUs and FLs were made and sorted based on *GFP* expression using a MoFlo cell sorter (Kevin Clark). Compared to WT embryos, 2.3% of AVU cells and 24.6% of FL cells were GFP^+, respectively (Figure 5.10). The purity of GFP^+ and GFP^- sorted populations are shown in the right-hand windows of Figure 5.10 (GFP^+ Sort) and show that the purity of GFP^+ sorted cells from the AVU and FL were 77.7% and 80.9% respectively in this sort. The purity of GFP^- sorted cells from the AVU and FL cells were 99.6% and 97.7% respectively.

Transplantation assays were performed by injecting 1.5 to 2 embryo-equivalent numbers of GFP^- or GFP^+ sorted AVU or FL cells into the tail

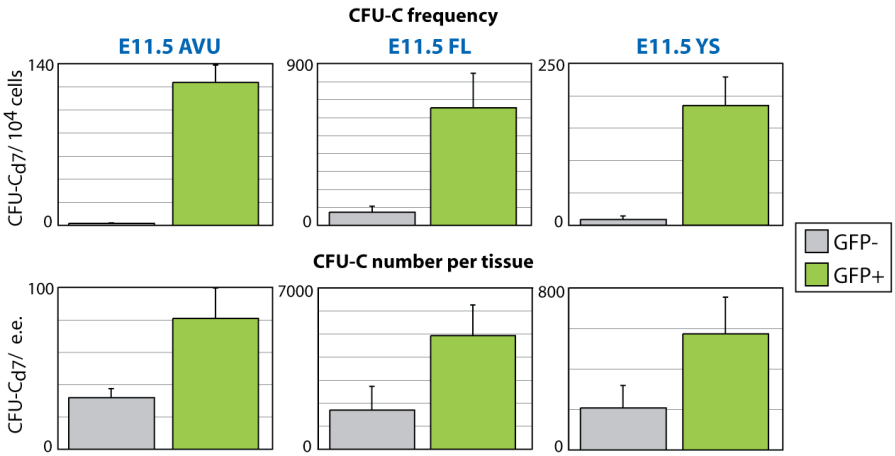


Figure 5.9. Results of CFU-C assays using AVU, FL and YS cells from E11.5 hsp68*GFP*+110 ln5 TG embryos. The top three graphs show the frequency of CFU-Cs generated from sorted GFP+ and GFP- cells per 10^4 cells for each tissue after 7 days in culture (CFU-C_{d7}/10^4 cells). The bottom three graphs show the number of CFU-Cs generated from GFP+ and GFP- cells per embryo equivalent for each tissue after 7 days in culture (CFU-Cd_7/ e.e.). Results are the mean ± SD of three experiments.

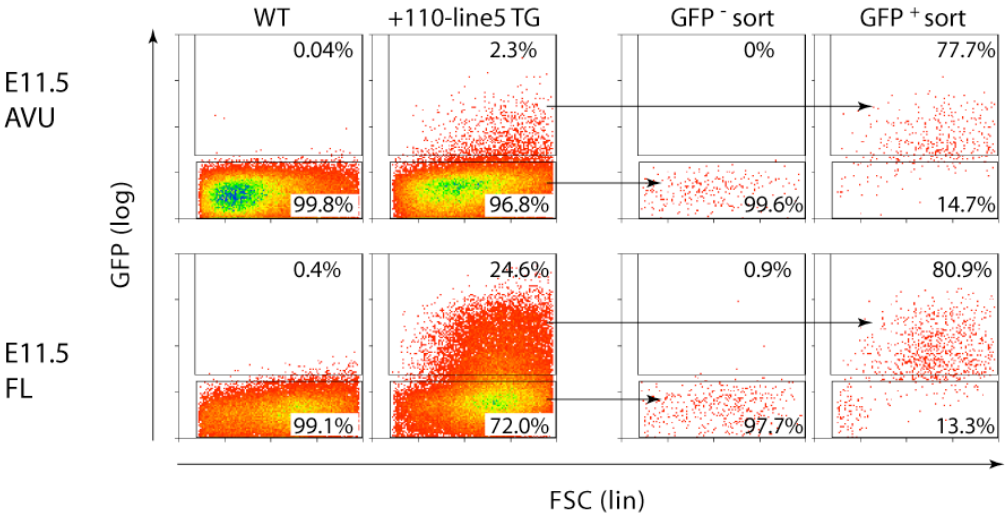


Figure 5.10. Representative FACS plots of cell sorting experiments using cells from E11.5 hsp68*GFP*+110 line 5 TG and WT embryos. The x-axis indicates the size of cells by forward scatter (FSC) and the y-axis represents *GFP* expression. The top row of plots represents cells sorted from the AVU whilst the bottom row represents cells from the FL. Plots for both the AVU and FL cells on the left hand side of the figure show the percentage of cells that are GFP+ (top gate, e.g. 2.3% for TG E11.5 AVU) and GFP- (bottom gate, e.g. 96.8% for TG E11.5 AVU) after sorting from both the WT and TG tissues. The plots on the right hand side of the figure represent purity checks of the GFP- and GFP+ sorted cell populations, to gauge what percentage of the sorted population are of that phenotype (indicated by arrows coming from right hand plot gates). A percentage of 100% would indicate a 100% pure population with no contamination from the other cell population.

veins of lethally irradiated mice. Transplanted mice were then bled 1-3 months later. Genomic DNA was isolated from blood samples and tested for the presence of the *GFP* transgene by PCR. In the case of mice transplanted with GFP^- or GFP^+ sorted AVU cells, 67% of those mice transplanted with GFP^- AVU cells (2/3) showed donor reconstitution. However, 0% (0/3) of mice transplanted with GFP^+ AVU cells showed donor reconstitution suggesting that the +110 enhancer does not target HSCs in these experiments. Results for the FL remain for be verified. So far, reconstitution has only been verified in one adult that received GFP- cells. This suggests that the +110 is also inactive in FL HSCs.

5.6 Megakaryocyte differentiation from E14.5 hsp68*GFP*+110 fetal livers

To characterize the activity of the +110 enhancer in a definitive hematopoietic cell type, megakaryocytes were chosen. Runx1 is known to be involved in the generation of megakaryocytes (Goldfarb, 2007, 2009; Pencovich et al., 2011) but less is known about Runx1's role in the gene-regulatory pathways that dictate the formation of these cells from the megakaryocyte-erythroid progenitor (MEP). In collaboration with the Groner group (Weizmann Institute, Israel), which has an interest in Runx1's role megakaryocyte lineage specification, megakaryocytes were differentiated from E14.5 hsp68*GFP*+110 ln 5 TG fetal livers. This was done in parallel with cells differentiated from E14.5 hsp68*GFP*+23+24 ln 3 TG fetal livers, which had not been done before.

In this assay, single cell suspensions from the FLs of E14.5 TG hsp68*GFP*+110 ln5 and hsp68*GFP*+23+24 ln3 embryos were made and cultured in the presence of TPO supernatent for four days to induce megakaryocyte differentiation. After four days, cells were incubated with an antibody against CD41, which, in addition to early hematopoietic cells, marks megakaryocytes. Cells were analyzed by FACS to assess the percentage of megakaryocytes to which the +110 and +23+24 enhancers mediate *GFP* expression (Figure 5.11). The $CD41^{+}GFP^{-}$ population (i.e megakaryocytes in which the enhancers were not active) made up 68.3% of the hsp68*GFP*+110 derived megakaryocytes, whereas in the hsp68*GFP*+23LF population only 7.7% of megakaryocytes were GFP^{-}. Therefore, in the $CD41^{+}GFP^{+}$ populations, the +110 enhancer mediated expression to only 31.7% of megakaryocytes whilst the +23LF enhancer was active in 92.3% of the cells.

5.7 Analysis of +110 enhancer activity in adult hematopoietic organs

Runx1 is expressed in the majority of myeloid and lymphoid cells in the adult mouse (North et al., 2004). Therefore, to assess if the +110 enhancer contributes to *Runx1* expression in hematopoietic cells of the adult mouse, we examined +110-mediated *GFP* expression in cells of the spleen and bone marrow of an adult hsp68*GFP*+110 ln 5 TG mouse. The same was done with corresponding tissues from a WT mouse to provide controls. Spleen cells were incubated with antibodies against TER119, Mac1, and B220, to select for $TER119^{+}$ erythroid cells, $Mac1^{+}$ myeloid cells (macrophages, granulocytes and dendritic cells) and $B220^{+}$ B-cells. In the $TER119^{+}$ erythroid

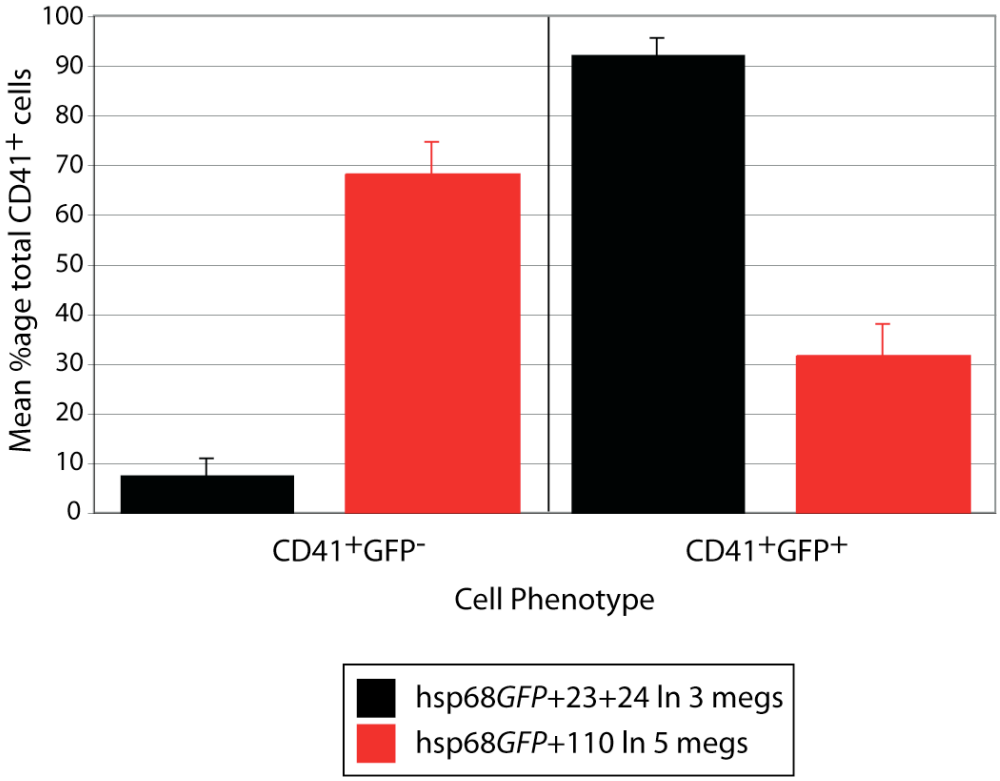


Figure 5.11. Analysis of GFP$^+$ and GFP$^-$ megakaryocytes (megs) cultured from E14.5 hsp68*GFP*+23+24 and hsp68*GFP*+110 TG fetal livers. Fetal livers were cultured for 4 days in the presence of TPO to induce meg differentiation. After 4 days, cells were incubated with an antibody against CD41 and analysed by FACS. Live cells were deemed to be megs by being positive for CD41 expression. To assess the ability of both the +23+24 and the +110 enhancers to direct *GFP* expression to these CD41$^+$ megs, expression of *GFP* in megs derived from both types of TG fetal liver was assessed, the results of which are shown in the bar graph above. Data in the left-hand panel represent the mean percentage of megs to which the enhancers did <u>not</u> direct *GFP* expression ([CD41$^+$GFP$^-$] hsp68*GFP*+23+24 ln 3 megs [black bar] and hsp68*GFP*+110 ln 5 [red bar]). Data in the right-hand panel represent the mean percentage of megs to which the enhancers did direct *GFP* expression ([CD41$^+$GFP$^+$] hsp68*GFP*+23+24 ln 3 megs [black bar] and hsp68*GFP*+110 ln 5 [red bar]). Data are the mean of 3 experiments. Error bars represent the SD of the mean.

population the +110 was active in less than half of the population (42.6% of the cells were GFP^+ with 57.4% GFP^-). In the $Mac1^+$ population slightly fewer cells were GFP^+ at 28.1% of the population (71.9% GFP^-) whilst in the $B220^+$ B-cell population, the +110 enhancer was only active in 10% of the cells with 90% being GFP^- (Figure 5.12 top panel). Bone marrow cells were incubated with antibodies against Gr1 and CD31 in order to select for $Gr1^+$ mature granulocytes, $CD31^+$ lymphoid progenitors, and $CD31^+Gr1^+$ myeloid progenitor cells. The +110 enhancer was active in just less than half of the $Gr1^+$ mature granulocyte and monocyte population (47.2% of cells GFP^+ vs. 52.8% GFP^-). It was less active in $CD31^+$ lymphoid progenitors (23.7% of cells GFP^+ vs. 76.3% GFP^-) and in $CD31^+Gr1^+$ myeloid progenitors the +110 enhancer was active in 36.7% of cells, with 63.3% of cells GFP^- (Figure 5.12, bottom panel).

5.8 Chapter Discussion

The aim of this chapter was to begin to characterize functionally, +110 enhancer-targeted cells in the embryo and in adult mice. The combination of assays used (described in section 5.1) has generated interesting results regarding the function of this novel enhancer. FACS experiments have shown that the +110 enhancer targets both EryPs and transient definitive cells in the E8.5 embryo. In E11.5 embryos, it also targets the majority of VE-$Cadherin^+TER119^-CD41^+CD45^+$ hematopoietic cells in the AVU, FL and YS. Interestingly however, the +110 enhancer does not target HSCs in the AVU of hsp68*GFP*+110 ln5 TG embryos, in contrast to the +23+24 enhancer,

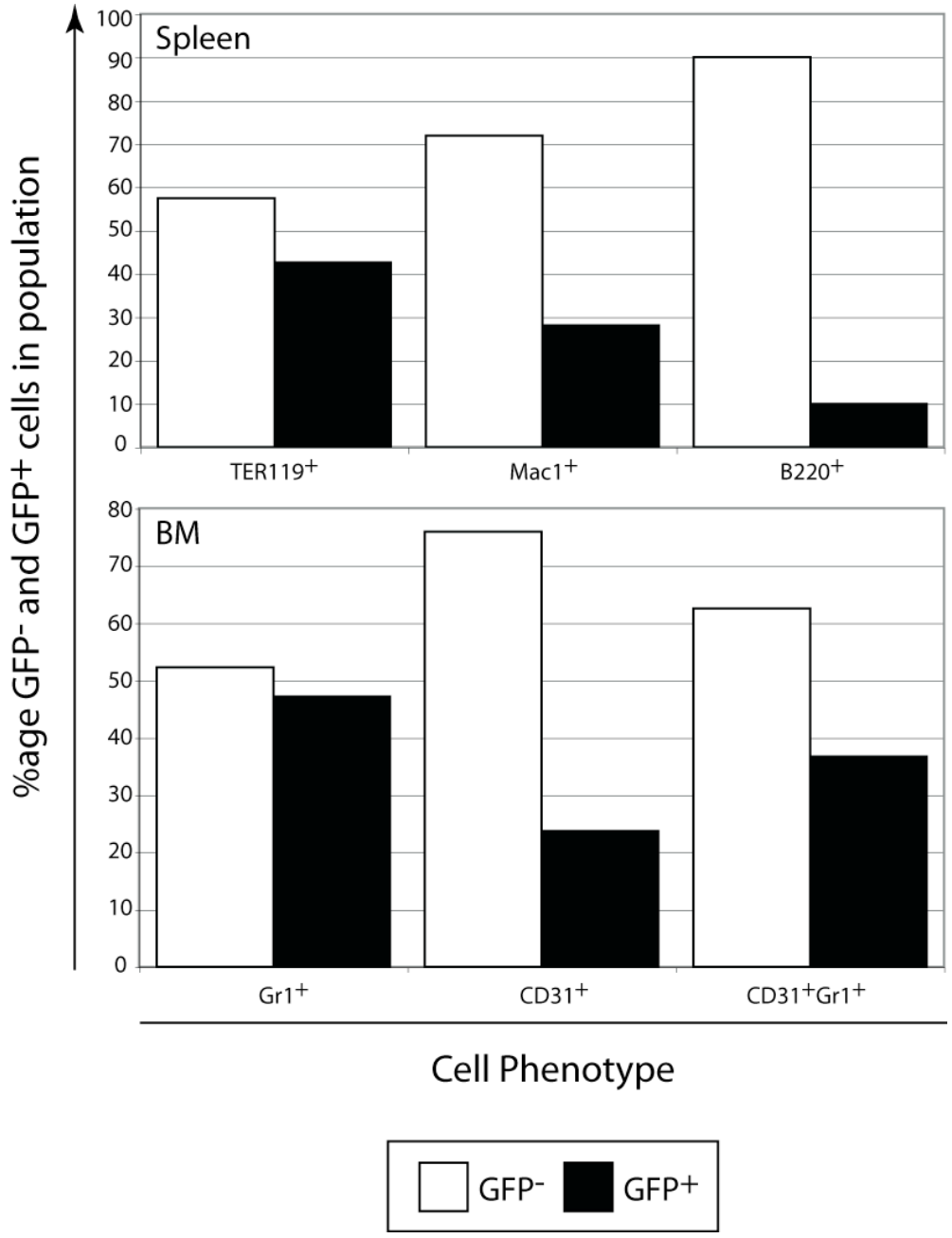


Figure 5.12. Summary of *GFP* expression in hematopoietic cell lineages within the spleen and bone marrow of an adult hsp68*GFP*+110 ln5 TG mouse. The top panel represents the percentage of GFP^- and GFP^+ cells within $TER119^+$ erythroid, $Mac1^+$ monocyte and $B220^+$ B-cell populations of the spleen. The bottom panel represents the percentage of GFP^- and GFP^+ cells within $CD31^+$ lymphoid progenitor, $Gr1^+$ mature granulocytes and monocytes and $CD31^+Gr1^+$ myeloid progenitors of the bone marrow (BM). White bars represent GFP^- populations and black bars represent GFP^+ populations. Results are from the analysis of one adult hsp68*GFP*+110 ln5 TG mouse.

which does (Bee et al., 2009a; Nottingham et al., 2007). Despite this, CFU-C data demonstrate that +110 enhancer-targeted cells are enriched for CFU-Cs and that the +110 targets the majority of clonogenic progenitors in the AVU, FL and YS of E11.5 hsp68*GFP*+110 ln5 TG embryos. The +110 enhancer was also found to be active in a small proportion of megakaryocytes in comparison to the +23+24 enhancer (which was active in the majority) and in sub-populations of adult hematopoietic cell populations. Interestingly, the +110 only marks 0.5% of the endothelial cells in the AVU, versus approximately 5% of cells marked by the +23+24 enhancer, suggesting that the +110 enhancer may not be active in HE. Collectively, these data suggest that the +110 enhancer is active later in hematopoietic development than the +23+24 enhancer, increasing as cells mature along the hematopoietic lineages downstream of the HSC.

Future experiments to define the differences between the +110 and +23+24 enhancer sequences will be important. This will allow comparison of putative binding sites for upstream TFs, which may determine the different activities of the enhancers during developmental hematopoiesis. In HPC-7 ES-derived progenitor cells for example, both enhancers are bound by the same nine hematopoietic TFs (Runx1, Gata2, Scl, Lmo2, Fli1, Gfi1b, PU.1, Erg and Lyl1). In contrast however the +110 enhancer does not bind Meis1, whereas the +23+24 enhancer does (Wilson et al., 2010) (Chapter three, table 3.1). Whether binding of Meis1 is key to the differences in expression mediated by the +23+24 vs. the +110 is unknown. Indeed, binding of TFs to sequences does not necessarily implicate that they have function at those sites. However,

distinct differences between TF binding at enhancers may be more indicative of a role in regulating one enhancer vs. another in a particular cell type. Mutation of putative TF binding sites will provide more direct evidence of what regulates the +110 enhancer, and has already been initiated for the +23 enhancer (Nottingham et al., 2007).

Chapter Six:

Chromosome Conformation Capture (3C):

Physical interactions of regulatory elements with the P1 and P2 promoters

6.1 Chapter Aim

The overall aim of this thesis has been to identify and characterize hematopoietic-specific *Runx1* enhancers, to understand what controls blood specification and commitment during development. Future experiments aim to extend the use of chromatin assays to characterize physical interactions in *cis* between the *Runx1* promoters and other genomic regions, which can be indicative of *cis*-regulation. This will be done using the 4C technique, which is a non-biased method used to detect such interactions. This technique would therefore provide a library of all putative *Runx1 cis*-regulatory elements that physically interact with the P1 and P2 promoters for *in vivo* analysis. As a precursor to 4C studies, this chapter aims to establish the 3C technique. This technique differs from 4C as it quantifies interactions between genomic regions defined by the investigator. I will investigate interactions between the *Runx1* P1 and P2 promoters and putative and known regulatory elements within the *Runx1* syntenic region, identified in previous chapters. Some additional putative elements will also be investigated. This will be done in four different cell types, to assess any cell-type specific differences in promoter-regulatory element interaction.

6.2 The Chromosome Conformation Capture (3C) assay

The Chromosome Conformation Capture (3C) assay involves the chemical cross-linking of chromatin to fix in place genomic regions that are physically

interacting with each other. The chromatin is then digested with a restriction enzyme to generate fragments that contain genomic elements of interest (e.g. *cis*-regulatory elements). These genomic fragments are then ligated together so that, after reversal of the chemical cross-links, interacting restriction fragments form one DNA sequence. The interaction between two fragments is then quantified by qPCR using Taqman probes and primers. A forward (F) primer is designed to bind the 'partner' fragment and the reverse primer (R) and FAM-TAMRA probe (P) are designed to bind the 'anchor' fragment. The latter fragment is the one that remains constant in the experiment, whilst the partners interacting with it vary, as specified by the investigator. This experimental process is illustrated in Figure 6.1, where the interactions between a hypothetical promoter (anchor) and enhancer (partner) are detected by 3C (Figure 6.1) (Dekker, 2006; Dekker et al., 2002; Splinter et al., 2004).

The interaction frequency between genomic regions is inversely proportional to the distance between them. This must be considered when looking at interactions between an anchor sequence and partner sequences, as random collisions due to proximity of genomic fragments can be mistaken for specific interactions. Figure 6.2 illustrates this point. For example, in the absence of specific looping events that would bring fragments together, non-specific interactions are detectable between fragments up to 40 Kb apart. After a distance of 40 Kb between fragments, random interactions should not be expected (Figure 6.2, graph, open circles). In the case of a specific interaction (e.g. between a promoter and an enhancer) a local peak of interaction would

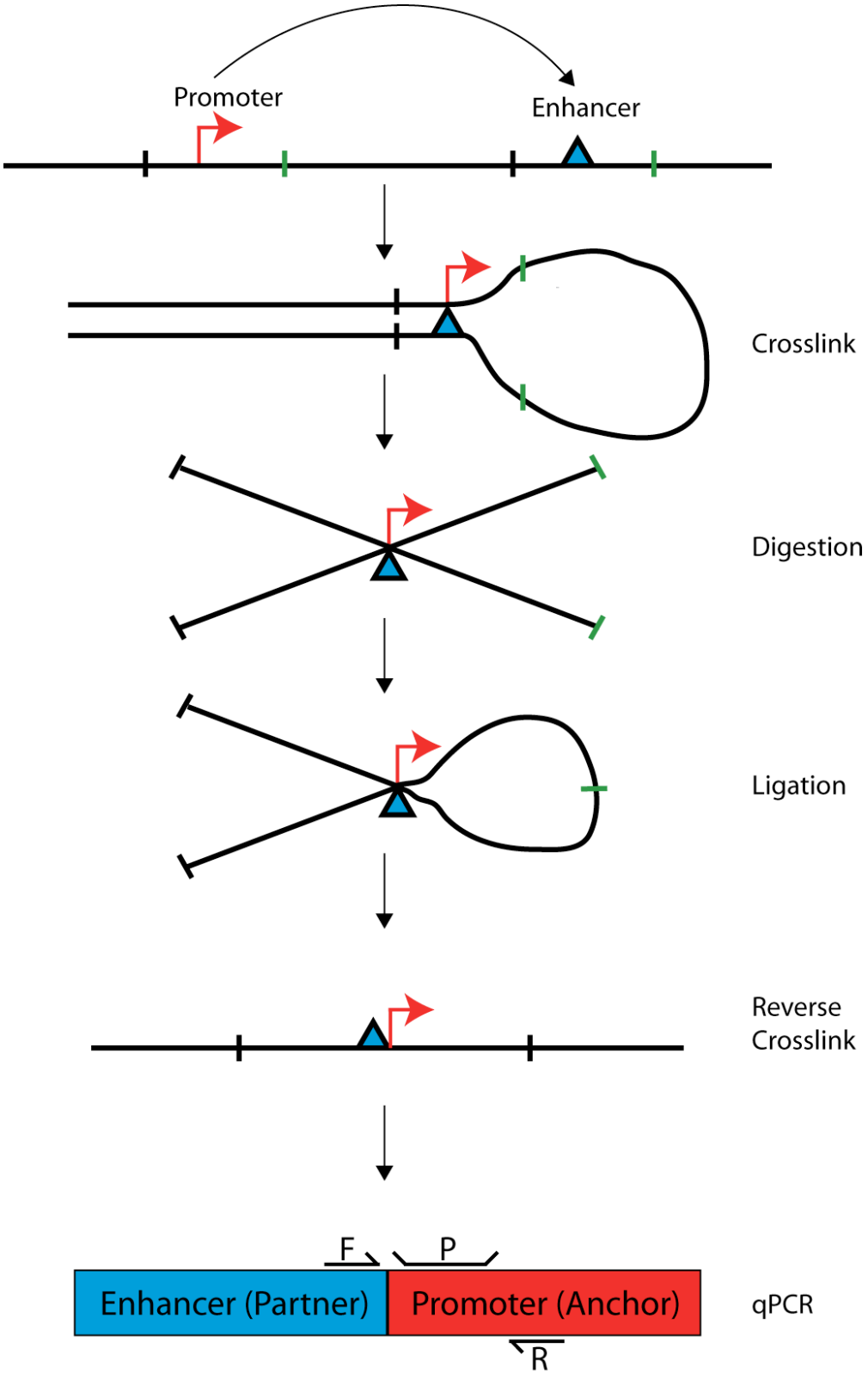


Figure 6.1. For legend, see overleaf.

Figure 6.1. Schematic representation of the Chromosome Conformation Capture (3C) assay. Interactions between genomic regions e.g. a promoter (red arrow) and an enhancer (blue triangle) are fixed by chemical cross-linking. Next chromatin is digested with a restriction enzyme to generate fragments that contain the regions of interest (restriction sites are indicated by green lines). Digested chromatin is then ligated together and chemical crosslinks reversed. The resulting sequence of DNA is analyzed by qPCR using Taqman probes and primers, to quantify the interaction occurring between the enhancer and promoter. Forward primers are designed on the 'Partner' sequence (i.e. the enhancer) and the reverse primer and FAM-TAMRA probe designed on the 'Anchor' sequence (i.e. the promoter). Abbreviations: F, forward primer; R, reverse primer; P, FAM-TAMRA probe. Adapted from (Dekker, 2006).

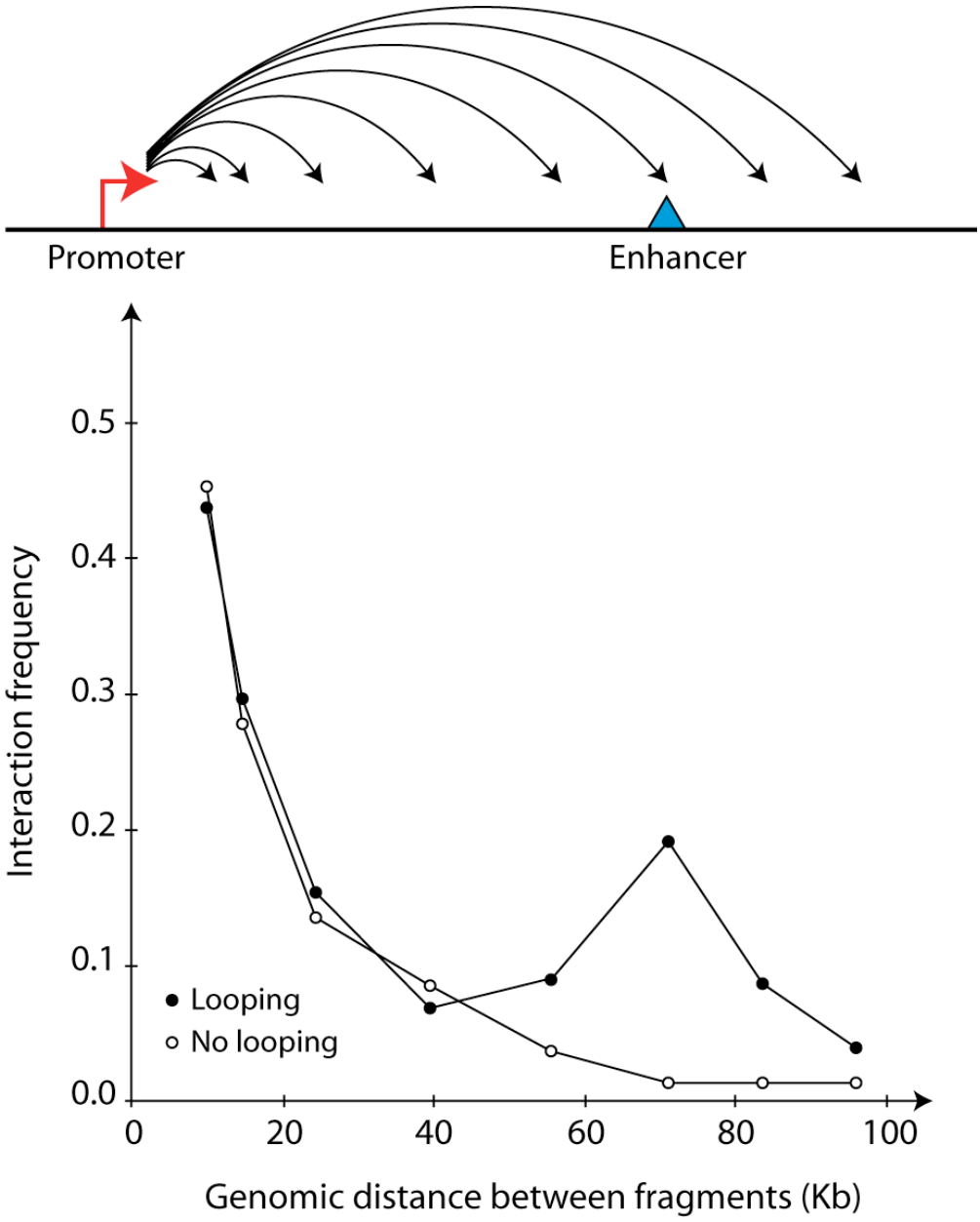


Figure 6.2. Theoretical model of chromosome conformation capture (3C). The top schematic represents 3C analysis to detect *cis* interactions between a gene promoter (red arrow) and a regulatory element (in this case an enhancer, blue triangle). Open circles in the graph below indicate predicted interactions between the promoter and sites located up to 100 Kb away in the absence of specific looping contacts. These interactions represent random collisions. Filled circles illustrate the expected pattern of interactions in the presence of a specific interaction between the promoter and the distant enhancer. The presence of the

specific promoter-enhancer interaction is apparent by the local peak in interaction frequencies. Adapted from (Dekker, 2006).

be detectable at distances where random collisions would not be expected, for example with 70 Kb distance between fragments (Figure 6.1, closed circles).

6.3 *Runx1* 3C

6.3.1 Rationale for use of different cell types in 3C assays

For *Runx1* 3C experiments, four *Runx1*-expressing cell lines were chosen to reflect different stages of hematopoietic commitment. Ideally, multiple stages from the non-hematopoietic, pluripotent ES cell stage through to committed hematopoietic cells of the erythroid and myeloid lineages would be examined (Figure 6.3). However, cell lines representing hemogenic endothelium and HSCs are currently not available. The J1 mouse ESC line was chosen as a source of *in vitro* ESCs. In these cells, *Runx1* is only expressed from the P2 promoter and at very low levels. 416B myeloid progenitor cells were used to represent the first myeloid cells *in vivo*. *Runx1* is expressed from both the P1 and P2 promoters in this cell line. Finally, un-induced mouse erythroleukaemia cells, or 'U-MEL' (pro-erythroblast equivalent, *α-globin* inactive) and induced mouse erythroleukaemia, 'I-MEL' cells (mature erythroblast equivalent, *α-globin* active) were used. Both U-MEL and I-MEL cells only express P1-derived transcripts (Figure 6.3). U- and I-MEL cells were also attractive cells to use as they have been used extensively in 3C studies of the *α-globin* locus by the Higgs group in our department (Vernimmen et al., 2007). This therefore gave us access to

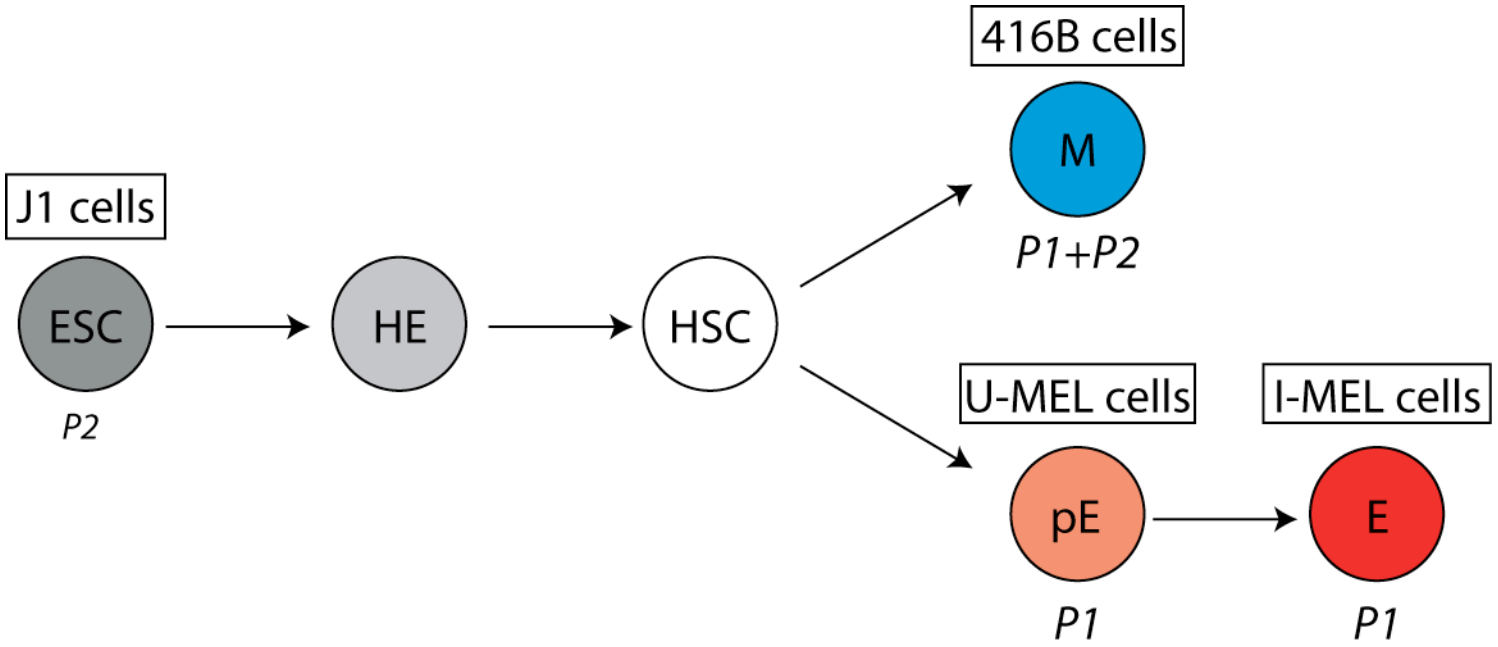


Figure 6.3. For legend, see overleaf.

Figure 6.3. A simplified schematic of developmental hematopoiesis highlighting the different cell types used in 3C assays. In the schematic, embryonic stem cells (ESC) undergo lineage restriction to form hemogenic endothelial cells (HE), which give rise to hematopoietic stem cells (HSC). HSCs give rise to myeloid cells (M). In addition the HSC gives rise to pro-erythroid cells (pE), which further commit to mature erythroid cells (E). Cell lines used in 3C experiments to represent cell types at the ESC, M, pE and E stages of hematopoiesis are indicated above the relevant cell in the schematic. These are: the J1 mouse embryonic stem cell line; the 416B mouse myeloid progenitor cell line; the U-MEL un-induced mouse erythroleukaemic cell line and the I-MEL, induced mouse erythroleukaemic cell line. Below each cell type used in 3C, the promoter(s) (P1, P2 or P1+P2) from which *Runx1* transcripts are derived is indicated. NB: in J1 cells, *Runx1* is transcribed from the P2 at a very low level compared to other cell types.

positive and negative control Taqman probes and primers (kind gift of Dr. Douglas Vernimmen). In addition, upon induction of U-MEL to I-MEL cells, the level of *Runx1* expression (as determined by qPCR) increases four-fold (Figure 6.4). Therefore, these four cell lines provided material in which the *Runx1* promoters are used differentially and to varying degrees, which would help highlight any specific differences or similarities in enhancer-promoter interactions depending on the cellular context.

6.3.2 Testing Runx1 3C Taqman probes and primers by qPCR

Runx1 3C probes and primers were designed using HindIII fragments containing the P1 as anchor sequence plus partners, or the 3'P2 HindIII fragment plus partners (for a full list of anchor:partner sequences and the rationale for their selection, see table 6.1). The 3'P2 fragment is the HindIII fragment directly 3' to the P2. This was used as probes and primers couldn't be designed using the P2 HindIII fragment, which is GC-rich. To check that probe and primer combinations worked specifically with the ligation products generated by 3C experiments, qPCR products were run on 8% TBE gels (Invitrogen) to check for single-band products (Figure 6.5). Although re-design of probe-primer combinations was carried out after preliminary experiments had revealed doublets of bands (i.e. generation of two products), some doublets are still seen in the qPCR products using the final combination of probes and primers (Figure 6.5). Indeed, some probe-primers did not generate PCR products (Figure 6.5). These were probe-primer combinations using the following HindIII fragments as partner sequences: Cbr3 5', -315,

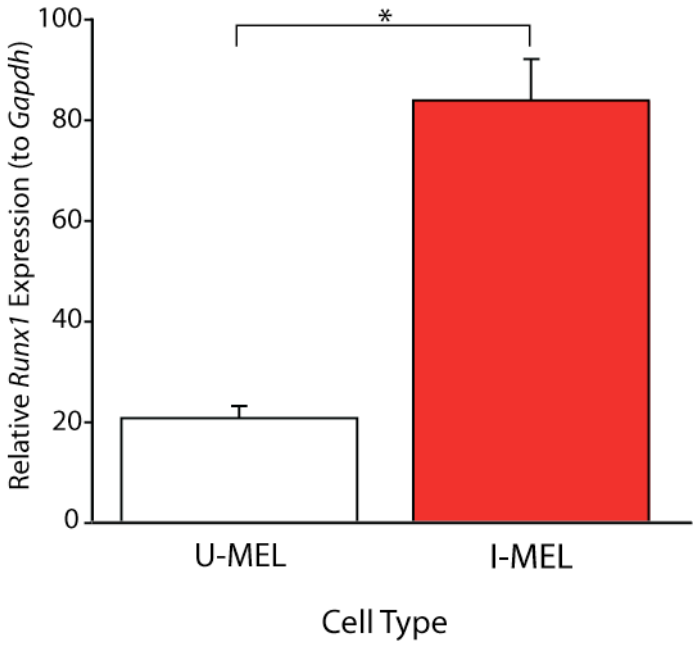


Figure 6.4. *Runx1* expression levels in U-MEL and I-MEL cells. qPCR was performed using cDNA reverse-transcribed from U-MEL and I-MEL cell RNA. The x-axis indicates the cell type, the y-axis is a scale of relative *Runx1* expression normalised to a *Gapdh* control. The relative expression of *Runx1* increases four-fold when cells are induced to become I-MEL cells (* p-value < 0.01, two-tailed Student t-test). Results are the average (± SD) of a total of four qPCR reactions (qPCR reactions repeated 2 times using 2 biological samples for each cell type).

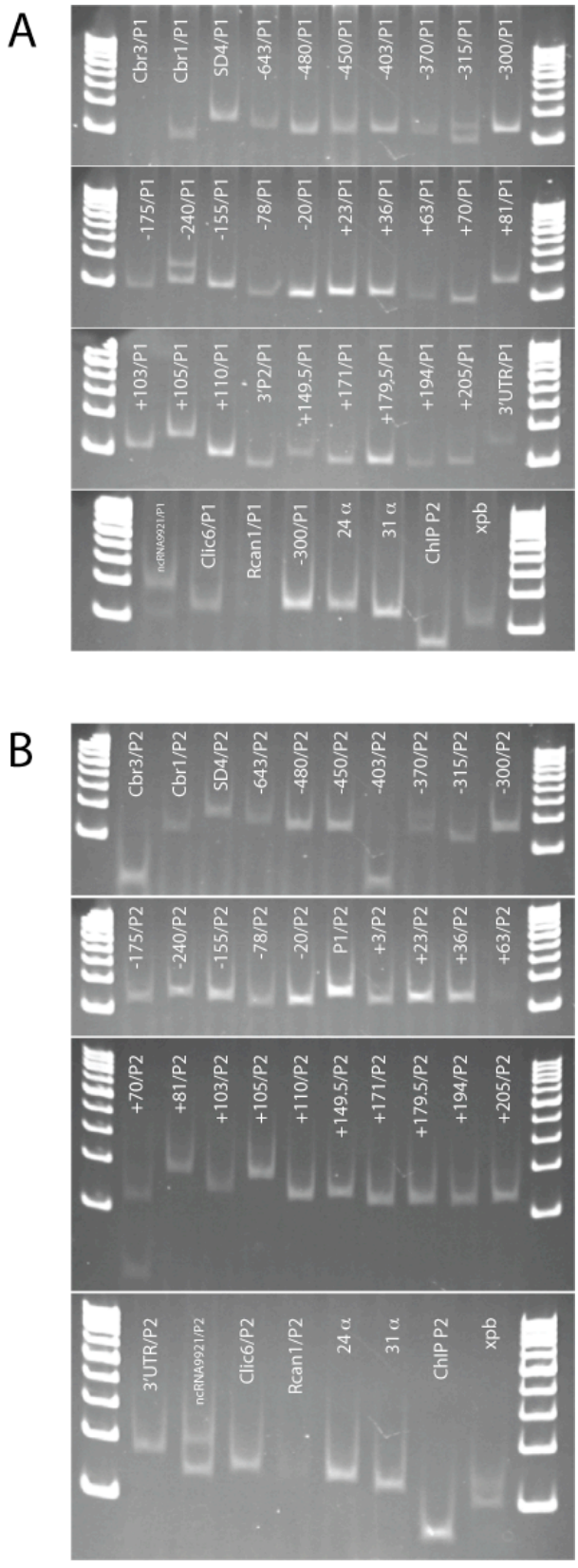


Figure 6.5. For legend, please see overleaf.

Figure 6.5. Acylamide gels of qPCR products from 3C experiments using the *Runx1* P1 and 3'P2 HindIII fragments as anchor sequences. (A) shows all P1 as anchor probe and primer combinations from an experiment using I-MEL cell material. The names of anchor probe and primer combinations (including controls in the last four lanes) are indicated above each lane. (B) shows all 3'P2 as anchor probe and primer combinations from an experiment using 416B cell material. The names of the probe and primer combinations are indicated as for A.

-240, +63, +70, ncRNA9921, and Rcan1 5'. These were therefore excluded from the final results.

6.4 Elements interacting with the P1 promoter HindIII fragment

Several elements were found to interact with the P1 promoter, some in a cell-type specific way. The results of the P1 anchor experiments represented as line graphs reveal five main peaks of interaction between the P1 HindIII fragment and HindIII fragments >40 Kb from the P1 (Figure 6.6). These peaks all have a fold enrichment of >2 and are: the -300, -78, +110, 3'P2 (+128) and +179.5. The xpb locus (which encodes a subunit of the basal transcription factor TFIIH) is assumed have a similar chromosome conformation in all cell types due to its ubiquitous role and is therefore used as an internal control to be sure that the assay has worked. These results are also shown to scale, as bar graphs in Figure 6.7 for the elements upstream of *Runx1* in the gene desert, and in Figure 6.8 for the elements within and downstream of the *Runx1* locus. The interaction between the P1 and the -300 is interesting for two reasons. Firstly, the distance between the elements is very large, making the chance of it being a random interaction highly unlikely (see figure 6.2). Secondly, the peaks of interaction are distinct only in the U- and I-MEL cells suggesting that these interactions may be erythroid-specific. This pattern is repeated with the +179.5 fragment where the interaction of the +179.5 with the P1 is also specific to U- and I-MEL cells. Interactions between the -78 and the P1 were seen in both J1 mES cells and I-MEL cells. Finally, interactions between the +110 and the P1 and the 3'P2 fragment and

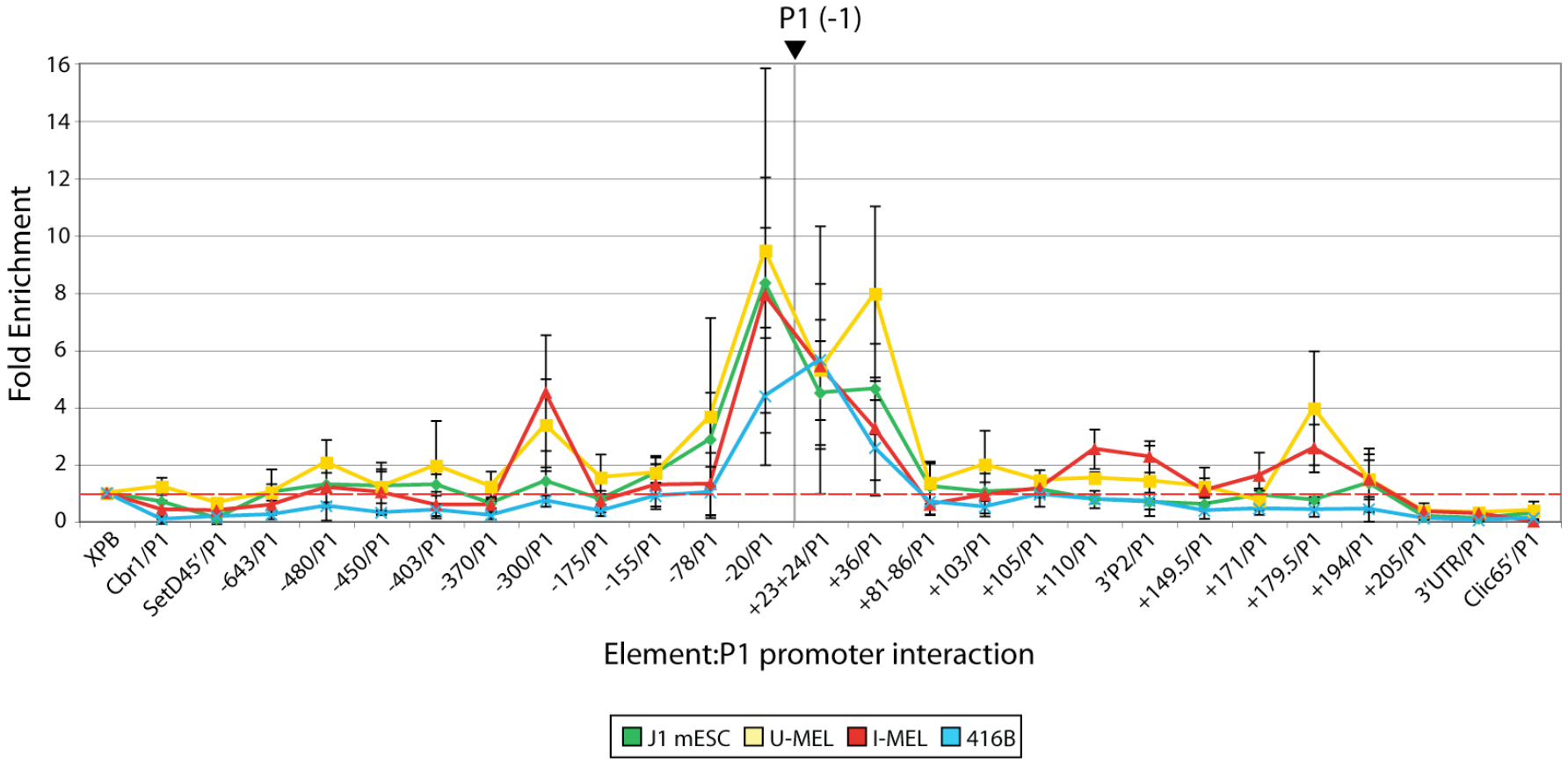


Figure 6.6 For legend please see overleaf.

Figure 6.6 Intrachromosomal interactions of putative and known *cis*-regulatory elements with the *Runx1* P1 promoter. Results are shown using 3C material prepared from the J1 mESC line (green lines), U-MEL cells (yellow lines), I-MEL cells (red lines) and 416B cells (blue lines). Results are the mean ± SD of at least 3 qPCR experiments using material from two separate 3C assays. The x-axis represents the interaction combination of a chosen element (partner) and the *Runx1* P1 promoter (anchor). Elements are numbered '+' or '-' in Kb from the *Runx1* translation start site, or by their genomic name. '3'UTR' represents the HindIII fragment containing the *Runx1* 3'UTR. The y-axis represents the fold enrichment of interaction of elements with the P1 promoter HindIII fragment (vertical grey line) after normalization to a genomic control (GAPDH) in comparison to the control xpb locus (fold enrichment of 1, red dashed horizontal line). Not to scale.

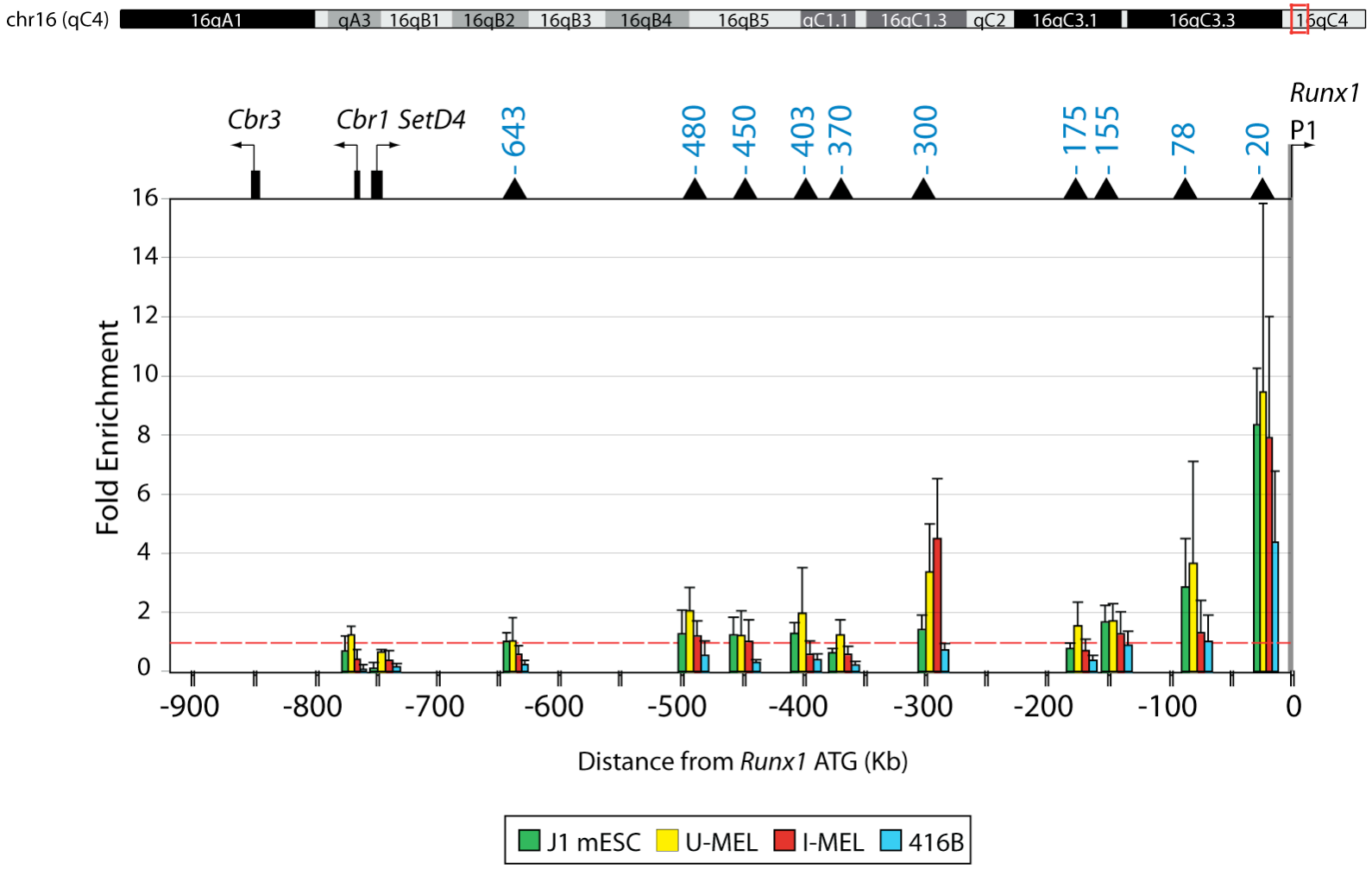


Figure 6.7. For legend please see overleaf.

Figure 6.7. Intrachromosomal interactions of putative *cis*-regulatory elements with the *Runx1* P1 promoter HindIII fragment. Results are shown using 3C material prepared from J1 mESC cells (green bars), U-MEL cells (yellow bars), and I-MEL cells (red bars) and 416B cells (blue bars). Results are the mean ± SD of at least 3 qPCR experiments using material from two separate 3C assays The x-axis represents the scale of the genomic region in Kb and the y-axis the fold enrichment of interaction of elements with the P1 promoter HindIII fragment (vertical grey line) after normalization to a genomic control (*Gapdh*) in comparison to the control xpb locus (fold enrichment of 1, red-dashed horizontal line). At the top of the figure is a chromosome ideogram indicating the region represented below (red box) which covers part of the *Runx1* locus (including the P1 and P2 promoters). The genes *Cbr3*, *Cbr1* and *SetD4* are shown upstream of *Runx1*, with the 3' limit of *SetD4 (-780)* marking the beginning of the gene desert, which continues up to the 5' limit of *Runx1*. The direction of transcription for each gene is indicated by the position of the promoters (arrows). Black triangles and blue numbers represent the location of putative regulatory elements.

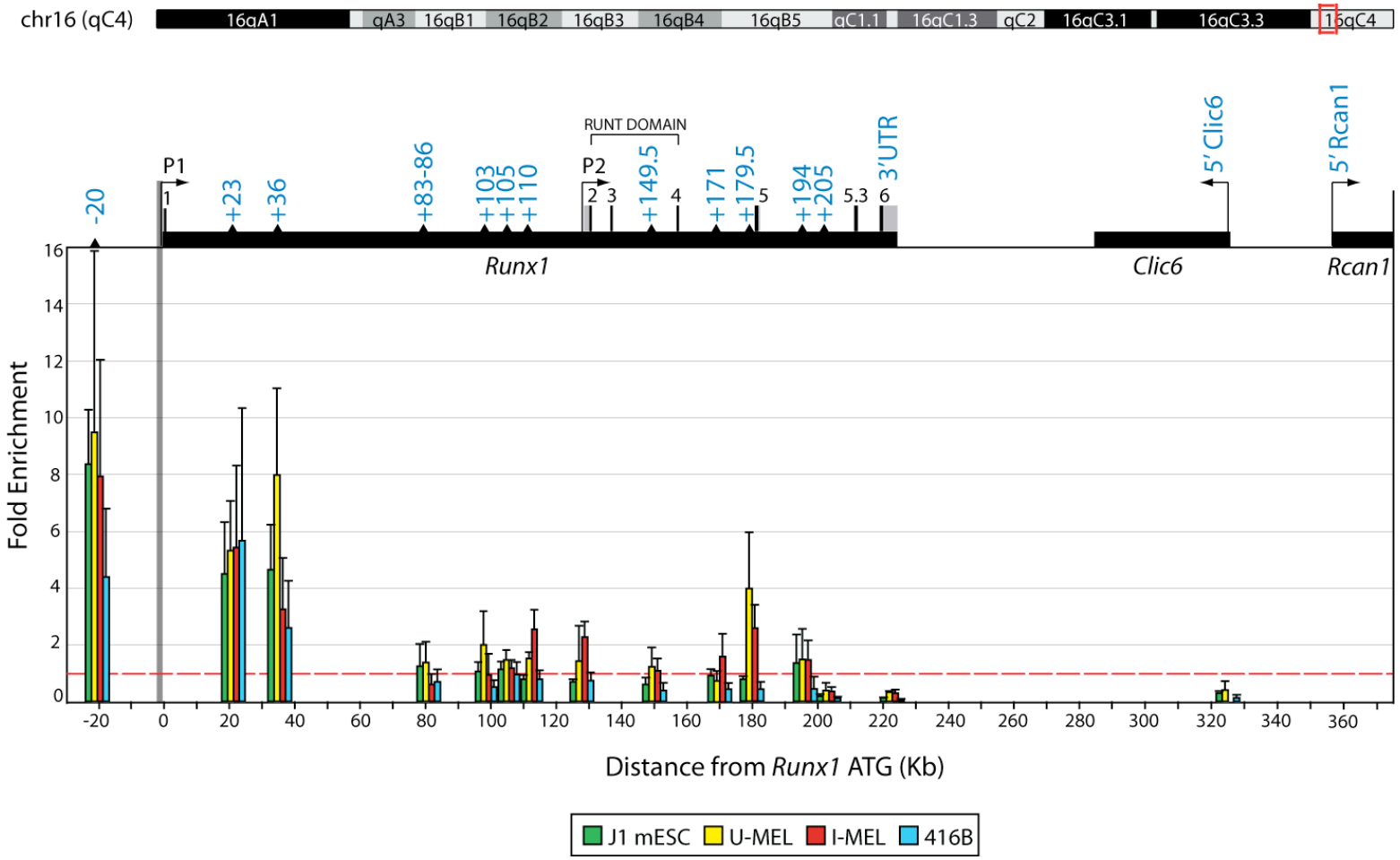


Figure 6.8. For legend please see overleaf.

Figure 6.8. Intrachromosomal interactions of putative *cis*-regulatory elements with the *Runx1* P1 promoter. Results are shown using 3C material prepared from J1 mESC cells (green bars), U-MEL cells (yellow bars), I-MEL cells (red bars) and 416B cells (blue bars). Results are the mean ± SD of at least 3 qPCR experiments using material from two separate 3C assays. The x-axis represents the scale of the genomic region in Kb and the y-axis the fold enrichment of interaction of elements with the P1 promoter HindIII fragment (vertical grey line) after normalization to a genomic control (*Gapdh*) in comparison to the control *xpb* locus (fold enrichment of 1, red-dashed horizontal line). At the top of the figure is a chromosome ideogram indicating the region represented below (red box). This region represents the *Runx1* locus, including the P1 and P2 promoters, exons (numbered black points) and UTRs (grey). The exons coding for the DNA-binding Runt domain are indicated. Black triangles and blue numbers represent the location of putative and known regulatory elements. The genes *Clic6* and *Rcan1* are represented downstream of *Runx1* with the direction of transcription indicated by their promoters (arrows). NB: the full length of *Rcan1* is not represented. The vertical grey line represents the position of the P1 anchor sequence with which other elements are interacting.

the P1 were only evident in I-MEL cells. Slightly less striking P1 interactions also appear with the -480 fragment and the -403 fragment in I-MEL cells. This may also be the case for the +103 element. Unfortunately, the proximity of the +23 to the P1 makes it impossible to say whether the strong interaction seen between these fragments is real. The same interaction would be highly likely to occur in the absence of chemical cross-linking, as illustrated in figure 6.2. Proximity effect may also explain the interaction seen between the P1 and the -20 and +36 fragments.

6.5 Interactions of elements with the 3'P2 promoter HindIII fragment

Preliminary results using U-MEL, I-MEL and 416B cells suggest that several elements are also interacting with the 3'P2 HindIII fragment. The results of the 3'P2 anchor experiments represented as line graphs, reveal five main peaks of interaction between the 3'P2 HindIII fragment and HindIII fragments >40 Kb from the 3'P2 (Figure 6.9). These peaks all have a fold enrichment of >2 and are: the -155, -20, the P1 promoter, the +3 and the +23/24 in U-MEL cells. As for the P1 anchor results, 3'P2 anchor results are also shown to scale, as bar graphs in Figure 6.10 for the elements upstream of *Runx1* in the gene desert, and in Figure 6.11 for the elements within and downstream of the *Runx1* locus. Most evident from these data, is the interaction between the P1 promoter fragment and the 3'P2 fragment, in all cell types. In addition, a U-MEL specific interaction occurs between the -450 enhancer HindIII fragment and the 3'P2. Proximity related interaction is evident using the 3'P2 as anchor as was the case for the P1 fragment. For

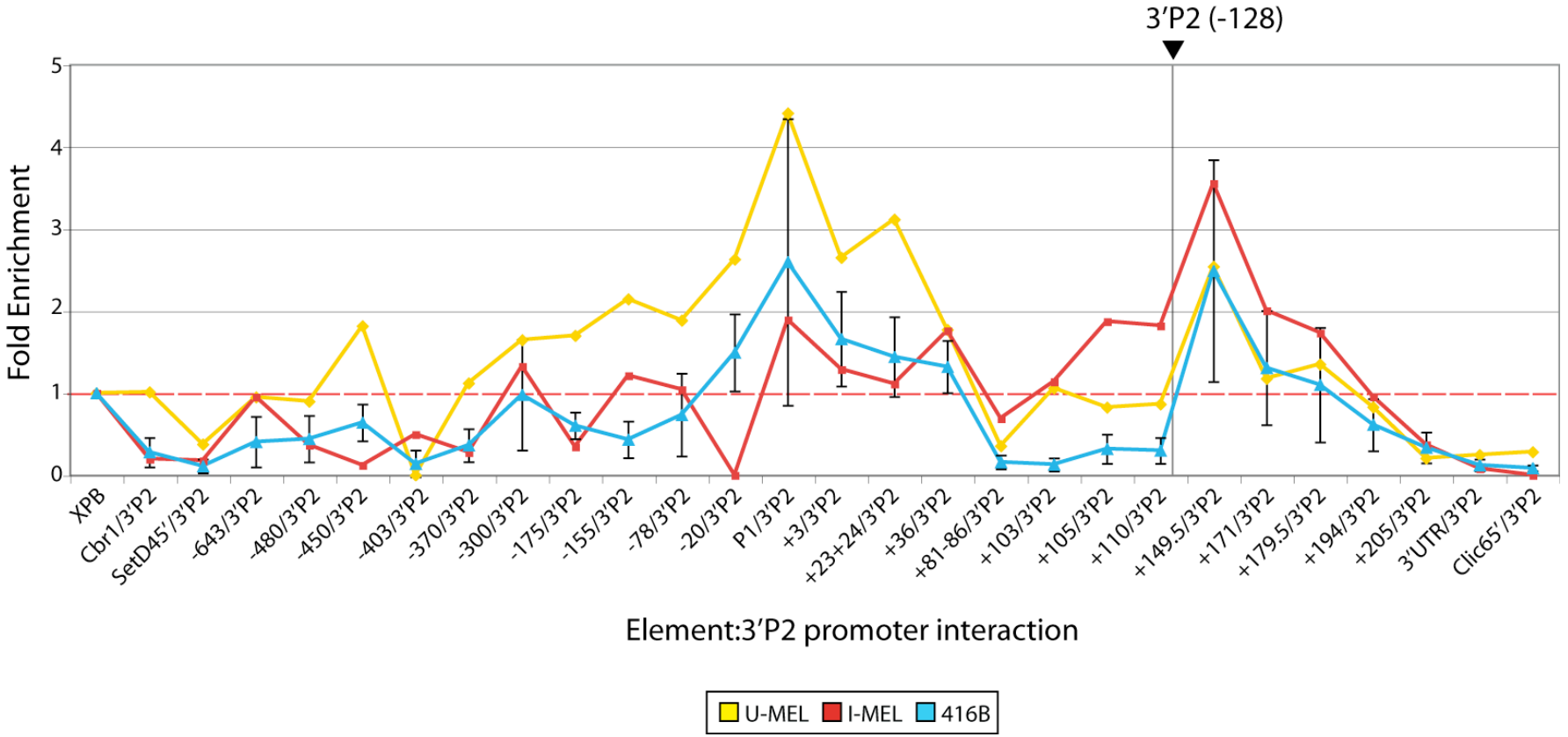


Figure 6.9. For legend, see overleaf.

Figure 6.9. Intrachromosomal interactions of putative and known *cis*-regulatory elements with the *Runx1* P2 promoter. Results are shown using 3C material prepared from the U-MEL cells (yellow bars, 1 x qPCR reaction), I-MEL cells (red bars, 1 x qPCR reaction) and 416B cells (blue bars, mean +SD of 3-4 x qPCR reactions). The x-axis represents the interaction combination of a chosen element (partner) and the *Runx1* 3'P2 promoter (anchor). Elements are numbered '+' or '-' in Kb from the *Runx1* translation start site, or by their genomic name. '3'UTR' represents the HindIII fragment containing the *Runx1* 3'UTR. The y-axis represents the fold enrichment of interaction of elements with the 3'P2 promoter HindIII fragment (vertical grey line) after normalization to a genomic control (GAPDH) in comparison to the control xpb locus (fold enrichment of 1, red dashed horizontal line). Not to scale.

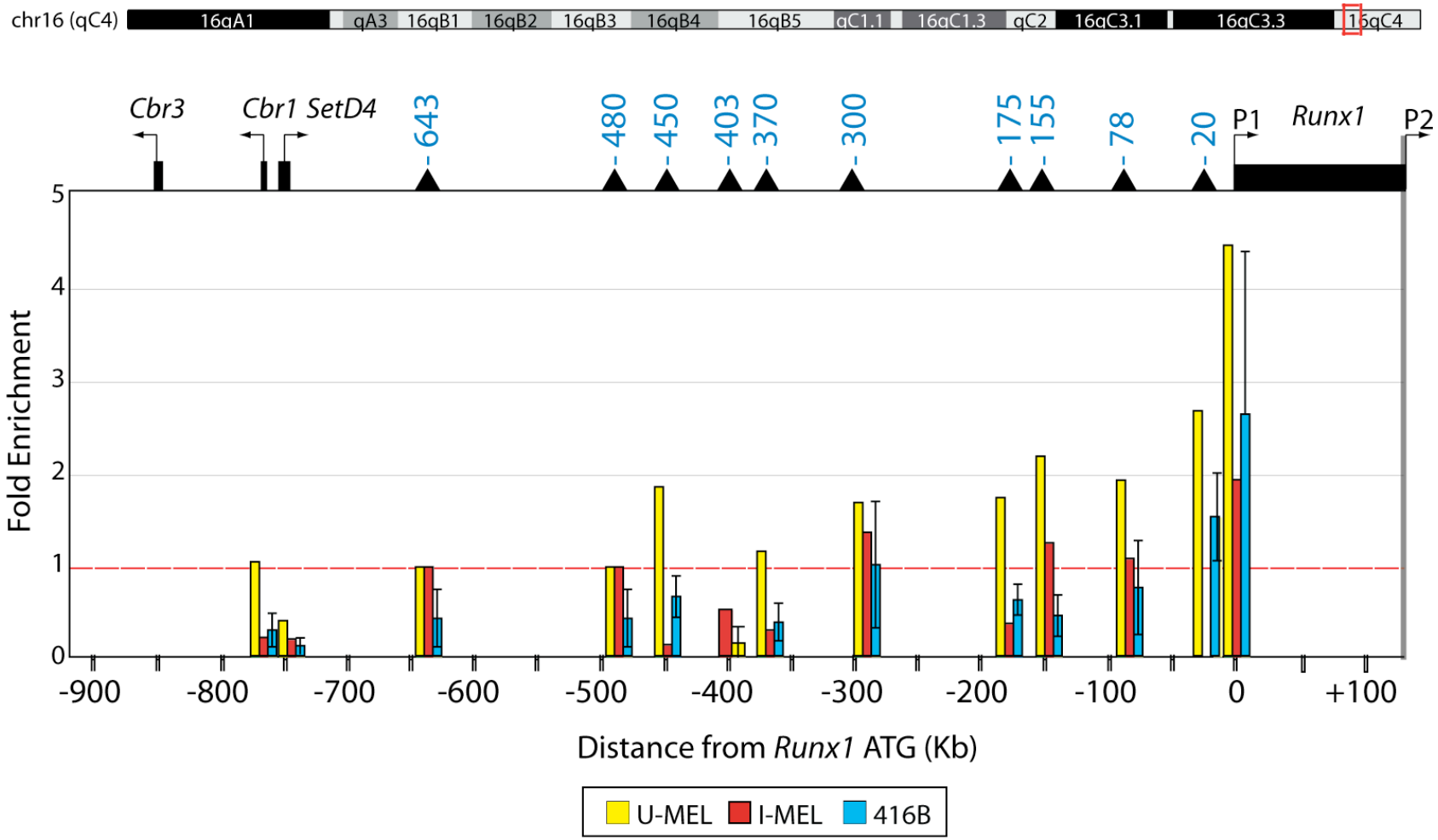


Figure 6.10. For legend please see overleaf.

Figure 6.10. Intrachromosomal interactions of putative *cis*-regulatory elements with the *Runx1* 3'P2 promoter HindIII fragment. Results are shown using 3C material prepared from U-MEL cells (yellow lines, 1 x qPCR reaction), I-MEL cells (red lines, 1 x qPCR reaction) and 416B cells (blue lines, +SD of 3-4 x qPCR reactions). The x-axis represents the scale of the genomic region in Kb and the y-axis the fold enrichment of interaction of elements with the 3'P2 promoter HindIII fragment (vertical grey line) after normalization to a genomic control (*Gapdh*) in comparison to the control *xpb* locus (fold enrichment of 1, red-dashed horizontal line). At the top of the figure is a chromosome ideogram indicating the region represented below (red box) which covers part of the *Runx1* locus (including the P1 and P2 promoters). The genes *Cbr3*, *Cbr1* and *SetD4* are shown upstream of *Runx1*, with the 3' limit of *SetD5* marking the beginning of the gene desert, which continues up to the 5' limit of *Runx1*. The direction of transcription for each gene is indicated by the position of the promoters (arrows). Black triangles and blue numbers represent the location of putative regulatory elements.

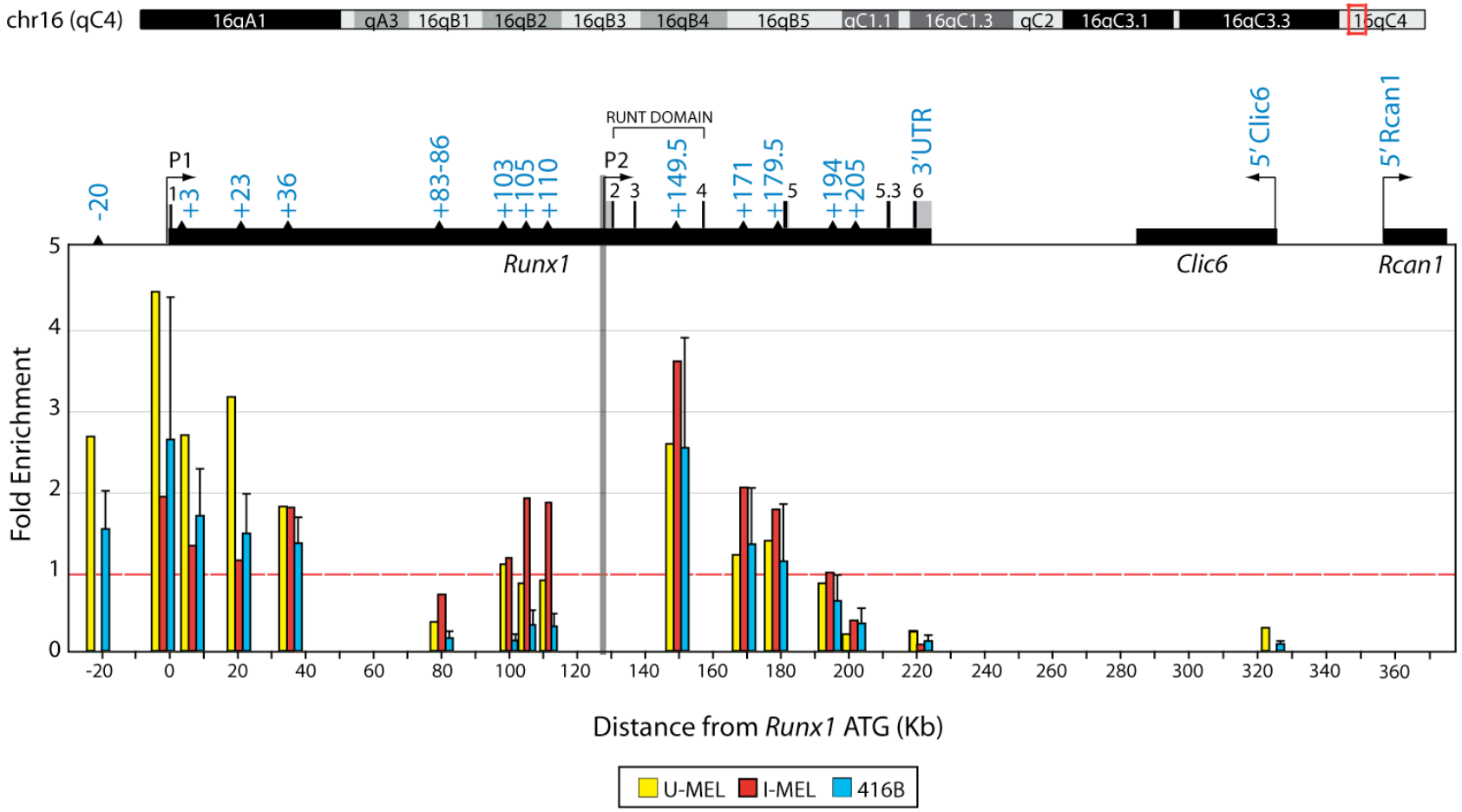


Figure 6.11. For legend please see overleaf.

Figure 6.11. Intrachromosomal interactions of putative *cis*-regulatory elements with the *Runx1* 3'P2 promoter HindIII fragment. Results are shown using 3C material prepared from U-MEL cells (yellow bars, 1 x qPCR reaction), I-MEL cells (red bars, 1 x qPCR reaction) and 416B cells (blue bars, mean +SD of 3-4 x qPCR reactions). The x-axis represents the scale of the genomic region in Kb and the y-axis the fold enrichment of interaction of elements with the 3'P2 promoter HindIII§ fragment (vertical grey line) after normalization to a genomic control (*Gapdh*) in comparison to the control *xpb* locus (fold enrichment of 1, red-dashed horizontal line). At the top of the figure is a chromosome ideogram indicating the region represented below (red box) which covers the *Runx1* locus, including the P1 and P2 promoters. The genes *Clic6* and *Rcan1* are represented downstream of *Runx1*. NB: the full length of *Rcan1* is not shown. The direction of transcription for each gene is indicated by the position of the promoters (arrows). Black triangles and blue numbers represent the location of putative and known regulatory elements.

example, the +149.5 element is just over 20 Kb downsteam of the 3'P2 fragment and appears to interact strongly with the 3'P2. A was the case for the +23 interaction with the P1 fragment this interaction cannot be classed as specific.

6.6 Chapter Discussion

The aim of this chapter was to use the 3C assay to investigate physical interactions between the *Runx1* P1 and P2 promoters and putative and known *cis*-regulatory elements in the *Runx1* syntenic region. Strong interactions were seen between the P1 and five elements within the *Runx1* syntenic region. This was also the case for the P2 and five elements within the *Runx1* syntenic region, which weren't all the same as those interacting with the P1. However, a common interaction was seen in both sets of experiments between the P1 and 3'P2 fragments regardless of which promoter was anchor/partner. This interaction was most evident when the 3'P2 was anchor, showing interaction with the P1 fragment in U-MEL, I-MEL and 416B cells. When the P1 was anchor, this interaction was only evident in I-MEL cells. These results seem surprising as the P2 does not generate transcripts in U-MEL or I-MEL cells, only the P1 does. Therefore, physical interaction between the P2 and the P1 in cells where the P2 is transcriptionally inactive, may highlight a role for the P2 in the *cis*-regulation of the P1, perhaps by bringing downstream regulatory elements into closer proximity with the P1.

These 3C studies have also highlighted interactions occurring in cell-specific ways. Firstly, in both the P1 and 3'P2 experiments, no interactions were seen between the promoters or any partner fragments in J1 mESCs. This was to be expected, as *Runx1* is mostly inactive in these cells, with P2-derived transcripts most likely representing leaky, basal transcription from the P2 promoter. Indeed, the P2 is bivalently marked in mES cells by both H3K4Me3 and H3K427Me3 histone modifications (Bernstein et al., 2006). Such bivalent marks have been found at several developmentally important genes in mES cells, and are thought to represent a 'poised' state where promoters are inactive, but are also ready to be transcribed when instructed to do so by the correct regulatory factors (Bernstein et al., 2006). The only interactions seen between the P1 or 3'P2 fragments and other partner fragments in J1 mESCs can be explained due to proximity to the anchor, including the interaction between the P1 and the +23 for example.

Interactions between the P1 fragment and the -300 fragment in U- and I-MEL cells is an example of a cell type specific interaction, as this interaction was absent in J1 mESC or 416B cells. The -300 element was originally identified due to its distinctive peak of H3K4Me2 enrichment in ChIP-chip assays using E12.5 FL cells (chapter 3, figure 3.8). Therefore, these 3C and ChIP-chip data would imply that the -300 could be an erythroid-specific *Runx1* enhancer. Preliminary experiments to test the *in vivo* activity of the -300 in E14.5 hsp68*LacZ*-300 F0 TGs however, did not reveal any X-gal staining in the FL of these embryos, which is a mostly erythroid cell population (not shown). It therefore seems unlikely that the -300 acts as an enhancer. However, it could

be that the -300 acts as a looping point or 'hinge' to bring gene desert elements into closer proximity with the P1, but further experiments need to be done to assess this claim.

With regards to *Runx1* enhancer elements characterized in chapter four (the -480, -450, -370, +3, +110 and +205), all enhancers apart from the -370 and the +205 interacted with either the P1 or 3'P2 fragments in at least one of the cell types tested. One reason for the lack of interaction between the promoters and the -370 or the +205 may be due to cell type specificity. However this is surprising for the +205 as it acts as an enhancer in 416B cells *in vitro* (Chapter three, Figure 3.11). The -370 mediates its functions in different cell types than those used in these experiments. Indeed, expression data from hsp68*LacZ*+205 F0 TGs in chapter four imply that the +205 enhancer is active in early embryos, in the hemogenic endothelium of the E8.5 YS and dorsal aortae, for example (chapter four, figure 4.10 B). If the +205 is a *Runx1* endothelial enhancer, then using an endothelial cell line such as MS1 for future 3C experiments may be informative in demonstrating this (although MS1 still doesn't represent a hemogenic endothelial cell line). Data regarding the -370 enhancer are less clear-cut, with *LacZ* expression only having been found in a few cells in the YS of E11.5 hsp68*LacZ*-370 F0 TGs. Like the +205, the -370 enhancer may be active in other cell types.

Finally, these 3C experiments have identified three new putative regulatory elements, including the -78 and the -155 elements in the gene desert and the +179.5 in the *Runx1* locus. The +179.5 is enriched for binding of several Ets

TFs, including Fli1 and Erg in HPC7 cells (Wilson et al., 2010) and was also selected to test in 3C assays as it was identified as a potential CNE in the Gumby analysis of the 270 Kb *Runx1* fragment (Wade Nottingham, unpublished). Testing these elements for *in vitro* and potentially *in vivo* enhancer activity will therefore be interesting future experiments.

In conclusion, the 3C technique has proved useful in identifying interactions that occur between the *Runx1* P1 and P2 promoters and selected elements in the *Runx1* syntenic region. Additional experiments are needed to extend and confirm the results shown, in particular by moving on to 4C studies. Currently the cell numbers required are prohibitive to 3C/4C experiments using small cell populations from embryonic tissues. However, as 3C/4C technology advances, adapting these studies to specific subsets of cells (e.g hemogenic endothelial vs. commited hematopoietic cells) of the embryo, should help us understand differences in the requirements of certain *cis*-regulatory elements for *Runx1* transcription in these cells. Such data would contribute to our ultimate aim of delineating the upstream inputs that activate the formation of hematopoietic stem and progenitor cells.

Chapter Seven:

General discussion and future directions

7.1 Summary of results

The aim of this thesis has been to identify and characterize *Runx1* enhancers with the long-term aim to further understand what regulates the emergence and commitment of hematopoietic cells during embryonic development. In chapter three, a combination of chromatin assays was used to identify putative *Runx1* enhancers in the *Runx1* syntenic region. This resulted in the identification of eleven candidate genomic regions of which five had *in vitro* enhancer activity in luciferase assays. In chapter four, seven elements were tested for *in vivo* enhancer activity using reporter gene F0 TG embryos. These seven elements were found to mediate *Runx1*-specific expression to specific aspects of developmental hematopoiesis. In chapter five, characterization of cells in which the +110 enhancer is active was initiated using a novel hsp68*GFP*+110 mouse line. Experiments showed that the +110 enhancer is active during the primitive and transient definitive waves of YS hematopoiesis in E8 hsp68*GFP*+110 ln5 TG embryos. Also, during the adult definitive wave in E11.5 embryos, the +110 enhancer targets the majority of clonogenic progenitors in the AVU, FL and YS but, interestingly, not HSCs. In chapter six, physical interactions between putative and known *cis*-regulatory elements and the P1 and P2 promoters were characterized by 3C. Several interactions were identified, including an interaction between the P1 and P2 promoters in all cell types tested.

7.2 Which *Runx1* enhancers are required for *Runx1* expression?

Data from this thesis show that different aspects of *Runx1* expression are mediated by several enhancer elements during developmental hematopoiesis. This is the first time since the identification of the *Runx1* +23-hematopoietic specific enhancer that other hematopoietic *Runx1* enhancers have been characterized, adding to our understanding of the complexity of *Runx1 cis*-regulation. Interestingly, regulation of gene expression by multiple enhancers has been reported for other genes involved in developmental hematopoiesis. For example, three enhancers mediate distinct aspects of *SCL* expression during mouse development (Gottgens et al., 2011). The -4 enhancer mediates endothelial-specific *SCL* expression (Sinclair et al., 1999), the +19 enhancer is active in HSCs and progenitors (Sanchez et al., 1999; Silberstein et al., 2005) and the +40 enhancer mediates erythroid-specific expression (Delabesse et al., 2005). The *cis*-regulation of genes involved in the development of other organ systems can also be regulated by multiple enhancer elements. This is the case for the gene *Myf5* for example, which plays an important role in skeletal muscle formation (Zammit et al., 2004). In conclusion therefore, such modular, *cis*-regulatory control seems to be a recurring theme for developmental TFs with complex spatiotemporal expression patterns and likely represents a means of tightly regulating or controlling these genes as and when needed.

In vivo analysis of *Runx1* enhancers in this thesis has shown that they are sufficient to mediate reporter gene expression to specific aspects of

endogenous *Runx1* expression *in vivo*, but it is unknown which of these enhancers are *necessary* for *Runx1* expression during developmental hematopoiesis. Future experiments using mES cells and mice in which enhancers have been deleted from the genome, will demonstrate any requirement of those enhancers for normal *Runx1* gene expression. Germline deletion of the +23+24 enhancer in mES cells is ongoing in the lab (Ana Santos, Joey Riepsaame). Interestingly, preliminary experiments suggest that *Runx1* expression may be negatively affected by heterozygous deletion of the +23+24 enhancer in mES cells (Ana Santos, unpublished observations), but this needs to be confirmed. It will also be important to test which enhancer, or combination of enhancers, is able to rescue the *Runx1* null hematopoietic phenotype, where embryos die at around E12.5 with FL anemia. Work by Ana Santos in the lab has already shown that full-length Runx1 cDNA transcribed under the control of the +23+23 enhancer, is able to part rescue FL hematopoiesis and delay the embryonic lethal phenotype to E18 in compound transgenic embryos. These embryos lack endogenous Runx1 but express full length *Runx1* cDNA under the control of the +23+24 enhancer. Extended survival of the embryos to E18 is conferred by a partial rescue of clonogenic progenitors, which are absent in *Runx1* null embryos (Ana Santos, unpublished). In conclusion therefore, work is underway to assess the importance of *Runx1* enhancers in the control of Runx1 expression and therefore the emergence of HSCs and progenitors *in vivo*.

7.3 What controls the distinct activities of *Runx1* enhancers?

Understanding what causes different enhancers to mediate specific aspects of *Runx1* expression is not yet understood for the enhancers identified in this study. A deletion series of the +23 enhancer into three sequence fragments showed that its enhancer activity was confined to a central 235 bp fragment (Nottingham et al., 2007). Deletion series of the +3, +110 and +205 enhancer sequences are currently being done in the lab to characterize the regions critical for their function (Talia Galeadi). In addition, it will be important to identify sequence motifs of putative binding sites for upstream TFs within the enhancers. Analysis of the +23 enhancer sequence revealed conserved putative binding motifs for GATA factors as well as for Runx, Scl (E-box motif), Myb, and three Ets factor motifs, amongst others (Nottingham et al., 2007). Targeted mutagenesis of these sites showed that +23 enhancer activity decreases significantly when the all three Ets sites, the Runx site and also the Gata site are mutated respectively, in *in vitro* luciferase assays. In *in vivo* experiments however, +23 enhancer activity was only diminished in F0 TG embryos that had mutated GATA or Ets sites. Mutation of the Runx motif had no effect on +23 activity *in vivo* (Nottingham et al., 2007). The +23 acts downstream of Gata2, Ets factors, and Scl which are all implicated in the specification of the hemangioblast cell type in *Xenopus* and zebrafish embryos (Liu et al., 2008). In addition, these TFs also constitute a recursively wired triad of TFs at the onset of hematopoiesis (Pimanda et al., 2007b), which is in line with the +23 enhancer being active in the initiation of *Runx1* at the onset of hematopoiesis.

Initial analysis of the +3, +110 and +205 enhancer sequences shows differences in the putative binding motifs that they contain. The +3 enhancer sequence contains conserved GATA, Ets and Runx motifs, which may be the reason that the +3 mediates reporter gene expression in a pattern similar to the +23 (i.e. at all sites of hematopoiesis in the embryo). The +110 enhancer is highly active in primitive erythrocytes in the E8 YS and mediates strong *GFP* expression to the FL of E14.5 hsp68*GFP*+110 ln5 TG embryos (a mostly erythroid population). The +110 also targets clonogenic progenitor cells in the AVU, FL and YS of E11.5 hsp68*GFP*+110 ln5 TG embryos. It is interesting therefore that the +110 enhancer sequence has multiple conserved GATA motifs, but no conserved Runx or E-box motifs. The TFs Gata1 and Gata2 are both expressed in the YS at E8 and are critical for primitive erythroid cell differentiation (Palis et al., 1999; Silver and Palis, 1997). Gata2 is also essential for the maintenance of erythroid cell numbers and the survival and proliferation of multipotent hematopoietic progenitors (Tsai et al., 1994). Whether Gata1 and Gata2 influence these processes through +110-mediated *Runx1* expression will be of interest for future experiments. Finally, the +205 enhancer sequence contains multiple Ets motifs in the 3' end of the sequence spanning the two hematopoietic DNAseI hypersensitive sites. Members of the Ets family of TFs are known to be key regulators of endothelial genes (Meadows et al., 2011). As the +205 targets *LacZ* expression to a sub-set of endothelial cells in the PAS region of E8.5 hsp68*LacZ*+205 F0 TGs, it may be that expression to these endothelial cells is regulated by binding of Ets TFs to these binding motifs. Mutagenesis studies, like those performed for the +23 enhancer, will be undertaken to

assess the importance of these motifs to the expression patterns mediated by these novel enhancers.

7.4 How many more *Runx1 cis*-regulatory elements are there?

Chromatin assays used to identify putative enhancers in this study revealed many regions within the *Runx1* syntenic region with putative enhancer activity based on epigenetic modifications, DNaseI hypersensitivity and TF binding data. In addition, 3C experiments have revealed three elements, not yet tested for enhancer activity, that interact with both P1 and P1 promoters (-78 and +179.5) or the P2 promoter (-155). New techniques including the 4C assay will help catalogue all interactions that occur between the *Runx1* P1 and P2 promoters and other regions of the genome in *cis*. This technique differs from 3C as ligated products are sequenced using primers on both edges of the anchor sequence. This inverse PCR technique reads into any sequence with which the anchor is interacting. Large-scale sequencing is then used to reveal the nature of interacting partner sequences. Using this technique in a range of cell types representing different types of hematopoietic cells will allow us to identify all *cis*-regulatory elements that function via interaction with the *Runx1* P1 and P2 promoters. It should be noted that these elements may include silencers, insulators and boundary elements, not just enhancers. In conclusion, identifying the full complement of *Runx1* hematopoietic enhancers will lead us to identify the full complement of TFs that converge on *Runx1* to regulate its expression and hence HSC emergence. By doing this in a 'developmental sequence' of cell lines and tissues we can aim to identify

the order in which these TFs act and therefore understand the order of cues that are needed to make blood.

7.5 *Runx1* enhancers as experimental tools

Two transgenic mouse lines generated in our lab have been studied in detail: i) hsp68*GFP*+23+24 ln3 and, ii) hsp68*GFP*+110 ln 5. A third hsp68*GFP*+205 line is currently being established (Chrissy Rode). The hsp68*GFP*+23+24 ln3 has already been instrumental in visualizing the requirement for the P1 and P2 promoters in hematopoietic cluster formation in the DA of E11.5 embryos (Bee et al., 2010). The hsp68*GFP*+23+24 ln3 has also been used recently in studies to establish when and where hematopoietic commitment occurs in the developing embryo. In the latter example, sorted +23+24-targeted cells from different aged TG embryos, have been used in co-culture assays and for analyis of gene expression, to establish the point in development when endothelial cells targeted by the +23-+24 enhancer begin to take on hematopoietic identity (Gemma Swiers, manuscript in preparation). These mouse models are also amenable to confocal microscopy studies where sections of TG embryos can be analysed for overlap of enhancer-mediated *GFP* expression and expression of cell specific antibodies, such as VE-Cadherin and CD45. This is a good visual experiment to distinguish between cell types in which the enhancers are active. FACS experiments to determine the phenotype of enhancer-targeted, GFP^+ cells, is also a powerful use of these mouse models, as shown in chapter five and (Bee et al., 2010). In conclusion, the use of enhancer-reporter transgenic mouse

lines to study the nature of cells to which *Runx1* enhancers mediate reporter gene expression have already proved vital in our understanding of *Runx1*'s role in developmental hematopoiesis.

Having these mouse models provides huge scope for future experiments aimed at characterizing the nature of cells targeted by the +110 and +205 in particular. It would also be interesting to generate transgenic mouse lines where an enhancer mediates expression to a reporter gene who's protein product fluoresces at a different wavelength to GFP e.g RFP. Crossing transgenic mice heterozygous for different reporter transgenes, would generate an overlap in expression patterns allowing us to determine cells in which both enhancers are active and those which are enhancer-specific. These experiments would help build a picture of the enhancers required to mediate the total *Runx1* expression pattern during development and ultimately aid in refining the cellular hierarchy of developing hematopoietic system.

7.6 Concluding remarks

Before this study, the +23 enhancer was the only hematopoietic-specific *Runx1* enhancer characterized. This study has used a combination of *in vitro* screening assays and *in vivo* transgenic experiments to identify seven new *Runx1* enhancers. These enhancers mediate different aspects of *Runx1* expression during developmental hematopoiesis and highlight new levels of complexity in *Runx1 cis*-regulation.

Appendix I.

ChIP-chip array formats

Arrays were designed in two formats using pre-existing Agilent probes. The first format was printed with 2 x 205K tiled arrays (i.e 2 duplicate arrays per slide) each containing the same probe coverage of the *Runx1, Runx2, Runx3, CBFβ* and *α-globin* genes and flanking regions (Table 3.1). The second format was a 4 x 44K tiled array (i.e. 4 replicate arrays per slide). These arrays were smaller in size and covered *Runx1* and *α-globin* coordinates only (Table 3.2). For both formats, the *α-globin* locus was included as a positive control.

Appendix II.

Shown below are the positions and sequences of HindIII fragments and Taqman probes and primers used in 3C experiments. Table A shows putative and known regulatory elements according to their distance from the *Runx1* ATG (in Kb). For each element, the start and end coordinate of the HindIII fragment within which the element is located and the size of this HindIII fragement is given. The position of the Taqman forward (F) primer, its sequence and the strand to which it binds are given. Table B shows the coordinates and sequences of the P1 and 3'P2 anchor HindIII fragment reverse primers and probes.

A.

	HindIII Fragment		Fragment Size (Kb)	Partner: F Primer		Partner Primer Sequence	Strand
Cbr3 Small	93,684,933	93,685,793	0.861	93684952	93684975	GCGTCTCAAGAACACTCAGTTTTG	-
Cbr1	93,606,483	93,610,951	4.468	93610920	93610941	ACATCAGGAGCCAAAGCAGATT	+
SetD4_5'	93,602,322	93,604,696	2.374	93604614	93604641	AGTACGAATTAGGATTATGGGAAGAAAT	+
-643	93,454,774	93,465,470	10.697	93465404	93465430	TTCATCAGAAGTAACTTTGGAGACATC	+
-480	93,310,009	93,313,338	3.33	93313287	93313306	GCCTTGCTGGAGGAAGAATG	+
-450	93,267,312	93,272,884	5.573	93267344	93267366	ACATAAGCAAGGACGAGGAGAGA	-
-403	93,236,464	93,241,963	5.499	93241899	93241924	CTTCACACGTTGTCTATCTCGTTTTT	+
-370	93,195,535	93,202,157	6.623	93202100	93202126	TTTGATTTACATATGCTCTGAGACACA	+
-315	93,173,880	93,181,817	7.938	93173891	93173917	GACAGTAGAGAACAGAGTTTGCAGACA	-
-300	93,125,159	93,132,683	7.525	93132619	93132644	AAGCTAACAGATTTTCCCAGAGAGTT	+
-175	93,000,694	93,002,614	1.920	93002579	93002607	TGTAACCCTTATACCACAAATATCACCTT	+
-240	93,062,101	93,064,769	2.669	93062128	93062152	CTCAAGGAAGACTTCCAGAGTAAGC	-
-155	92,978,793	92,981,048	2.256	92981005	92981030	ACCGAGTGATTCATTTAGCTACAGAA	+
-78	92,904,393	92,909,119	4.727	92909090	92909111	ATTAGTTCCCCACCTGTTCCAA	+
-20	92,842,322	92,851,876	9.555	92842330	92842353	CAGCAGAAATGGTAACCACATCTT	-
P1 Promoter (-1)	92,824,140	92,826,881	2.402	92824171	92824196	GCATCAGGATTTAAAGAGAAGAGGAA	-
+3	92,817,586	92,824,145	6.560	92817596	92817618	TGGACTGGTCTTGTTGTTCAGAA	-
+23/24	92,798,784	92,802,366	3.583	92798800	92798822	CACTCATGCTTTGGGTCAATGTT	-
+24	-	-	-	-	-	-	
+36	92,784,325	92,792,818	8.493	92792781	92792807	CCCACCAAAATGATTAAGGTATTCTG	+
+63	92,757,181	92,766,923	9.742	92766887	92766912	CTATTTTTAATAAGCCGTTCGTTTCC	+
+70	92,750,477	92,755,138	4.662	92750485	92750505	GCATCGAATGCTAACGCAGAA	-
+81/86	92,736,362	92,744,797	8.436	92736369	92736390	AGCATCGTTTGTGTCAATAGTATGAAC	-
+103	92,722,114	92,722,737	0.624	92722686	92722707	GAGCCTGGGATGGGAGAGTAAT	+
+105	92,717,787	92,721,253	3.467	92717846	92717870	ACTGCATGTGGGTGTTATTGTAAAC	-
+110	92,716,098	92,717,792	1.695	92716111	92716139	GAGAGGAATGATAAAACTGTGTCTGATAA	-
3'P2 Promoter (+128)	92,692,097	92,694,501	2.405	92694475	92694494	GCTGAGGCCTGGAACTACGA	+
+149.5	92,675,139	92,677,815	2.677	92675164	92675185	GCCCTCTCCCTATGTACTTCCA	-
+171	92,650,777	92,658,583	7.807	92650789	92650811	GGCATCCCATGAAGTTGTTAAGA	-
+179.5	92,637,419	92,646,084	8.666	92637434	92637457	GGAGGAGGGAGAGAATTTTTAACG	-
+194	92,627,697	92,634,257	6.560	92634222	92634244	TGGCAACATATGGTGGCTCTTAG	+
+204	92,612,580	92,626,885	14.306	92612592	92612619	GAGACATGTGAGTATCCAGACTAGTTGA	-
3' UTR	92,600,902	92,608,017	7.116	92600962	92600986	CCCTGTCCAAACTGTTACTAGCATT	-
+254	92,564,577	92,575,265	10.688	92575220	92575246	AATGCATGCTAATGTATATGTGACCAT	+
Clic6_5'	92,492,320	92,501,876	9.557	92501820	92501843	CTATTGTTCCAAGGTGGTTGGTTA	+
Rcan1_5'	92,464,988	92,470,567	5.579	92470524	92470548	TCTAGTCACTGGAACCCTTCACTGT	+

B.

	Anchor Primer		Strand	Anchor Primer Sequence	Anchor Probe		Probe sequence
P1 Promoter	92826819	92826839	+	CTTTTCCGTGAGTGGCTGATC	92826844	92826870	TGTCAGTAGGTTGCCAGCCTAGCCCTC
3'P2 Promoter	92694419	92694438	+	CCGGACAGCACTTGTCAAAA	92694467	92694493	CGTAGTTCCAGGCCTCAGCTGGATCTT

Appendix II. (A) Positions of 3C partner sequence HindIII fragments (UCSC, mm9) and of Taqman forward primer sequences within the HindIII fragment. (B) Positions of the P1 and 3'P2 HindIII anchor fragments and of Taqman primer and probe sequences within the HindIII fragment.

References

Adolfsson, J., Mansson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.J., Thoren, L.A., *et al.* (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell *121*, 295-306.

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193-197.

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

Bae, S.C., Yamaguchi-Iwai, Y., Ogawa, E., Maruyama, M., Inuzuka, M., Kagoshima, H., Shigesada, K., Satake, M., and Ito, Y. (1993). Isolation of PEBP2 alpha B cDNA representing the mouse homolog of human acute myeloid leukemia gene, AML1. Oncogene *8*, 809-814.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bee, T., Ashley, E.L., Bickley, S.R., Jarratt, A., Li, P.S., Sloane-Stanley, J., Gottgens, B., and de Bruijn, M.F. (2009a). The mouse Runx1 +23 hematopoietic stem cell enhancer confers hematopoietic specificity to both Runx1 promoters. Blood.

Bee, T., Liddiard, K., Swiers, G., Bickley, S.R., Vink, C.S., Jarratt, A., Hughes, J.R., Medvinsky, A., and de Bruijn, M.F. (2009b). Alternative Runx1 promoter usage in mouse developmental hematopoiesis. Blood Cells Mol Dis.

Bee, T., Swiers, G., Muroi, S., Pozner, A., Nottingham, W., Santos, A.C., Li, P.S., Taniuchi, I., and de Bruijn, M.F. (2010). Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. Blood *115*, 3042-3050.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

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

Bertrand, J.Y., Giroux, S., Golub, R., Klaine, M., Jalil, A., Boucontet, L., Godin, I., and Cumano, A. (2005). Characterization of purified intraembryonic hematopoietic stem cells as a tool to define their site of origin. Proc Natl Acad Sci U S A *102*, 134-139.

Bertrand, J.Y., and Traver, D. (2009). Hematopoietic cell development in the zebrafish embryo. Curr Opin Hematol *16*, 243-248.

Bigas, A., Robert-Moreno, A., and Espinosa, L. (2010). The Notch pathway in the developing hematopoietic system. Int J Dev Biol *54*, 1175-1188.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., *et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Boffelli, D., Nobrega, M.A., and Rubin, E.M. (2004). Comparative genomics at the vertebrate extremes. Nat Rev Genet *5*, 456-465.

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

Brudno, M., Chapman, M., Gottgens, B., Batzoglou, S., and Morgenstern, B. (2003). Fast and sensitive multiple alignment of large genomic sequences. BMC Bioinformatics *4*, 66.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327-339.

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

Cai, Z., de Bruijn, M., Ma, X., Dortland, B., Luteijn, T., Downing, R.J., and Dzierzak, E. (2000). Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. Immunity *13*, 423-431.

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

Chubb, J.R., and Bickmore, W.A. (2003). Considering nuclear compartmentalization in the light of nuclear dynamics. Cell *112*, 403-406.

Ciau-Uitz, A., Pinheiro, P., Gupta, R., Enver, T., and Patient, R. (2010). Tel1/ETV6 specifies blood stem cells through the agency of VEGF signaling. Dev Cell *18*, 569-578.

Ciau-Uitz, A., Walmsley, M., and Patient, R. (2000). Distinct origins of adult and embryonic blood in Xenopus. Cell *102*, 787-796.

Corbel, C., Salaun, J., Belo-Diabangouaya, P., and Dieterlen-Lievre, F. (2007). Hematopoietic potential of the pre-fusion allantois. Dev Biol *301*, 478-488.

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

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

Cumano, A., and Godin, I. (2007). Ontogeny of the hematopoietic system. Annu Rev Immunol *25*, 745-785.

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

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

De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. Blood *110*, 4503-4510.

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods *3*, 17-21.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Delabesse, E., Ogilvy, S., Chapman, M.A., Piltz, S.G., Gottgens, B., and Green, A.R. (2005). Transcriptional regulation of the SCL locus: identification of an enhancer that targets the primitive erythroid lineage in vivo. Mol Cell Biol *25*, 5215-5225.

Dexter, T.M., Scott, D., Teich, N.M. (1978). Infection of bone marrow cells in vitro with FLV: effects on stem cell proliferation, differentiation and leukemogenic capacity. Cell *12*, 355-364.

Dexter, T.M., Allen, T.D., Scott, D., and Teich N.M. (1979). Isolation and characterisation of a bipotential haematopoietic cell line. Nature (London) *277*, 471-474.

Dieterlen-Lievre, F. (1998). Hematopoiesis: progenitors and their genetic program. Curr Biol *8*, R727-730.

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

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

Erickson, P., Gao, J., Chang, K.S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J., and Drabkin, H. (1992). Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. Blood *80*, 1825-1831.

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

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Res *32*, W273-279.

Friend, C., Scher W., Holland., J.G., and Sato., T. (1971). Hemoglobin Synthesis in Murine Virus-Induced Leukemic Cells In Vitro: Stimulation of Erythroid Differentiation by Demethyl Sulfoxide. Proc Natl Acad Sci U S A *68*, 378-382.

Gekas, C., Dieterlen-Lievre, F., Orkin, S.H., and Mikkola, H.K. (2005). The placenta is a niche for hematopoietic stem cells. Dev Cell *8*, 365-375.

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

Ghozi, M.C., Bernstein, Y., Negreanu, V., Levanon, D., and Groner, Y. (1996). Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions. Proc Natl Acad Sci U S A *93*, 1935-1940.

Godin, I., and Cumano, A. (2002). The hare and the tortoise: an embryonic haematopoietic race. Nat Rev Immunol *2*, 593-604.

Goldfarb, A.N. (2007). Transcriptional control of megakaryocyte development. Oncogene *26*, 6795-6802.

Goldfarb, A.N. (2009). Megakaryocytic programming by a transcriptional regulatory loop: A circle connecting RUNX1, GATA-1, and P-TEFb. J Cell Biochem *107*, 377-382.

Gottgens, B., Ferreira, R., Sanchez, M.J., Ishibashi, S., Li, J., Spensberger, D., Lefevre, P., Ottersbach, K., Chapman, M., Kinston, S.*, et al.* (2010). cis-Regulatory remodeling of the SCL locus during vertebrate evolution. Mol Cell Biol *30*, 5741-5751.

Haar, J.L., and Ackerman, G.A. (1971). A phase and electron microscopic study of vasculogenesis and erythropoiesis in the yolk sac of the mouse. Anat Rec *170*, 199-223.

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

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone

modifications at human enhancers reflect global cell-type-specific gene expression. Nature.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

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

Hughes, J.R., Cheng, J.F., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., De Gobbi, M., de Jong, P., Rubin, E., and Higgs, D.R. (2005). Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc Natl Acad Sci U S A *102*, 9830-9835.

Imai, Y., Kurokawa, M., Izutsu, K., Hangaishi, A., Takeuchi, K., Maki, K., Ogawa, S., Chiba, S., Mitani, K., and Hirai, H. (2000). Mutations of the AML1 gene in myelodysplastic syndrome and their functional implications in leukemogenesis. Blood *96*, 3154-3160.

Inoue, K., Ozaki, S., Shiga, T., Ito, K., Masuda, T., Okado, N., Iseda, T., Kawaguchi, S., Ogawa, M., Bae, S.C., *et al.* (2002). Runx3 controls the axonal projection of proprioceptive dorsal root ganglion neurons. Nat Neurosci *5*, 946-954.

Ishida, W., Hamamoto, T., Kusanagi, K., Yagi, K., Kawabata, M., Takehara, K., Sampath, T.K., Kato, M., and Miyazono, K. (2000). Smad6 is a Smad1/5-induced smad inhibitor. Characterization of bone morphogenetic protein-responsive element in the mouse Smad6 promoter. J Biol Chem *275*, 6075-6079.

Iwasaki, H., and Akashi., K. (2007). Myeloid lineage commitment from the hematopoietic stem cell. Immunity *26*, 726-740.

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

Kagoshima, H., Shigesada, K., Satake, M., Ito, Y., Miyoshi, H., Ohki, M., Pepling, M., and Gergen, P. (1993). The Runt domain identifies a new family of heteromeric transcriptional regulators. Trends Genet *9*, 338-341.

Kingsley, P.D., Malik, J., Fantauzzo, K.A., and Palis, J. (2004). Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis. Blood *104*, 19-25.

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

Knezevic, K., Bee, T., Wilson, N.K., Janes, M.E., Kinston, S., Polderdijk, S., Kolb-Kokocinski, A., Ottersbach, K., Pencovich, N., Groner, Y., *et al.* (2011). A Runx1-Smad6 rheostat controls Runx1 activity during embryonic hematopoiesis. Mol Cell Biol *31*, 2817-2826.

Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R.T., Gao, Y.H., Inada, M., *et al.* (1997). Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell *89*, 755-764.

Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell *91*, 661-672.

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

Kumaravelu, P., Hook, L., Morrison, A.M., Ure, J., Zhao, S., Zuyev, S., Ansell, J., and Medvinsky, A. (2002). Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role

of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. Development *129*, 4891-4899.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V., and Lacaud, G. (2009). The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. Nature *457*, 892-895.

Landry, J.R., Kinston, S., Knezevic, K., de Bruijn, M.F., Wilson, N., Nottingham, W.T., Peitz, M., Edenhofer, F., Pimanda, J.E., Ottersbach, K.*, et al.* (2008). Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. Blood *111*, 3005-3014.

Lecuyer, E., and Hoang, T. (2004). SCL: from the origin of hematopoiesis to stem cells and leukemia. Exp Hematol *32*, 11-24.

Leder, A., Kuo, A., Shen, M.M., and Leder, P. (1992). In situ hybridization reveals co-expression of embryonic and adult alpha globin genes in the earliest murine erythrocyte progenitors. Development *116*, 1041-1049.

Levanon, D., Bernstein, Y., Negreanu, V., Bone, K.R., Pozner, A., Eilam, R., Lotem, J., Brenner, O., and Groner, Y. (2011). Absence of Runx3 expression in normal gastrointestinal epithelium calls into question its tumour suppressor function. EMBO Mol Med.

Levanon, D., Bettoun, D., Harris-Cerruti, C., Woolf, E., Negreanu, V., Eilam, R., Bernstein, Y., Goldenberg, D., Xiao, C., Fliegauf, M.*, et al.* (2002). The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. EMBO J *21*, 3454-3463.

Levanon, D., Brenner, O., Negreanu, V., Bettoun, D., Woolf, E., Eilam, R., Lotem, J., Gat, U., Otto, F., Speck, N.*, et al.* (2001). Spatial and temporal expression pattern of Runx3 (Aml2) and Runx1 (Aml1) indicates non-redundant functions during mouse embryogenesis. Mech Dev *109*, 413-417.

Levanon, D., Glusman, G., Bettoun, D., Ben-Asher, E., Negreanu, V., Bernstein, Y., Harris-Cerruti, C., Brenner, O., Eilam, R., Lotem, J.*, et al.* (2003). Phylogenesis and regulated expression of the RUNT domain transcription factors RUNX1 and RUNX3. Blood Cells Mol Dis *30*, 161-163.

Levanon, D., and Groner, Y. (2004). Structure and regulated expression of mammalian RUNX genes. Oncogene *23*, 4211-4219.

Levanon, D., Negreanu, V., Bernstein, Y., Bar-Am, I., Avivi, L., and Groner, Y. (1994). AML1, AML2, and AML3, the human members of the runt domain gene-family: cDNA structure, expression, and chromosomal localization. Genomics *23*, 425-432.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Li, Q.L., Ito, K., Sakakura, C., Fukamachi, H., Inoue, K., Chi, X.Z., Lee, K.Y., Nomura, S., Lee, C.W., Han, S.B.*, et al.* (2002). Causal relationship between the loss of RUNX3 expression and gastric cancer. Cell *109*, 113-124.

Liu, F., Walmsley, M., Rodaway, A., and Patient, R. (2008). Fli1 acts at the top of the transcriptional network driving blood and endothelial development. Curr Biol *18*, 1234-1240.

Look, A.T. (1997). Oncogenic transcription factors in the human acute leukemias. Science *278*, 1059-1064.

Lozzio, C.B., and Lozzio, B.B. (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood *45*, 321-334.

Luc, S., Buza-Vidas, N., and Jacobsen, S.E. (2008). Delineating the cellular pathways of hematopoietic lineage commitment. Semin Immunol *20*, 213-220.

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

Ma, X., de Bruijn, M., Robin, C., Peeters, M., Kong, A.S.J., de Wit, T., Snoijs, C., and Dzierzak, E. (2002). Expression of the Ly-6A (Sca-1) lacZ transgene in mouse haematopoietic stem cells and embryos. Br J Haematol *116*, 401-408.

Mar, B.G., Amakye, D., Aifantis, I., and Buonamici, S. (2011). The controversial role of the Hedgehog pathway in normal and malignant hematopoiesis. Leukemia.

McGrath, K.E., and Palis, J. (2005). Hematopoiesis in the yolk sac: more than meets the eye. Exp Hematol *33*, 1021-1028.

Meadows, S.M., Myers, C.T., and Krieg, P.A. (2011). Regulation of endothelial cell development by ETS transcription factors. Semin Cell Dev Biol.

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

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

Medvinsky, A.L., Gan, O.I., Semenova, M.L., and Samoylina, N.L. (1996). Development of day-8 colony-forming unit-spleen hematopoietic progenitors during early murine embryogenesis: spatial and temporal mapping. Blood *87*, 557-566.

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

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P*., et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mikkola, H.K., Fujiwara, Y., Schlaeger, T.M., Traver, D., and Orkin, S.H. (2003). Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo. Blood *101*, 508-516.

Mikkola, H.K., Gekas, C., Orkin, S.H., and Dieterlen-Lievre, F. (2005). Placenta as a site for hematopoietic stem cell development. Exp Hematol *33*, 1048-1054.

Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. Proc Natl Acad Sci U S A *88*, 10431-10434.

Moore, M.A., and Metcalf, D. (1970). Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo. Br J Haematol *18*, 279-296.

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

Nakagawa, M., Ichikawa, M., Kumano, K., Goyama, S., Kawazu, M., Asai, T., Ogawa, S., Kurokawa, M., and Chiba, S. (2006). AML1/Runx1 rescues Notch1-null mutation-induced deficiency of para-aortic splanchnopleural hematopoiesis. Blood *108*, 3329-3334.

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

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

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

North, T.E., Goessling, W., Peeters, M., Li, P., Ceol, C., Lord, A.M., Weber, G.J., Harris, J., Cutting, C.C., Huang, P.*, et al.* (2009). Hematopoietic stem cell development is dependent on blood flow. Cell *137*, 736-748.

North, T.E., Stacy, T., Matheny, C.J., Speck, N.A., and de Bruijn, M.F. (2004). Runx1 is expressed in adult mouse hematopoietic stem cells and differentiating myeloid and lymphoid cells, but not in maturing erythroid cells. Stem Cells *22*, 158-168.

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

Ogawa, E., Inuzuka, M., Maruyama, M., Satake, M., Naito-Fujimoto, M., Ito, Y., and Shigesada, K. (1993a). Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2 alpha. Virology *194*, 314-331.

Ogawa, E., Maruyama, M., Kagoshima, H., Inuzuka, M., Lu, J., Satake, M., Shigesada, K., and Ito, Y. (1993b). PEBP2/PEA2 represents a family of transcription factors homologous to the products of the Drosophila runt gene and the human AML1 gene. Proc Natl Acad Sci U S A *90*, 6859-6863.

Ogawa, M. (1993). Differentiation and proliferation of hematopoietic stem cells. Blood *81*, 2844-2853.

Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G., and Downing, J.R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell *84*, 321-330.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631-644.

Ottersbach, K., and Dzierzak, E. The placenta as a haematopoietic organ. (2010). Int J Dev Biol *54*, 1099-1106.

Ottersbach, K., and Dzierzak, E. (2005). The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. Dev Cell *8*, 377-387.

Otto, F., Thornell, A.P., Crompton, T., Denzel, A., Gilmour, K.C., Rosewell, I.R., Stamp, G.W., Beddington, R.S., Mundlos, S., Olsen, B.R.*, et al.* (1997). Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell *89*, 765-771.

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

Patterson, L.J., Gering, M., and Patient, R. (2005). Scl is required for dorsal aorta as well as blood formation in zebrafish embryos. Blood *105*, 3502-3511.

Pencovich, N., Jaschek, R., Tanay, A., and Groner, Y. (2011). Dynamic combinatorial interactions of RUNX1 and cooperating partners regulates megakaryocytic differentiation in cell line models. Blood *117*, e1-14.

Pimanda, J.E., Donaldson, I.J., de Bruijn, M.F., Kinston, S., Knezevic, K., Huckle, L., Piltz, S., Landry, J.R., Green, A.R., Tannahill, D.*, et al.* (2007a). The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. Proc Natl Acad Sci U S A *104*, 840-845.

Pimanda, J.E., Ottersbach, K., Knezevic, K., Kinston, S., Chan, W.Y., Wilson, N.K., Landry, J.R., Wood, A.D., Kolb-Kokocinski, A., Green, A.R.*, et al.* (2007b). Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. Proc Natl Acad Sci U S A *104*, 17692-17697.

Pinto, do O.P., Kolterud, A., and Carlsson L. (1998). Expression of the LIM-homeobox gene LH2 generates immortalized Steel factor-dependent multipotent hematopoietic precursors. EMBO J *17*, 5744-5756.

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

Prabhakar, S., Poulin, F., Shoukry, M., Afzal, V., Rubin, E.M., Couronne, O., and Pennacchio, L.A. (2006). Close sequence comparisons are sufficient to identify human cis-regulatory elements. Genome Res *16*, 855-863.

Rampon, C., and Huber, P. (2003). Multilineage hematopoietic progenitor activity generated autonomously in the mouse yolk sac: analysis using angiogenesis-defective embryos. Int J Dev Biol *47*, 273-280.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E.*, et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Rhodes, K.E., Gekas, C., Wang, Y., Lux, C.T., Francis, C.S., Chan, D.N., Conway, S., Orkin, S.H., Yoder, M.C., and Mikkola, H.K. (2008). The emergence of hematopoietic stem cells is initiated in the placental vasculature in the absence of circulation. Cell Stem Cell *2*, 252-263.

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

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

Roh, T.Y., Wei, G., Farrell, C.M., and Zhao, K. (2007). Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res *17*, 74-81.

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

Sanchez, M., Gottgens, B., Sinclair, A.M., Stanley, M., Begley, C.G., Hunter, S., and Green, A.R. (1999). An SCL 3' enhancer targets developing endothelium together with embryonic and adult haematopoietic progenitors. Development *126*, 3891-3904.

Schones, D.E., and Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. Nat Rev Genet *9*, 179-191.

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

Silberstein, L., Sanchez, M.J., Socolovsky, M., Liu, Y., Hoffman, G., Kinston, S., Piltz, S., Bowen, M., Gambardella, L., Green, A.R.*, et al.* (2005). Transgenic analysis of the stem cell leukemia +19 stem cell enhancer in adult and embryonic hematopoietic and endothelial cells. Stem Cells *23*, 1378-1388.

Silver, L., and Palis, J. (1997). Initiation of murine embryonic erythropoiesis: a spatial analysis. Blood *89*, 1154-1164.

Sinclair, A.M., Gottgens, B., Barton, L.M., Stanley, M.L., Pardanaud, L., Klaine, M., Gering, M., Bahn, S., Sanchez, M., Bench, A.J.*, et al.* (1999). Distinct 5' SCL enhancers direct transcription to developing brain, spinal cord, and endothelium: neural expression is mediated by GATA factor binding sites. Dev Biol *209*, 128-142.

Smith, A.M., Sanchez, M.J., Follows, G.A., Kinston, S., Donaldson, I.J., Green, A.R., and Gottgens, B. (2008). A novel mode of enhancer evolution: the Tal1 stem cell enhancer recruited a MIR element to specifically boost its activity. Genome Res *18*, 1422-1432.

Splinter, E., Grosveld, F., and de Laat, W. (2004). 3C technology: analyzing the spatial organization of genomic loci in vivo. Methods Enzymol *375*, 493-507.

Steiner, R., and Vogel, H. (1973). On the kinetics of erythroid cell differentiation in fetal mice. I. Microspectrophotometric determination of the hemoglobin content in erythroid cells during gestation. J Cell Physiol *81*, 323-338.

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

Swiers, G., Speck, N.A., and de Bruijn, M.F. (2010b). Visualizing blood cell emergence from aortic endothelium. Cell Stem Cell *6*, 289-290.

Tahirov, T.H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K., Shiina, M., Sato, K., Kumasaka, T., Yamamoto, M.*, et al.* (2001). Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta. Cell *104*, 755-767.

Taniuchi, I., Osato, M., Egawa, T., Sunshine, M.J., Bae, S.C., Komori, T., Ito, Y., and Littman, D.R. (2002). Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. Cell *111*, 621-633.

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

Taoudi, S., Morrison, A.M., Inoue, H., Gribi, R., Ure, J., and Medvinsky, A. (2005). Progressive divergence of definitive haematopoietic stem cells from the endothelial compartment does not depend on contact with the foetal liver. Development *132*, 4179-4191.

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

Till, J.E., McCulloch, E.A., and Siminovitch, L. (1964). A Stochastic Model of Stem Cell Proliferation, Based on the Growth of Spleen Colony-Forming Cells. Proc Natl Acad Sci U S A *51*, 29-36.

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

Tsai, F.Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W., and Orkin, S.H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature *371*, 221-226.

Tsai, F.Y., and Orkin, S.H. (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood *89*, 3636-3643.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J *26*, 2041-2051.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Wang, H., Zou, J., Zhao, B., Johannsen, E., Ashworth, T., Wong, H., Pear, W.S., Schug, J., Blacklow, S.C., Arnett, K.L.*, et al.* (2011). Genome-wide analysis reveals conserved and divergent features of Notch1/RBPJ binding in human and murine T-lymphoblastic leukemia cells. Proc Natl Acad Sci U S A *108*, 14908-14913.

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

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A.H., and Speck, N.A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging

in the central nervous system and blocks definitive hematopoiesis. Proc Natl Acad Sci U S A *93*, 3444-3449.

Wang, S., Wang, Q., Crute, B.E., Melnikova, I.N., Keller, S.R., and Speck, N.A. (1993). Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. Mol Cell Biol *13*, 3324-3339.

Warren, A.J., Bravo, J., Williams, R.L., and Rabbitts, T.H. (2000). Structural basis for the heterodimeric interaction between the acute leukaemia-associated transcription factors AML1 and CBFbeta. EMBO J *19*, 3004-3015.

Wilkinson, R.N., Pouget, C., Gering, M., Russell, A.J., Davies, S.G., Kimelman, D., and Patient, R. (2009). Hedgehog and Bmp polarize hematopoietic stem cell emergence in the zebrafish dorsal aorta. Dev Cell *16*, 909-916.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schutte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E.*, et al.* (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell *7*, 532-544.

Wilson, N.K., Miranda-Saavedra, D., Kinston, S., Bonadies, N., Foster, S.D., Calero-Nieto, F., Dawson, M.A., Donaldson, I.J., Dumon, S., Frampton, J.*, et al.* (2009). The transcriptional programme controlled by the stem cell leukaemia gene Scl/Tal1 during early embryonic haematopoietic development. Blood.

Woolf, E., Xiao, C., Fainaru, O., Lotem, J., Rosen, D., Negreanu, V., Bernstein, Y., Goldenberg, D., Brenner, O., Berke, G.*, et al.* (2003). Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis. Proc Natl Acad Sci U S A *100*, 7731-7736.

Yoder, M.C., Hiatt, K., Dutt, P., Mukherjee, P., Bodine, D.M., and Orlic, D. (1997). Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. Immunity *7*, 335-344.

Yokomizo, T., Hasegawa, K., Ishitobi, H., Osato, M., Ema, M., Ito, Y., Yamamoto, M., and Takahashi, S. (2008). Runx1 is involved in primitive erythropoiesis in the mouse. Blood *111*, 4075-4080.

Zammit, P.S., Carvajal, J.J., Golding, J.P., Morgan, J.E., Summerbell, D., Zolnerciks, J., Partridge, T.A., Rigby, P.W., and Beauchamp, J.R. (2004). Myf5 expression in satellite cells and spindles in adult muscle is controlled by separate genetic elements. Dev Biol *273*, 454-465.

Zeigler, B.M., Sugiyama, D., Chen, M., Guo, Y., Downs, K.M., and Speck, N.A. (2006). The allantois and chorion, when isolated before circulation or chorio-allantoic fusion, have hematopoietic potential. Development *133*, 4183-4192.

Zovein, A.C., Hofmann, J.J., Lynch, M., French, W.J., Turlo, K.A., Yang, Y., Becker, M.S., Zanetta, L., Dejana, E., Gasson, J.C.*, et al.* (2008). Fate tracing reveals the endothelial origin of hematopoietic stem cells. Cell Stem Cell *3*, 625-636.