Identification of novel Runx1 targets involved in HSC development

Florian Bonkhofer



A thesis submitted for the degree of

Doctor of Philosophy

Weatherall Institute of Molecular Medicine

St John's College

University of Oxford

Trinity Term 2017

Identification of novel Runx1 targets involved in HSC development

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY, TRINITY TERM 2017

FLORIAN BONKHOFER, ST JOHN'S COLLEGE, UNIVERSITY OF OXFORD

Haematopoietic stem and progenitor cells (HSPCs) are *de novo* generated within in the ventral aspects of the embryonic dorsal aorta (DA). Cells of this haemogenic endothelium (HE) will eventually undergo an endothelial to haematopoietic transition (EHT) that involves cell budding out of the aortic wall. Despite the detailed description of the cellular events, the exact haemogenic lineage path and the underlying molecular mechanism that establish full haematopoietic competence are still not entirely understood.

The transcription factor Runx1 is critical for the emergence of HSPCs and shows expression in the zebrafish HE as early as 24 hpf. To facilitate a detailed analysis of the transient HE population I generated a *TgBAC(runx1P2:Citrine)* reporter line under the control of the endogenous *runx1* promoter on a bacterial artificial chromosome (BAC). Double-transgenic reporter lines for *runx1* and the endothelial marker *kdrl* allowed us to isolate specifically cells of the DA away from the whole endothelial population, which could be further sub-divided into HE and non-haemogenic cells. Genome-wide expression analysis within the respective tissues and upon Runx1 loss of function enabled the identification of HE-specific Runx1-regulated genes. Hereby, the *gfi1ab* gene appeared as the functional homologue of the murine *Gfi1*. I show that in zebrafish, EHT is orchestrated through a conserved Runx1-Gfi1-Lsd1 axis. The cellular functions of the remaining Runx1 targets imply that maturation into fully functional HSCs depends on epigenetic regulation due to the up-regulation of *de novo* DNA-methyltransferases, as well as on factors that allow the developing HSCs to respond to extrinsic cues from haematopoietic niches.

Lastly, it became evident that the early HE expresses *dll4* at similar levels to the rest of the aortic endothelium, indicating a common lineage path. In the absence of RUNX1 the HE remains essentially arterial and persists as an integrated part of the DA.

Declaration

I confirm that the work in this thesis is my own unless otherwise indicated. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Any text, data and representation that is not my own work are indicated as such, acknowledged and references appropriately.

Chapter 3

- The double fluorescence ISH (Figure 3-9) and the combined immunofluorescence and fluorescent ISH (Figure 3-17) has been carried out by Monika Krecsmarik

Chapter 4

- RNA sequencing was carried out by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (WTCHG, Churchill Hospital, Oxford)
- The bioinformatic analysis of the RNAseq data was performed in collaboration with Rossella Rispoli

The work for this thesis has been performed by the author at the University of Oxford. This work has not been submitted for any other degree in this or any other University

---

Florian Bonkhofer

Publications

Monteiro, R., Pinheiro, P., Joseph, N., Peterkin, T., Koth, J., Repapi, E., **Bonkhofer, F**., Kirmizitas, A. and Patient, R. (2016). Transforming Growth Factor β Drives Hemogenic Endothelium Programming and the Transition to Hematopoietic Stem Cells. *Dev. Cell* 38, 358–370.

Acknowledgements

First and foremost, I would like to thank my supervisors Roger and Marella for taking me on as a student. I really enjoined the atmosphere you created and the freedom you granted me to explore science while following my interests. Thank you for the great support during my time here in Oxford.

This project had a lot of contributions from many sides and I say "Thank You" to everyone for your support. Tatjana Sauks-Spengler and Upeka Senanayake for kindly teaching me how to perform BAC recombineering and Kevin Clark for the best FAC-sorting experiences in live. A massive thanks to the BMS team including Andy, Kelly, Greg, Adam, Alison and the rest of the team, you are the backbone of all our work.

I will also take the opportunity to thank the whole Patient-laboratory, a bunch of great people that made my stay an exciting journey full of joy and fun. Thanks go to Mukesh for organising the coffee breaks, as well as to Arif and Rui. I enjoyed our great conversations and I had so many good laughs during our breaks. Thanks go to Caroline, Tomek and Michael for being amazing DPhil companions, the future is awaiting us. Jana, you are a great desk neighbour and I believe that everyone would agree with me when I say that our working space was always organised and clean, how it should be for proper Germans. Lei, it is cheering that you are smiling so much and creating a positive spirit. This brings me to Aldo, who unfortunately works only with the 2nd best model organisms there is; I learned a lot from you, more than you might think. I want to thank you Rossella, for making this project a success. Monika, it was a pleasure to work with you. Thank you, Linda, for organizing everything. Lorna, I hope you will have a great time in the lab. Last but not least, special thanks go to Tess and Phil, my lovely supervisors who always lend me an ear when I needed to talk about my thoughts, ideas and worries. You guys kept me on track.

There are so many more people to mention who accompanied me on my journey. Big thanks go to Vanessa, Tom, Lena and Greg who started this Oxford adventure together with me. I big shout out to Ingrid for whom I had the privilege to supervise. *Danke, Lars und Christine für eure Lebensfreude, es tat so gut euch hier kennengelernt zu haben.*

Ulla und Theo, Marius und Benedikt, danke dass ich euch meine Familie nennen darf. Die Zeit zusammen wird immer kostbarer. Ich freue mich schon darauf wieder näher bei euch zu sein, wann auch immer das sein wird. Vielen Dank Opa, dafür, dass du ein Beweis dafür bist, dass man bis zum Ende lieben kann. Vielen Dank Oma, das du die Kraft hattest noch auf mich zu warten, sodass ich Dir Lebewohl sagen konnte.

Vielen Dank an euch, Susanne und Juergen, Ralf und Bouchra, dafür, dass ihr meine Familie so sehr erweitert habt. Ich bin gern bei euch.

Mein letzter und wichtigster Dank geht an dich, Kathie. Wir haben Oxford zusammen überlebt, wir werden auch alles was weiterhin auf uns zukommen wird gemeinsam meistern. Ich bin so froh, dass ich dich habe und dass es jemanden gibt der wichtiger als man selbst ist. Lass uns jetzt gemeinsam unsere Zukunft in vollen Zügen genießen. Zu zweit. Füreinander da. Danke.

Table of Contents

Declaration II
Publications III
Acknowledgements IV
Table of Contents V
Table of figures IX
Table of tables XI
Abbreviations XII

- Chapter 1 - 1
Introduction
1.1 The haematopoietic system 2
1.1.1 Haematopoietic stem cells (HSCs) and the haematopoietic tree 2
1.1.2 Current and future clinical applications for HSC transplantations 4
1.2 Embryonic haematopoiesis 5
1.2.1 Yolk sac (YS) haematopoiesis 7
1.2.2 Intra-embryonic haematopoiesis 8
1.3 The zebrafish as a model system 10
1.4 Haematopoiesis in zebrafish 11
1.5 The aortic haemogenic endothelium 14
1.6 The transcription factor Runx1 18
1.6.1 Role of RUNX1 during haematopoiesis 20
1.6.2 Mechanisms of Runx1 action 22
Objectives and general experimental strategy 26
Thesis aims 28

- Chapter 2 - 29
Materials and Methods
2.1 Materials 30
2.1.1 Solutions 30
2.1.2 Morpholinos 31
2.1.3 Chemical inhibitors 31
2.1.4 Primers 32
2.1.5 Primers for quantitative real-time PCR 32
2.1.6 Primers used for making gRNA, MO and CRISPR/Cas9 validation, genotyping and characterisation of BACs 32
2.1.7 Primers used for BAC cloning 33
2.1.8 Primers for ISH 33
2.1.9 TaqMan probes 34
2.1.10 Plasmids with antisense probe constructs for ISH 35
2.1.11 BACs and Plasmids for BAC recombineering 36
2.2 Zebrafish husbandry and manipulation of embryos 37
2.2.1 Zebrafish lines 37
2.2.2 Zebrafish maintenance 37
2.2.3 Breeding, embryo collection and maintenance 37

2.2.4 Micro-injections of zebrafish embryos ... 37
2.2.5 Pharmacological treatment of embryos ... 38
2.2.6 Dechorionation and fixation of zebrafish embryos ... 38
2.2.7 Fin-clipping of zebrafish embryos ... 38
2.2.8 Fluorescence-activated cell sorting (FACS) ... 38
2.3 Extraction, preparation and manipulation of DNA ... 39
2.3.1 Bacterial culture ... 39
2.3.2 Transformation of chemically competent *Escherichia coli* ... 39
2.3.3 Isolation and extraction of plasmid DNA ... 40
2.3.4 Genomic DNA extraction from zebrafish embryos ... 40
2.3.5 Genomic DNA extraction from fin-clips ... 40
2.3.6 Determining nucleic acid concentration ... 40
2.3.7 Polymerase Chain Reaction (PCR) ... 40
2.3.8 Purification of PCR amplified DNA ... 41
2.3.9 Restriction enzyme digest ... 41
2.3.10 Agarose gel electrophoresis ... 41
2.3.11 Purification of gel extracted DNA ... 41
2.3.12 Cloning of PCR products ... 42
2.3.13 BAC cloning ... 42
2.3.14 Sanger DNA sequencing ... 43
2.4 Extraction, preparation and manipulation of RNA ... 43
2.4.1 RNA isolation from tissue samples ... 43
2.4.2 RNA Sample Preparation from FAC-sorted cells for RNA sequencing and quality/quantity determination ... 43
2.4.3 Sample preparation from FAC-sorted cells for Fluidigm BioMark gene expression analysis ... 44
2.4.4 Reverse Transcription of RNA ... 45
2.4.5 *In vitro* synthesis of capped mRNA ... 45
2.4.6 *In vitro* synthesis of antisense RNA-probes ... 45
2.4.7 In vitro synthesis of guide RNA (gRNA) ... 46
2.4.8 Whole mount *in situ* hybridisation (ISH) ... 47
2.4.9 Double fluorescent ISH ... 47
2.4.10 fluorescent ISH with immunohistochemistry staining ... 48
2.4.11 Quantitative Polymerase Chain Reaction (qPCR) ... 48
2.5 Microscopy ... 49
2.5.1 Bright-field microscopy ... 49
2.5.2 Wide-field fluorescence imaging ... 49
2.5.3 (Live) confocal imaging ... 50
2.5.4 Image analysis and representation ... 50
2.6 Bioinformatic tools ... 50
2.6.1 Genomic, transcript, and protein sequence information ... 50
2.6.2 Analysis of homology ... 50
2.6.3 Intersection of gene lists. ... 51
2.6.4 RNAseq analysis ... 51
2.6.5 GO term and KEGG pathway analysis. ... 52

- Chapter 3 - ... 53

Generation of a novel *runx1* reporter line that allows for the reliable isolation of HE cells

3.1 Introduction ... 54
3.1.1 Chapter aims ... 56
3.2 Results ... 57
3.2.1 The zebrafish *runx1* locus ... 57
3.2.1.1 Splicing of the zebrafish P1 and P2 derived *runx1* isoforms involves a non-species conserved intron ... 58
3.2.1.2 *Runx1* expression in the zebrafish HE is initiated via the P2 promoter ... 61

3.2.2 Generation of a novel *runx1* reporter line based on a zebrafish BAC ... 63
3.2.2.1 Characterisation of zebrafish *runx1*-containing BACs ... 63
3.2.2.2 The 97a02 BAC allows for the generation of a novel *runx1P2* reporter line. ... 66
3.2.2.3 Generation of a stable *TgBAC(runx1P2:Citrine)* reporter line ... 68
3.2.3 In depth analysis of reporter expression in *TgBAC(runx1P2:Citrine)* embryos ... 70
3.2.3.1 Transgene expression during early haematopoiesis recapitulates *runx1* expression in the PLM ... 70
3.2.3.2 Transgene expression during definitive haematopoiesis recapitulates *runx1* expression in the HE ... 73
3.2.3.3 The *runx1P2:Citrine* transgene is Tal1b and Gata2b dependent but Runx1 independent ... 75
3.2.4 FACS based isolation of the HE ... 77
3.2.4.1 An optimised FACS gating strategy allows for the isolation of the HE and additionally of the DA roof ... 77
3.2.4.2 Quantification of FACS isolated DA and HE cells ... 80
3.2.5 Analysis of the *runx1P2* reporter expression in the DA roof ... 82
3.2.5.1 Citrine fluorescence in the DA roof is caused by residual protein ... 82
3.2.5.2 Residual Citrine protein in the DA roof is caused by transgene expression from around 18-22 hpf ... 83
3.3 Discussion ... 86
3.3.1 Species-specific differences in the P2 N-terminus and the zebrafish *runx1* promoter usage ... 86
3.3.2 The 97a02 BAC covers the complete first *runx1* intron and contains most likely all haematopoietic enhancers. 87
3.3.3 Expression kinetics of the *runx1* reporter places upstream signal activity at around 22 hpf or earlier ... 88
3.3.4 *Runx1P2* promoter activity is independent of Runx1 ... 89
3.3.5 *Runx1P2* reporter activity in the early DA suggests a close relationship between HE and aortic endothelium ... 90
3.3.6 The ++HIGH fraction contains purified HE ... 92
3.3.7 How the refined gating strategy will influence further analysis ... 93
3.4 Conclusion and directions ... 94

- Chapter 4 - ... 95
Genome wide expression analysis of the HE and identification of potential Runx1 target genes
4.1 Introduction ... 96
4.1.1 Chapter aims ... 98
4.2 Results ... 99
4.2.1 Experimental design for the RNAseq analysis of the zebrafish HE ... 99
4.2.1.1 Using a *runx1* MO for the functional analysis of HE ... 99
4.2.1.2 Detailed experimental design for the generation of the RNAseq samples ... 100
4.2.2 Clustering analysis confirms the haemogenic and aortic identity of the ++HIGH and ++LOW populations ... 103
4.2.3 Identification of HE-specific genes ... 108
4.2.3.1 DEG analysis between ++LOW and ++HIGH populations identifies known and novel HE-specific genes.. 108
4.2.3.2 Zebrafish HE-specific genes contain a core set of conserved haematopoietic genes ... 111
4.2.4 Identification of Runx1-regulated genes ... 113
4.2.4.1 DEG analysis between *runx1* morphants and control embryos identifies novel Runx1 targets ... 113
4.2.4.2 Identified Runx1-regulated genes contain direct RUNX1 targets in murine haematopoietic progenitor cells ... 116
4.2.4.3 Newly identified Runx1-regulated genes contain RUNX1 targets in t(8;21) AML cells ... 117
4.2.5 Identification of high-confidence HE-specific Runx1-regulated genes ... 118
4.3 Discussion ... 122
4.3.1 The executed RNAseq experiment allowed for a robust analysis of DEGs and controlled for MO off-targets . 122
4.3.2 The zebrafish HE shows an early loss of the endothelial programme including *dll4* ... 123
4.3.3 *Runx1* positively regulates expression of *dnmt3* genes which might have implications for AML ... 125
4.3.4 Runx1 keeps the erythroid programme in check and presumably represses endothelial genes via *gfi1ab* ... 127
4.3.5 High-confidence Runx1 targets include genes with up-regulated expression during definitive haematopoiesis 128
4.4 Conclusion and directions ... 129

- Chapter 5 – 131

Detailed analysis of the consequences of the loss of Runx1 on the HE *in vivo*

5.1 Introduction 132

5.1.1 Chapter aims 134

5.2 Results 135

5.2.1 Identification of high-confidence potential Runx1 target genes 135

5.2.2 Expression analysis of high-confidence potential Runx1 target genes by ISH 137

5.2.3 Set-up of a qPCR based validation screen 141

5.2.3.1 Design of the BioMark-based expression screen 141

5.2.3.2 qPCR expression analysis of haematopoietic and endothelial genes 143

5.2.3.3 Analysis of potential *runx1* MO off-targets in *runx1* mutants 146

5.2.4 Validation of high-confidence potential Runx1 targets by qPCR 149

5.2.4.1 Expression dynamics of *dnmt* family members 149

5.2.4.2 Expression dynamics of high-confidence Runx1 target genes 152

5.2.4.3 Expression dynamics of non-ISH validated high-confidence Runx1 target genes 156

5.2.5 EHT in zebrafish is dependent on a conserved Runx1-Gfi1-Lsd1 axis 157

5.2.6 The HE retains an arterial programme in the absence of Runx1 162

5.3 Discussion 167

5.3.1 Successful screening of Runx1 target genes 167

5.3.2 Runx1 orchestrates EHT through activation of *gfi1ab* and repression of *sox17* and Dll4-Notch signalling 169

5.3.3 Runx1 establishes an epigenetic state in haematopoietic progenitor cells through regulation of *dnmt3* genes 172

5.3.4 Runx1 regulates genes with reported functions in the haematopoietic system already during HE stages 174

5.3.5 Runx1 regulates expression of stem cell niche interaction factors 176

5.3.6 The HE is initially *Dll4*$^{+}$ and maintains an arterial programme in *runx1* mutants 178

5.4 Conclusion 181

- Chapter 6 - 182

Final conclusion

- Chapter 7 - 186

Appendix

- Chapter 8 - 198

Bibliography

Table of figures

- Chapter 1 -

Figure 1-1 The haematopoietic developmental hierarchy 3
Figure 1-2 Overview of embryonic haematopoiesis in mouse..... 6
Figure 1-3: Haematopoietic waves during murine embryonic development..... 8
Figure 1-4: Haematopoietic sites in zebrafish. 12
Figure 1-5: Structure of the human RUNX1 protein..... 19
Figure 1-6: Experimental strategy to study global gene expression of the HE in vivo. 27

- Chapter 3 -

Figure 3-1: Annotation of the zebrafish *runx1* gene locus. 58
Figure 3-2: Species specific differences between the human *RUNX1* and zebrafish *runx1* gene. 60
Figure 3-3: Analysis of the zebrafish *runx1* promoter usage. 62
Figure 3-4: Characterization of zebrafish BACs containing different regions of the runx1 locus. 64
Figure 3-5: Generation of runx1P2 reporter constructs based on the two BACs 97a02 and 14i20..... 66
Figure 3-6: Transient expression analysis of embryos after injection with the recombineered *97a02* BAC. 68
Figure 3-7: *Citrine* expression in stable *TgBAC^{97a02}(runx1P2:Citrine)* reporter lines..... 69
Figure 3-8: *Citrine* transgene expression in the PLM of *TgBAC(runx1P2:Citrine)* embryos. 71
Figure 3-9: Citrine fluorescence analysis during primitive haematopoiesis of *TgBAC(runx1P2:Citrine)* embryos. 72
Figure 3-10: *Citrine* expression analysis during definitive hematopoiesis in double transgenic *TgBAC(runx1P2:Citrine), Tg(kdrl:mCherry)* embryos. 74
Figure 3-11: Analysis of *Citrine* transgene regulation in *TgBAC(runx1P2:Citrine)* embryos. 75
Figure 3-12: Analysis of Citrine fluorescence in *TgBAC(runx1P2:Citrine)* embryos on a *runx1$^{W84X/W84X}$* mutants background. 77
Figure 3-13: Isolation of HE cells from double transgenic *TgBAC(runx1P2:Citrine), Tg(kdrl:mCherry)* embryos via FACS. 78
Figure 3-14: qPCR analysis of HE and endothelial marker genes in FACS isolated cell populations of double transgenic *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* embryos..... 80
Figure 3-15: Quantification of FACS isolated cell populations of double-transgenic *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos at 28-30 hpf. 81
Figure 3-16: Analysis of Citrine fluorescence in the DA roof of *TgBAC(runx1P2:Citrine)* embryos..... 82
Figure 3-17: Combined immunofluorescence staining against Citrine protein and fluorescent ISH for *Citrine* mRNA. 83
Figure 3-18: Citrine fluorescence analysis in 18-22 hpf *TgBAC(runx1P2:Citrine)* embryos. 84
Figure 3-19: *Runx1* and *Citrine* mRNA expression analysis in 18-24 hpf *TgBAC(runx1P2:Citrine)* embryos. 85
Figure 3-20: Suggested model about the initiation of *Citrine* expression in *TgBAC(runx1P2:Citrine)* embryos..... 92

- Chapter 4 -

Figure 4-1: Validation of a *runx1*-MO for functional studies of the HE by RNAseq. 99
Figure 4-2: Timeline of the most important events during definitive haematopoiesis in the zebrafish HE..... 100
Figure 4-3: Experimental strategy to study the zebrafish HE by RNAseq..... 101
Figure 4-4: Comparison of three kits for the isolation of RNA from cells after FACS. 102
Figure 4-5: Consensus clustering of RNAseq data from Citrine$^{+/-}$mCherry^{+} endothelial populations. 104
Figure 4-6: Heatmaps of the 8 gene clusters resulting from consensus clustering as shown in Figure 4-5. 105
Figure 4-7: Detailed assessment of HE marker gene expression as detected by RNAseq 108
Figure 4-7: Identification of zebrafish HE-specific genes. 110

Figure 4-8: Identification of HE-specific genes with an evolutionarily conserved expression during murine haematopoiesis. 112
Figure 4-9: Identification of Runx1-regulated genes. 114
Figure 4-10: Comparison of Runx1-regulated genes with RUNX1 ChIPseq targets in HPC-7 cells. 117
Figure 4-11: Comparison of Runx1-regulated genes with RUNX1 and/or RUNX1/ETO ChIPseq targets in AML cells. 118
Figure 4-12: Identification of high-confidence zHE-specific Runx1-activated genes. 119

- Chapter 5 -

Figure 5-1: Selection logic for the identification of high-confdence potential Runx1 target genes. 136
Figure 5-2: ISH based expression screen for high-confidence potential Runx1 target genes. 139
Figure 5-3: ISH based expression screen for additional potential Runx1 target genes. 140
Figure 5-4: Setup of the Biomarkbased expression screen. 142
Figure 5-5: Gene expression analysis of haematopoietic and endothelial control genes. 144
Figure 5-6: Expression of *gata2a* in the HE is repressed by Runx1 146
Figure 5-7: Gene expression analysis of potential *runx1* MO off-target genes. 148
Figure 5-8: Gene expression analysis of DNA methyl-transferases. 150
Figure 5-9: Haemogenic expression of *dnmt3* isoforms is dependent on Runx1. 151
Figure 5-10: Analysis of *dnmt* genes for expression in the CHT region. 151
Figure 5-11: Gene expression analysis of four high-confidence potential Runx1 target genes identifies *angpt1*, *gfi1ab*, *pik3cd* and *irf1b* as bona fide Runx1 targets. 152
Figure 5-12: Haemogenic expression of *irf1b* and *angpt1* is dependent on Runx1 153
Figure 5-13: Deatailed expression dynamics of *irf1b*, *pik3cd* and *angpt1* within the HE by ISH. 154
Figure 5-14: *Irf1b* and *pik3cd*/*pik3r1* are specific Runx1 targets within their respective gene families. 155
Figure 5-15: Gene expression analysis identifies *mpl* and *syk* as bona fide Runx1 targets. 157
Figure 5-16: Expression analysis of *gfi1* isoforms during definitive haematopoieis in the DA and the CHT. 158
Figure 5-17: *Gfi1ab* expression in the HE can be visualized in *gfi1aa* morphants and is Runx1-dependent. 159
Figure 5-18: EHT in zebrafish is dependent on a conserved Runx1-Gfi1-Lsd1 axis. 161
Figure 5-19: Gene expression analysis of aretrial and venous marker genes reveals an arterial origin of the HE. 163
Figure 5-20: The HE of *runx1*$^{W84X/W84X}$ mutants retains an arterial programme and expression of the HE marker sox17. 164
Figure 5-21: The HE in *runx1* mutants remains integrated into the DA scaffold. 166
Figure 5-22: Role of Runx1 during the transition from HE to HSPCs. 168
Figure 5-23: Regulation of *gfi1aa* and *gfi1ab* in the HE. 171
Figure 5-24: A higher Jag1/Dll4 signalling ratio in the HE is accomplished by a Runx1-dependent down-regulation of *dll4*. 172
Figure 5-25: Model depicting the lineage relationship of the arterial and haemogenic endothelium. 179

Table of tables

- Chapter 2 -

Table 2-1: Solutions. 30
Table 2-2: Chemical inhibitors. 31
Table 2-3: qPCR primer pairs 32
Table 2-4: Primers for testing MO efficiency and genotyping. 32
Table 2-5: BAC recombineering primers 33
Table 2-6: Primer pairs for general ISH probes. 33
Table 2-7: TaqMan probes. 34
Table 2-8: ISH probe plasmids. 35
Table 2-9: BAC cloning plasmids. 36

- Chapter 3 -

Table 3-1: Genomic location of BAC ends as determined by BAC-end sequencing. 65

- Chapter 4 -

Table 4-1: GO term analysis among the genes of cluster 1 to 8 from Figure 4-6. 107
Table 4-2: Lists of *runx1* MO off-targets in the zebrafish HE 115
Table 4-3: GO term analysis on Runx1-activated and Runx1-repressed genes. 116
Table 4-4: KEGG Pathway analysis on the 85 HE-specific Runx1-activated genes. 120
Table 4-5: GO term analysis on the 85 HE-specific Runx1-activated genes. 120
Table 4-6: List of HE-specific Runx1-activated genes with an evolutionarily conserved expression during mouse haematopoiesis. 121

- Chapter 5 -

Table 5-1: List of the 16 identified high-confidence potential Runx1 target genes. 136
Table 5-2: List of genes that were probed for expression by ISH. 137
Table 5-3: List of genes that were analysed using the BioMark-based expression screen. 143
Table 5-4: List of runx1-MO off-targets genes anaysed. 147
Table 5-5: Summary of the performed screens to validate HE-specific Runx1-activated genes. 167

Abbreviations

AGM	aorta–gonad–mesonephros
ALM	anterior lateral plate mesoderm
AML	acute myeloid leukaemia
BAC	bacterial artificial chromosome
BM	bone marrow
BMP	Bone Morphogenetic Protein
ChIPseq	chromosome immunoprecipitation sequencing
CHT	caudal haematopoietic tissue
CLP	common lymphoid progenitor
CMP	common myeloid progenitor
DA	dorsal aorta
dpf	days post fertilisation
E	embryonic day
EC	endothelial cells
EHT	endothelial to haematopoietic transition
EMP	erythroid-myeloid progenitor
ESC	embryonic stem cell
ETP	early thymocyte progenitor
FACS	fluorescence-activated cell sorting
FGF	Fibroblast Growth Factor
FL	foetal liver
GMP	granulocyte/macrophage progenitor
HCC	hematopoietic cluster cells
HE	haemogenic endothelium
HEC	haemogenic endothelial cells
HSC	haematopoietic stem cell
HSPC	haematopoietic stem and progenitor cell
hpf	hours post fertilisation
IAC	intra-aortic cluster
iPS	induced pluripotent stem
ICM	intermediate cell mass
ISH	*in situ* hybridisation
ISV	Inter-somitic vessel
kb	kilo base
LMPP	lympho-myeloid primed progenitor
LSK	$Lin^-/Sca\text{-}1^+/c\text{-}KIT^+$
LT-HSC	long-term repopulating haematopoietic stem cell
MEP	megakaryocyte erythroid progenitor
MO	morpholino

MPP	multipotent progenitor
OP	Olfactory placode
PBI	posterior blood island
PLM	posterior lateral plate mesoderm
PTU	phenylthiourea
qPCR	quantitative polymerase chain reaction
RBN	Rohon-Beard neurons
RD	Runt domain
RNAseq	RNA sequencing
ST-HSC	short-term repopulating haematopoietic stem cell
TF	Transcription factor
TGFβ	Transforming Growth Factor Beta
Wt	Wild type
YS	yolk sac
++HIGH	*kdrl:mCherry*$^{+}$*;runx1P2*^{HIGH}
++MEDIUM	*kdrl:mCherry*$^{+}$*;runx1P2*^{MEDIUM}
++LOW	*kdrl:mCherry*$^{+}$*;runx1P2*^{LOW}

- Chapter 1 -

Introduction

1.1 The haematopoietic system

The haematopoietic system comprises a plethora of blood cell types, each with distinct and essential functions. Important tasks include the transport of nutrients, tissue oxygenation and immune defence. Blood cells circulate via the cardiovascular system, that pumps blood to tissues and organs throughout the body. Mature blood cells have – depending on their cell type - variable life spans, that range from several years to only a few days. Consequently, cells of the haematopoietic system have to be continuously replaced throughout life, a process known as haematopoiesis. It is estimated that the human body produces approximately one trillion (10^{12}) new blood cells daily. The maintenance of the blood system relies on self-renewing haematopoietic stem cells (HSCs) that reside in the bone marrow (BM) niche of adult mammals.

1.1.1 Haematopoietic stem cells (HSCs) and the haematopoietic tree

Multipotent HSCs give rise to all mature blood lineages through a balance of differentiation and self-renewal, and are defined functionally by their capacity to reconstitute the entire blood system of a lethally irradiated recipient (Micklem et al., 1972; Till and McCulloch, 1961). This definition includes the ability of transplanted HSCs to home to their appropriate haematopoietic niches. Functional HSCs can be enriched for based on the presence or absence of specific cell surface markers. The lack of surface markers for mature blood cells (Lin^-) and expression of *Sca-1* and *c-Kit* defines, in mouse, the so-called $Lin^-/Sca\text{-}1^+/c\text{-}KIT^+$ (LSK) population (Figure 1-1) (Okada et al., 1992; Spangrude et al., 1988a). Based on their self-renewal ability, HSCs can be further classified into long-term reconstituting HSCs (LT-HSCs) that have the ability to sustain life-long reconstitution of all mature lineages, whilst short-term HSCs (ST-HSCs) have limited reconstitution capacity *in vivo* (Adolfsson et al., 2001; Morrison and Weissman, 1994).

As for other stem cell systems, HSCs depend on their microenvironment for the regulation of self-renewal and differentiation. In adults, HSCs are found in the BM niche (Crane et al., 2017). It was suggested that HSCs interact with thrombopoietin-producing osteoblasts through which they are

maintained in a quiescent state (Yoshihara et al., 2007). Other studies found HSCs adjacent to a vascular niche (Kiel and Morrison, 2006; Kiel et al., 2005). It is possible that HSCs reside in various sites within the BM and that their function depends on their precise location. Whilst many aspects of haematopoiesis are conserved across species, the location of the niche is not and in zebrafish HSCs reside in the kidney, for instancs (Traver et al., 2003).

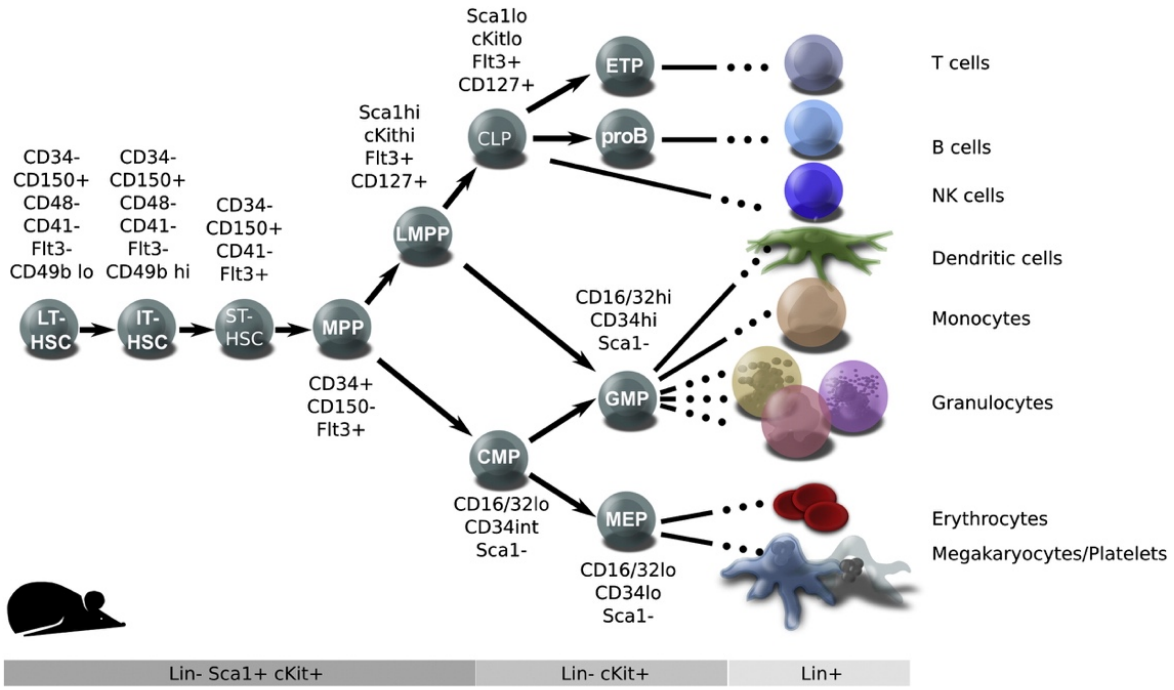


Figure 1-1 The haematopoietic developmental hierarchy

Depicted is the murine haematopoietic tree as it is most widely accepted. At the top of the haematopoietic hierarchy lie the HSCs that can be distinguished based on their repopulating potential into long term- (LT-) intermediate term- (IT-) or short term- (ST-) HSCs. Terminally differentiated cells are shown on the right, and inferred lineage relationships are depicted with arrows. It can be expected that the lineage relationships will be retraced due to technical advances including single cell sequencing (Perié and Duffy, 2016). MPP, multi-potent progenitor; LMPP, lymphoid primed MPP; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; ETP, early thymocyte progenitor; GMP, granulocyte macrophage progenitor; MEP, megakaryocyte erythroid progenitor; Adapted from Doulatov et al. (2012).

HSCs sit atop a hierarchy of progenitors that eventually commit to further maturation along specific pathways with restricted lineage potential (Orkin, 2000). Haematopoiesis is usually depicted in a (oversimplified) hierarchical fashion (Figure 1-1). The actual model of HSC lineage restriction proposes that HSCs and multipotent progenitors (MPPs) gradually lose the ability to self-renew, whilst still retaining the multi-lineage potential. This is followed by subsequent stepwise lineage

restriction to more committed progenitors and the final generation of terminally differentiated cells (Doulatov et al., 2012). Such models will most likely be refined by new technical advances including single cell sequencing (Perié and Duffy, 2016).

1.1.2 Current and future clinical applications for HSC transplantations

Stem cell-based therapies hold great promise in regenerative medicine. Transplantation of HSCs has been successfully applied in the clinic for many years, for both malignant and non-malignant diseases (Gratwohl et al., 2011; Staal et al., 2011). However, when performing allogeneic BM transplantation, there is a risk of developing graft-versus-host disease in which immune cells of the donated graft attack the patient's tissue (Hale et al., 1998; Le Blanc et al., 2008; Petersdorf, 2013). On the other hand, immunogenic reactions of the host after transplantation of foreign tissue can lead to a transplant rejection, thus further limiting its broader utility. In addition to the shortage of cells from matching donors, the numbers of adult HSCs in the tissues used for transplantation are low, creating a major challenge for their clinical use (Hofmeister et al., 2007). Due to these complications, there is an unmet clinical need to either expand the HSC population prior to transplantation or to generate patient-specific HSCs *in vitro*, negating the need for immunologically matched donors (Murry and Keller, 2008).

The potential of embryonic stem cells (ESCs) to give rise to any cell type of the body raised new prospects for biomedical research as well as for regenerative medicine (Murry and Keller, 2008). Building on this potential, the discovery that somatic cells can be reprogrammed into induced pluripotent stem (iPS) cells with similar properties to ESCs will eventually make patient-specific transplants feasible (Takahashi and Yamanaka, 2006). The original reprogramming protocol was based on the viral introduction of a defined set of transcription factors (TF). To circumvent potential adverse effects caused by insertional mutagenesis or by miss-expression of tumour suppressors and oncogenes, more recent reprogramming strategies are based on integration-free vector systems, synthetic mRNA or on direct protein delivery (Stadtfeld and Hochedlinger, 2010).

Haematopoietic development of human ESCs and iPS cells has been demonstrated by multiple groups (Ditadi et al., 2015; Kaufman et al., 2001; Kennedy et al., 2007). Despite recent advances in establishing culture conditions that allow for the development of progenitor cells with lymphoid potential, it is still not possible to generate transplantable HSCs without the help of genetic manipulations (Clarke et al., 2013; Sturgeon et al., 2014). Other strategies involve a direct cell reprogramming and trans-differentiation of various cell types into haematopoietic (progenitor) cells (Elcheva et al., 2014; Pereira et al., 2013; Sandler et al., 2014; Szabo et al., 2010). Recently, two studies reported for the first time the successful generation of HSCs *in vitro* (Lis et al., 2017; Sugimura et al., 2017). These protocols still require the stable integration of a set of TF-encoding genes. It can be expected that further optimizations will include the use of integration-free vector systems as well as a refinement of the culture conditions to establish an efficient and stable production of transplantable HSCs. With this goal in mind, it is crucial to elucidate the cellular processes and molecular programs that drive HSC *de novo* generation in the embryo.

1.2 Embryonic haematopoiesis

The BM niche is seeded by HSCs that were generated during embryogenesis. The mechanisms of how HSCs are *de novo* generated during development are under intense study but our knowledge remains incomplete. Blood is a very mobile tissue and analysis is further complicated by the complexity of embryonic haematopoiesis, which occurs at multiple times and tissues during development (Dzierzak and Speck, 2008). Haematopoiesis has been most extensively studied in mouse due to the establishment of colony-forming and transplantation assays, with which both the presence and the cellular potential of emerging progenitors can be determined (Cumano and Godin, 2007; Frame et al., 2013). The first blood cells in the murine conceptus are found in the blood islands of the extra-embryonic yolk sac (YS) in close association with the developing vasculature (Figure 1-1 A) (Medvinsky et al., 2011). Later, haematopoietic sites can be found intra-embryonically including the aorta–gonad–mesonephros (AGM) region and the foetal liver (FL). It became clear that

haematopoiesis initiates in several waves, each with a different output of blood cells (Figure 1-2). The cellular outcome of such blood cells can be broadly divided into primitive or definitive cells.

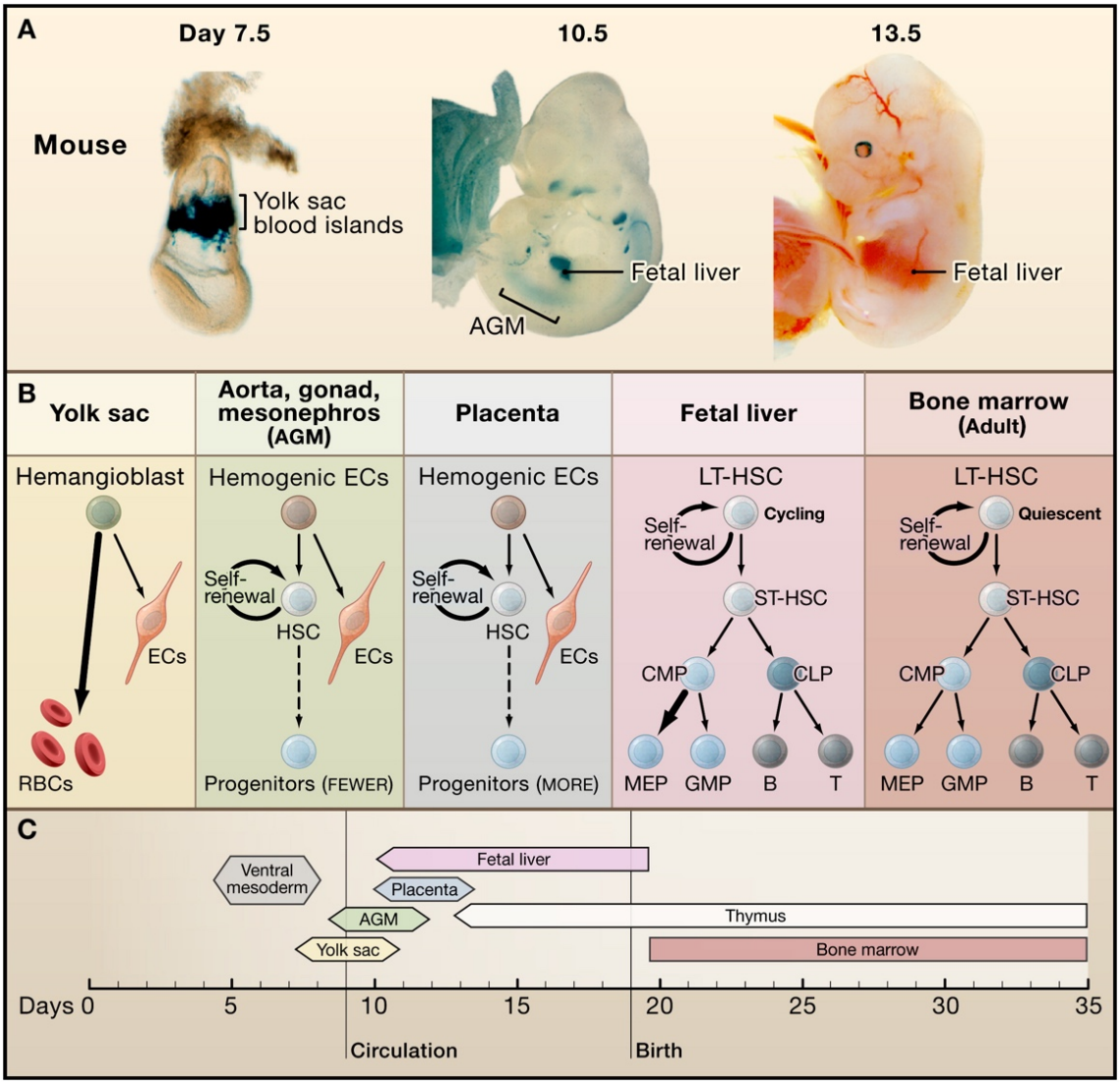


Figure 1-2 Overview of embryonic haematopoiesis in mouse.

Haematopoiesis during embryogenesis occurs in many sites that are both spatially and temporily separated. A) Haematopoiesis initiates first extra-embryonically in the yolk sac (YS) blood islands around E7.5. Another site of blood production is found intra-embryonically in the aorta-gonad-mesonephros (AGM) region around E10.5. Haematopoiesis then shifts to the foetal liver (FL). B) Each haematopoietic site favours specific blood lineages. C) Schematic highlighting the developmental time windows for each haematopoietic site. RBC, red blood cell; EC, endothelial cell; LT-HSC, long term HSC; ST-HSC, short term HSC; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte-erythroid progenitor; GMP, granulocyte-macrophage progenitor; B, B-cell; T, T-cell. Reproduced from Orkin and Zon (2008).

1.2.1 Yolk sac (YS) haematopoiesis

Haematopoiesis emerges in three consecutive waves, with the first wave initiating in the extra-embryonic YS around E7.25 mainly producing primitive erythrocytes (Figure 1-2 A, Figure 1-3) (Palis et al., 1999). The term "primitive" was based on the large size and the nucleated status of such erythrocytes. Further studies revealed that additionally macrophage and megakaryocyte lineages are associated temporally with the generation of primitive erythrocytes in the YS (Palis et al., 1999; Tober et al., 2007). Early generated blood cells are mainly required to meet the demands of the rapidly growing embryo as it becomes too large for oxygenation through diffusion and to provide innate immunity (Frame et al., 2013).

A second wave of haematopoiesis is marked by the emergence of erythro-myeloid progenitors (EMPs) from a so called haemogenic endothelium (HE) located in the blood island regions of the YS around E8.25 (Figure 1-3) (Palis et al., 1999; Palis et al., 2001; Wong et al., 1986). Such cells are classified as "definitive" since they generate enucleated erythrocytes and have a higher proliferative potential. However, it has been suggested that both the primitive and EMP-derived macrophages persist into adulthood, where they are found as tissue-resident macrophages populating the central nervous system or other organs (Alliot et al., 1999; Gomez Perdiguero et al., 2014; Herbomel et al., 2001; Kierdorf et al., 2013; Xu et al., 2015). The importance of EMPs is highlighted by the observation that the loss of EMPs causes embryonic lethality, even when transplantable HSCs are present (Chen et al., 2011). In contrast, in the absence of HSCs, the presence of EMPs enables mouse embryos to survive until birth (Chen et al., 2011). Strikingly, such studies and others have demonstrated that lymphoid potential can be generated independently of HSCs, as early as E8.25 before the onset of circulation (Böiers et al., 2013; Chen et al., 2011; Luc et al., 2012; Yokota et al., 2006; Yoshimoto et al., 2011; Yoshimoto et al., 2012).

Whereas these first two waves of haematopoiesis occur extra-embryonically in the YS, the third wave is initiated intra-embryonically and involves the *de novo* generation of HSCs (Figure 1-3). The term "definitive haematopoiesis" is sometimes ambiguously utilized whereby some authors are exclusively refering to repopulating HSC activity, whereas others use the term to describe the outcome of both the second and third wave of haematopoiesis. Here in this work, I will make use of the broader

meaning of the term "definitive", as suggested by others (Frame et al., 2013), since it accurately reflects the robust definitive erythroid outcome detected in the YS as well. To precisely distinguish between these different waves of haematopoiesis, terms like "second/third wave", "YS/intra-embryonic" or "HSC-generating" will be used as required.

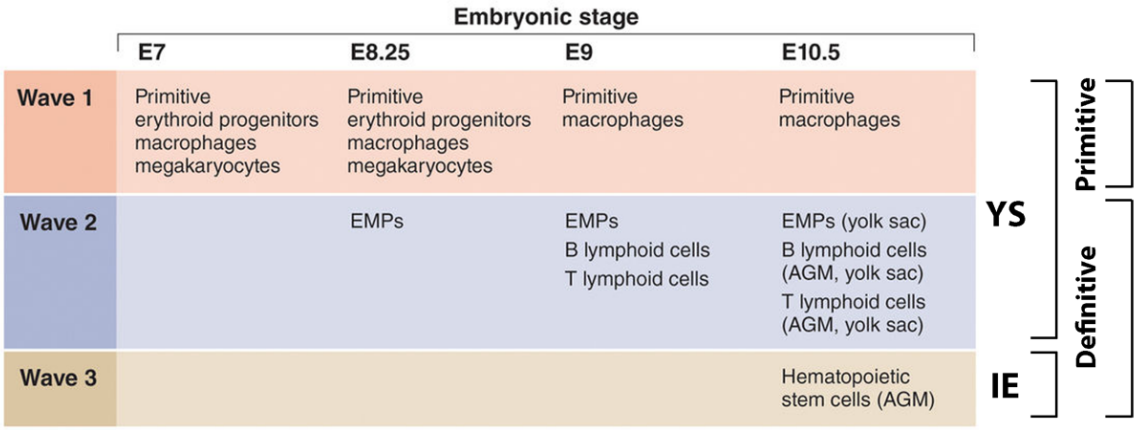


Figure 1-3: Haematopoietic waves during murine embryonic development.

Haematopoiesis occurs in three waves that partly overlap both spatially and temporally. During the first wave initiated in the YS primitive erythroid progenitors, embryonic macrophages and megakaryocytes are generated. The second wave initiates also in the YS around 1 day later while primitive haematopoiesis is still ongoing. Definitive erythro-myeloid progenitors (EMPs) as well as cells with lymphoid potential are generated. The third wave of haematopoieis occurs intra-embryonically mainly in the aorta-gonad-mesonephros (AGM) region *de novo* generating transplantable HSCs. Haematopoietic progenitor cells from all sites seed the foetal liver (FL), which remains the primary site of hematopoiesis until HSCs colonize the bone marrow (BM) during late gestation. IE: intra-embryoni; YS: yolk sac. Adapted from Yoder (2014).

1.2.2 Intra-embryonic haematopoiesis

The third wave of haematopoiesis initiates intra-embryonically with the emergence of haematopoietic intra-aortic clusters (IAC) and similar clusters in the vitelline and umbilical arteries around E10.5 (Figure 1-3) (de Bruijn, 2000; De Bruijn et al., 2002; Gordon-Keylock et al., 2013; Medvinsky and Dzierzak, 1996; Müller et al., 1994). Such IACs contain *de novo* generated HSCs that emerge, similar to EMPs, from an HE (Boisset et al., 2010; North et al., 2002). The aortic HE is found in the ventral aspects of the dorsal aorta (DA) (De Bruijn et al., 2002; North et al., 2002; Taoudi and Medvinsky, 2007). The process in which these cells undergo further haematopoietic differentiation, accompanied by the budding out of the DA lining, is called the endothelial to haematopoietic transition

(EHT) (Boisset et al., 2010; Kissa and Herbomel, 2010; Swiers et al., 2013a). Newly formed HSCs are contained in a population co-expressing haematopoietic and vascular markers including CDH5 (VE-cadherin), CD45, c-KIT, Sca-1 and CD31 (Bertrand et al., 2005; De Bruijn et al., 2002; Taoudi, 2005).

Even though further studies identified the placenta and other tissues, including the head, as sites where HSCs can be localized (Gekas et al., 2005; Li et al., 2012; Ottersbach and Dzierzak, 2005), *ex vivo* modelling demonstrated an autonomous potential to generate HSCs solely for the DA-containing AGM region and major extraembryonic vessels (Medvinsky and Dzierzak, 1996; Rybtsov et al., 2014). However, cells that harbour the capacity to engraft neonatal recipients can be detected already around E9.5, not only in the AGM region but also in the YS, suggesting that such cells can fully mature within a foetal environment *in vivo* (Yoder et al., 1997a; Yoder et al., 1997b). Similarly, $CDH5^+$ cells of the AGM region of E9.5-E10.5 embryos could be converted into transplantable HSCs *in vitro*. Based on such optimized culturing conditions, a hierarchical directionality for the developing HSC lineage could be established. Thus, pro-HSCs (E9.5 $CDH5^+CD41^{lo}CD45^-$) mature into pre-HSCs type-I (E10.5 $CDH5^+CD41^{lo}CD45^-$) and further into pre-HSC type-II (E10.5 $CDH5^+CD45^+$) (Rybtsov et al., 2011; Rybtsov et al., 2014; Taoudi et al., 2008). Both *in vivo* and *ex vivo* analysis have detected transplantable HSCs only in single digits in the murine conceptus between E10.5 and E11.5 (Gekas et al., 2005; Kumaravelu et al., 2002; Taoudi et al., 2008). Over a period of the next 24 hours their numbers increase exponentially when the site of haematopoiesis switches to the FL (Kumaravelu et al., 2002). Even though an expansion of the HSC population might involve proliferation by symmetric cell division, it was shown that the predominant mechanism is the acquisition of the HSC potential by maturing pre-HSCs (Kumaravelu et al., 2002; Taoudi et al., 2008). The FL itself does not generate haematopoietic cells *de novo* but is rather involved in further maturation and expansion of HSCs (Gribi et al., 2006; Kumaravelu et al., 2002; Takeuchi et al., 2002). Until late gestation, the FL remains the primary site of haematopoiesis, when HSCs are mobilized into the bloodstream and home to the BM compartment at E17, the site of adult haematopoiesis (Coskun and Hirschi, 2010; Yoder, 2014)

1.3 The zebrafish as a model system

The zebrafish (*Danio rerio*) has become a powerful model organism for the study of embryonic development and represents one of the most advanced fish models due to its robustness and easily accessible genome (Dooley and Zon, 2000). These and other properties make the zebrafish model suitable for genetic analysis by forward or reverse genetic screens (Patton and Zon, 2001). A plethora of mutant phenotypes have been uncovered, many of which resemble human clinical disorders, thus complementing studies in higher vertebrate systems (Jing and Zon, 2011). Furthermore, the zebrafish model has contributed to several examples of successful phenotype-based drug screens (MacRae and Peterson, 2015).

Advantages provided by the zebrafish model system are the relative ease with which they can be maintained and bred, as well as the external development of the embryos. Due to their rapid development, organogenesis is completed within 5 days (Kimmel et al., 1995). Zebrafish embryos can be precisely staged and their transparency enables *in vivo* analysis with fluorescent markers in a non-invasive manner (Clark et al., 2011; Kimmel et al., 1995), hereby greatly facilitating the observation and imaging of developmental processes. The large number of embryos generated at weekly intervals provides researchers with a good supply of material and makes large scale screens feasible (MacRae and Peterson, 2015). Embryos are relatively robust and therefore amenable to microinjection and genetic modifications. In this way, genes can be easily knocked down with antisense morpholinos, or overexpressed by simply introducing RNA or DNA (Summerton and Weller, 1997). The use of transposons for efficient transgenesis and the functionality of the CRISPR/Cas9 system have further extended the zebrafish genetic toolbox (Hwang et al., 2013; Suster et al., 2009).

Importantly, it has become increasingly clear that the genes and molecular signalling pathways controlling developmental processes are overwhelmingly conserved between fish and mammals (Chen and Zon, 2009; de Jong and Zon, 2005; Ransom et al., 1996; Zhang and Liu, 2011). In this respect, the zebrafish has proved particularly powerful and is widely used to study embryonic haematopoiesis (Chen and Zon, 2009)

1.4 Haematopoiesis in zebrafish

The process of haematopoiesis is generally conserved throughout vertebrate evolution and the zebrafish has become an excellent model system to study HSC development (Davidson and Zon, 2004; de Jong and Zon, 2005; Gering and Patient, 2005). The cellular events during early haematopoiesis, as well as during *de novo* HSC generation in the embryonic DA, have been described in detail (Bertrand et al., 2010a; Gering and Patient, 2005; Kissa and Herbomel, 2010; Kissa et al., 2008; Murayama et al., 2006; Wilkinson et al., 2009). Similar to mammals, zebrafish haematopoiesis occurs in three independent waves with the distinction that in zebrafish all the processes happen within the embryo proper.

The primitive haematopoietic wave occurs around 12 to 24 hpf at two anatomically distinct sites that are the anterior lateral mesoderm (ALM) and the posterior lateral mesoderm (PLM) (Davidson and Zon, 2004). These tissues are characterized by bilateral stripes of specialised mesoderm (Figure 1-4 A). Primitive myeloid cells are predominantly generated in the ALM that will also give rise to the head vasculature (Herbomel et al., 1999; Lieschke et al., 2002), whereas the PLM is the source of primitive erythrocytes and additionally contains precursors of the axial vessels of the trunk and tail (Detrich et al., 1995; Fouquet et al., 1997; Gering et al., 1998). The genetic programme underlying erythropoiesis is conserved in that it requires most of the genes also utilised during primitive haematopoiesis in mammals including *tal1* (*scl*), *gata1* and *lmo2* (Dooley et al., 2005; Lyons et al., 2002; Patterson et al., 2007). Likewise, the generation of primitive myeloid cells in the ALM is dependent on *spi1b* (*pu.1*) and *cebp1* (Hsu et al., 2004; Lyons et al., 2001). During the primitive wave, cells of the PLM including the primitive erythrocytes and cardinal vein (CV) precursors migrate inwards and reach the midline by around 18 somites (17 hpf), where they will form the intermediate cell mass (ICM), a structure that is functionally analogous to the yolk sac blood islands in the mouse (Detrich et al., 1995). The migration of these cells is preceded by prospective DA angioblasts that reach the midline first at around 15 somites (16 hpf), where they eventually coalesce to a cord structure underneath the notochord and hypochord and above the future ICM (Fouquet et al., 1997). Of note, DA and CV cells can already be distinguished by the 7–12 somite stage (13–15 hpf) (Zhong et al.,

2001). By 24 hpf, a primitive vascular loop is established and primitive erythrocytes enter circulation through the venous part of the system.

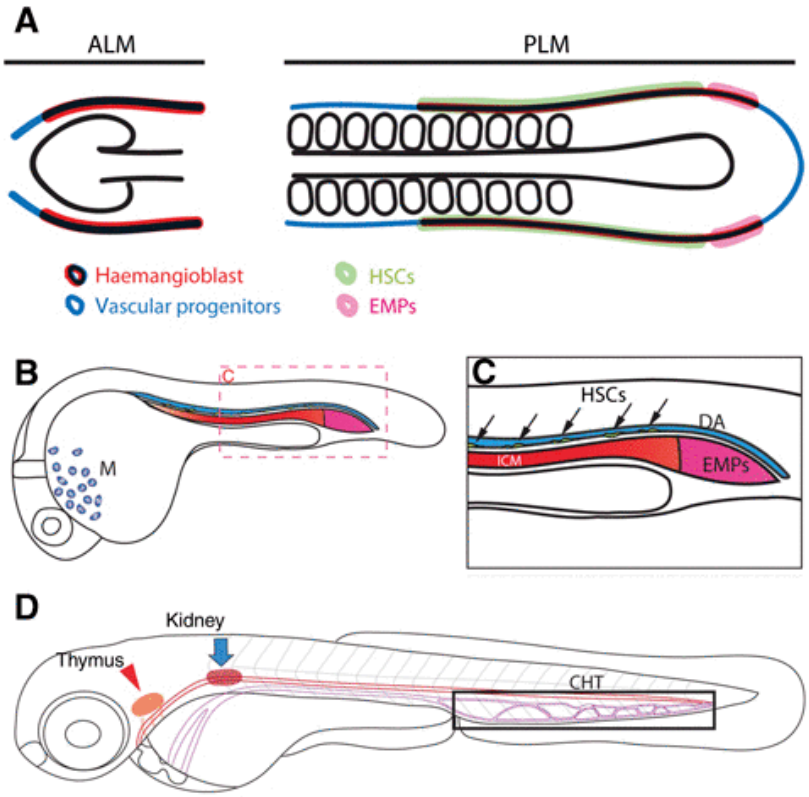


Figure 1-4: Haematopoietic sites in zebrafish.

A) Schematic fate map of the anterior lateral mesoderm (ALM) and posterior lateral mesoderm (PLM) at 10 somites represented as a flat mounted embryo with anterior to the left, viewed from the dorsal aspect. B) Schematic of the developing trunk around 24 hpf with a magnification view in (C) indicating the location of the intermediate cell mass (ICM) that predominantly contains PLM-derived primitive erythrocytes. Definitive HSPCs emerge from the ventral wall of the dorsal aorta (DA) located right above the ICM. Between 24-30 hpf erythro-myeloid progenitors (EMPs) arise from the posterior most region of the PLM (which becomes the posterior blood island (PBI)). D) HSPCs that emerge out of the aortic HE migrate to the caudal haematopoietic tissue (CHT) as an intermediate niche. From 3 dpf onwards HSPCs migrate first to the thymus to give rise to T-lymphocytes, and from 4 dpf they colonize the kidney which is the site of adult haematopoiesis. Adapted from Monteiro et al. (2011).

Definitive haematopoiesis initiates in zebrafish with the second haematopoietic wave between 24-30 hpf, when EMPs that harbour a limited proliferative potential arise from the most posterior PLM that becomes the posterior blood island (PBI) (Bertrand et al., 2007). EMPs are defined by the co-expression of fluorescent transgenes driven by the *lmo2* and *gata1* promoters (Bertrand et al., 2007).

These cells are presumably directly derived from *lmo2*$^{+}$ PLM derivatives and are analogous to yolk sac-derived murine EMPs (Bertrand et al., 2007; Bertrand et al., 2010b). EMPs were never observed to populate the thymus or kidney, indicating a lack of lymphoid potential. In contrast to the development of HSCs, the development of such EMPs was shown to be independent of the Notch signalling pathway (Bertrand et al., 2010b).

The third wave of zebrafish haematopoiesis initiates with the emergence of haematopoietic stem and progenitor cells (HSPCs) arising from the haemogenic endothelium located in the ventral wall of the DA (Bertrand et al., 2010a; Kalev-Zylinska et al., 2002; Kissa and Herbomel, 2010; Lam et al., 2010; Zhang and Rodaway, 2007). Since it is difficult to discriminate between true HSCs and multi-potent progenitors in zebrafish, these cells are here cumulatively termed HSPCs. The aortic HE in zebrafish utilises a similar genetic programme to that seen in the mouse and can be identified by the expression of genes including *gata2b*, *runx1* and *ikzf1* (*ikaros*) (Butko et al., 2015; Gering and Patient, 2005). From around 32 hpf onwards such HSPCs are generated through the process of EHT. During this process, expression of the haematopoietic marker *itga2b* (*CD41*) can be detected in fluorescent reporter lines (Bertrand et al., 2008; Lin et al., 2005). While in the murine system, an inwards oriented budding of the HE is observable followed by the formation of IAC (Boisset et al., 2010), in zebrafish cells of the HE egress out of the DA into the sub-aortic space (Bertrand et al., 2010a; Kissa and Herbomel, 2010; Lam et al., 2010). From there, HSPCs enter circulation by passing through the wall of the CV that is just beneath the DA. These striking differences from the murine system might reflect species-specific requirements, probably caused by the rapid development of zebrafish. In addition, the zebrafish DA is simply too thin in diameter to harbour IAC, which would otherwise clog the DA lumen. To date it is not known if differentiation processes that occur during the IAC stage in mouse and other mammals are not required for zebrafish HSC development or if they are initiated at different times, either already during the HE stage or later in subsequent haematopoietic niches. Through blood circulation, HSCs home to and enter the caudal haematopoietic tissue (CHT), a vascular plexus that fulfils equivalent maturational purposes for HSC development as the FL in mammals (Murayama et al., 2006). Within the CHT the HSPC population expands and partially undergoes a limited lineage commitment into erythroid and myeloid but not lymphoid progenitors (Monteiro et al., 2011;

Murayama et al., 2006). Consequently, no progenitor marker is specific to the HSC including *cmyb* (Murayama et al., 2006). However, intensities of fluorescent transgenes under control of the *itga2b* promoter is used to identify HSPCs, with low levels reflecting HSPCs and high levels thrombocytes (Bertrand et al., 2008; Lin et al., 2005). After proliferation and maturation, HSPCs leave the CHT and seed subsequent haematopoietic niches, either to the thymus around 3.5 dpf or the kidney marrow around 4.5 dpf (Kissa et al., 2008; Murayama et al., 2006). The thymus is the tissue where HSPCs generate T-lymphocytes, whereas true HSCs in the kidney sustain life-long haematopoiesis (Murayama et al., 2006).

Interestingly, one study reported that some early HSPCs that had just egressed from the DA, enter the pronephric ducts from where they colonize the pronephros as an alternative migration route (Bertrand et al., 2008). Recently, live imaging has also highlighted that some cells of the ventral DA bud apically into the DA lumen, a process seen only up to 40 hpf (Thambyrajah et al., 2016). Furthermore, new strategies of fate mapping revealed that adult microglia (primary immune cells of the central nervous system, similar to macrophages) arise from the ventral DA in a Runx1-dependent manner, suggesting that the HE is a heterogeneous tissue giving rise to blood progenitors with different potentials (Xu et al., 2015).

1.5 The aortic haemogenic endothelium

Intensive research has revealed that definitive blood progenitors (both EMPs and HCSc) emerge during a narrow developmental window from an endothelial precursor with haematopoietic potential that is termed “haemogenic endothelium” (Bertrand et al., 2010a; Boisset et al., 2010; de Bruijn, 2000; Kissa and Herbomel, 2010; Lam et al., 2010; Swiers et al., 2013a). While the HE that generates HSCs is always found in the lining of the major embryonic arteries (de Bruijn, 2000), EMPs emerge prior to the establishment of arborized arterial and venous vasculature in the YS which is generally independent of arterial identity (Frame et al., 2016). Likewise, exploiting the advantages of the amphibian model system, it was demonstrated that haematopoietic lineages in the ventral blood island

(equivalent to the YS) and AGM-derived haematopoietic cells originate from distinct blastomeres (Ciau-Uitz et al., 2000). Thus, both HSCs and trunk vasculature - including the part of the DA that eventually will give rise to the HE - emerge from a common blastomere, indicating a close lineage relationship (Ciau-Uitz and Patient, 2016; Ciau-Uitz et al., 2000). These accumulating data suggest that not all HE cells are created from the same endothelial precursor population and that different types of HE exist; one that leads to the emergence of transient progenitor cells (YS HE) and one that can generate HSCs (aortic HE) (Hirschi, 2012).

Mouse studies have shown that haematopoietic clusters that protrude from endothelium are also found in vitelline and umbilical arteries, YS, placenta and surprisingly in the head and heart (Gordon-Keylock et al., 2013; Li et al., 2012; Nakano et al., 2013; Rhodes et al., 2008; Yokomizo and Dzierzak, 2010; Zovein et al., 2008), however, the aortic HE as a site for HSC *de novo* generation is found to be conserved across vertebrates and is by far the best understood (Antas et al., 2013; Swiers et al., 2013a). Here, dissection studies localized the aortic HE to the ventral DA, a position that has also proved to be conserved across species (Ciau-Uitz et al., 2000; Gering and Patient, 2005; Richard et al., 2013; Taoudi and Medvinsky, 2007). Aortic HE expresses arterial genes such as *Sox17* and *Notch1* (Lizama et al., 2015) and time-lapse imaging studies visualized the transition of endothelium into blood in real-time *in vitro* (Eilken et al., 2009) as well as by live imaging *in vivo* in zebrafish (Bertrand et al., 2010a; Kissa and Herbomel, 2010; Lam et al., 2010) and *ex vivo* in mouse aortic explants (Boisset et al., 2010).

In addition to its endothelial phenotype and morphology, HE can be functionally defined by its capacity to form both haematopoietic offspring and endothelial tubules in culture (Nishikawa et al., 1998a; Nishikawa et al., 1998b). Phenotypically, HE cells express the endothelial surface markers CD31, VE-Cadherin (CDH5) and KDR (FLK1) (Nishikawa et al., 1998b; Shalaby et al., 1997), as well as haematopoietic markers including CD41, CD45 and Ter119 (Corbel and Salaün, 2002; Ferkowicz, 2003; Kina et al., 2000; Ledbetter and Herzenberg, 1979; Mikkola et al., 2003). However, the most prominent marker for HE is the TF RUNX1, whose expression precedes and is essential for the emergence of HSCs (Chen et al., 2009; North et al., 1999). Studies that analysed the transcriptional regulation of *Runx1* identified that +23 enhancer-driven GFP expression marks all HE cells in the

embryo (Bee et al., 2010; Nottingham et al., 2007). While these described cellular features are all common for both YS and aortic HE, some crucial differences have been identified, further supporting the notion that not all HE cells are created equally: by using different tissue-specific rescue cassettes in *CBFb*$^{-/-}$ embryos, that phenocopy *Runx1*$^{-/-}$ mutants lacking EMP as well as HSC generation (Chen et al., 2009; North et al., 1999), it was shown that a *Tek* (*Tie2*)-controlled expression of CBFb-GFP rescues EMPs but not HSCs, while a *Ly6a* controlled expression did not recue EMPs but HSCs (Chen et al., 2011).

Ly6a encodes the rodent specific cell surface marker Sca-1 that is not found in other vertebrate species including human, frog, chicken or zebrafish (Spangrude et al., 1988b), but that marks all murine HSCs in the FL and BM (De Bruijn et al., 2002; Ottersbach and Dzierzak, 2005). Even though Sca-1 is present on only a subset of newly emerging HSCs in the E11.5 AGM region, expression of GFP from the *Ly6a* regulatory sequences marks progenitors with lymphoid potential and all functional HSCs in the DA, FL and placenta (De Bruijn et al., 2002; North et al., 2002; Ottersbach and Dzierzak, 2005). Ly6a-GFP^{+} cells are found in a subset of the aortic endothelium and have been tracked by live imaging to bud into the lumen of the DA in slice cultures from E10.5 foetuses (Boisset et al., 2010). By contrast, *Ly6a* driven transgene expression is not detected in EMPs (Li et al., 2014), highlighting molecular differences between YS and aortic HE.

At the transcriptional level a number of TFs are essential for the development of blood and HSCs, including members of the ETS family as well as GATA2, TAL1 and RUNX1 (Ciau-Uitz et al., 2010). While *Runx1* expression is detected in established HE and is required for the EHT process, critical functions of the other TFs are placed upstream of RUNX1. GATA2 and TAL1 are recurrently used during embryonic and adult haematopoiesis (Liao et al., 1998; Porcher et al., 1996; Porcher et al., 1999; Tsai et al., 1994), and can interact with ETS TFs to generate self-regulated genetic circuits (Göttgens et al., 2002; Patterson et al., 2007; Pimanda et al., 2007a; Wilson et al., 2010a). TAL1 was shown to be critically required for generating HE (Patterson et al., 2005; Zhen et al., 2013). Likewise, GATA2 was shown to be essential for formation of HSCs and shown to regulate *Runx1* expression (Butko et al., 2015; Robert-Moreno et al., 2008). Haemogenic expression of *Runx1* driven by the +23

enhancer element integrates the GATA, ETS and TAL1 network, thus converging most of the discovered haematopoietic transcriptional inputs (Nottingham et al., 2007).

While the murine model was valuable to define the HE and its properties, the zebrafish model proved useful to determine the signals required to generate the aortic HE. Here, the HE can simply be identified by the expression of *runx1* in *kdrl*$^{+}$ cells of the ventral DA (Gering and Patient, 2005). Such studies have shown that Hedgehog signaling is required at multiple stages of blood cell formation (Gering and Patient, 2005), including arterial specification. Inhibition of hedgehog signaling severely reduces the number of *runx1*$^{+}$ cells in the DA. Additionally, VEGF and Notch signaling act upstream of *runx1* by controlling arterial specification (Burns et al., 2005; Gering and Patient, 2005). A transient Notch activation during embryogenesis using an inducible transgenic system expanded *runx1* expression (Burns et al., 2005). The regulation by Notch might be more complex, since different Notch signaling strengths were shown to define an arterial or haemogenic outcome (Gama-Norton et al., 2015). Furthermore, different Notch inputs might be required at multiple stages. In zebrafish, early somite-derived Notch signals through the ligands Dlc and Dld, regulated by Wnt16 (Clements et al., 2011), establish HSC fate as vascular precursors migrate across the ventral face of the somites (Kobayashi et al., 2014). Later, the Notch ligand Jag1a, which is a target of TGFβ signaling (Monteiro et al., 2016), is required for haemogenic expression of *runx1*, by controlling expression of its upstream regulator Gata2b (Butko et al., 2015; Robert-Moreno et al., 2008).

One of the localized signals that might define the ventral positioning of the HE within the DA is BMP signaling. Expression of the *Bmp4* ligand is found in the mesenchyme ventral to the DA in multiple species (Durand et al., 2007; Marshall et al., 2000; Pimanda et al., 2007b). In zebrafish, inhibition of Bmp4 signaling blocks both the initiation and maintenance of *runx1* expression (Wilkinson et al., 2009), while increased Bmp4 signaling causes an extended expression of *runx1* into the dorsal regions of the DA (Pouget et al., 2014). The activity of BMP in the sub-aortic mesenchyme was shown to be regulated by FGF signaling (Pouget et al., 2014). While localized BMP signaling sets up the HE, its later down-regulation during HSC development was shown to be required (Souilhol et al., 2016). Unlike Hedgehog and Notch signaling, Bmp4 appears to be unnecessary for arterial specification (Wilkinson et al., 2009).

Surprising discoveries in the field were that also the inflammatory response pathway and the nervous system are required for HSC emergence. In the case of the inflammatory signalling, primitive neutrophils have been determined as a source of TNFα that leads to an activation of NF-κB and the Notch signalling pathway via TNFR2 in developing HSPCs (Espín-Palazón et al., 2014). Similarly, Tlr4 was shown to be required for NF-κB expression enhancing the Notch pathway (He et al., 2015). Furthermore, inflammatory cytokines including IFNγ and IFNα impact HSC numbers and maturation (Kim et al., 2016; Li et al., 2014; Sawamiphak et al., 2014). In the case of the nervous system, signalling via catecholamine secreted from sympathetic nerves that develop adjacent to the DA has also been shown to regulate HSC development (Fitch et al., 2012).

Due to its restricted expression in the zebrafish DA, *runx1* has been widely used as a readout for functional screens in studies that aimed to identify novel factors regulating HSPC generation (Burns et al., 2009; Huang et al., 2013; North et al., 2007; North et al., 2009). In this way, the role of the prostaglandin E2 pathway in HSPC formation was elucidated (North et al., 2007). Furthermore, such studies revealed that haemodynamic forces are required to maintain the HSC programme but dispensable for its initiation (Adamo et al., 2009; North et al., 2009; Wang et al., 2011). This effect is at least partly mediated by blood flow-induced expression of the TF *klf2a* that regulates Nitric Oxide (NO) signalling required for arterial maturation and continuation of the HSC programme (Wang et al., 2011).

1.6 The transcription factor Runx1

The family of Runt-related TFs (RUNX) consists of three members - RUNX1, RUNX2 and RUNX3 - that have tissue-specific non-redundant functions throughout development (Ito et al., 2008; Komori and Kishimoto, 1998; Okuda et al., 1996; Wang et al., 1996a) and are key regulators of lineage-specific gene expression (Chuang et al., 2013; de Bruijn and Dzierzak, 2017). RUNX proteins form a functional heterodimeric complex with the co-factor CBFb that binds to the highly-conserved Runt domain (RD) and enhances the DNA-binding ability (Ogawa et al., 1993; Tahirov et al., 2001). All

RUNX proteins bind to the same DNA motif (PyGPyGGTPy) through their RD located in the N-terminal region (Kamachi et al., 1990; Melnikova et al., 1993; Speck, 1995), and are involved in both gene activation and repression through recruitment of common transcriptional regulators (Levanon and Groner, 2004; Lichtinger et al., 2012; Wheeler et al., 2000; Wilson et al., 2010a). Downstream of the RD, the protein structures of the RUNX family members reveal less homology, probably accounting for functional differences.

RUNX1 is also known as acute myeloid leukemia 1 protein (AML1) or core-binding factor subunit alpha-2 (CBFa2) and has pleiotropic roles during development. During haematopoiesis RUNX1 is essential for the generation of HSCs, and is involved in the differentiation of myeloid and lymphoid lineages (Kurokawa, 2006; Link et al., 2010; North et al., 1999). RUNX1 is also reported to have CBFb-independent functions and is able to weakly bind DNA on its own (Bresciani et al., 2014; Hoogenkamp et al., 2009; Wang et al., 2014). In addition to complexing with CBFb, RUNX1 can also interact with a multitude of other co-factors and chromatin modifiers (Figure 1-5).

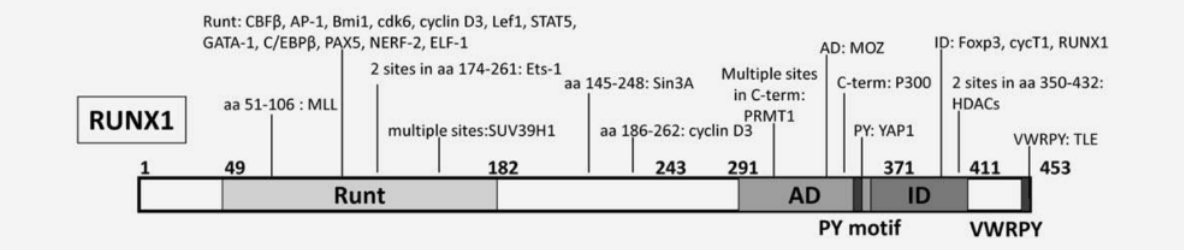


Figure 1-5: Structure of the human RUNX1 protein.
Schematic summarizing the location of functional domains of the RUNX1 protein as well as interacting proteins. The numbers indicate amino acid positions. Runt: Runt domain required for DNA binding and heterodimerization with co-factor CBFb. AD: activation domain, ID: inhibitory domain. RUNX1 interactants not shown here include ELF2, ELF4, HIPK1, HIPK2, SMRT, TAZ, MyoD, SWI-SNF, CHIP, TAL1, TRAF6 and SMAD proteins. Adapted from Chuang et al. (2013)

RUNX1 gene expression is initiated by two alternative promoters, P1 (distal) and P2 (proximal) (Ghozi et al., 1996; Levanon et al., 1996; Marsman et al., 2014; Telfer and Rothenberg, 2001), generating a series of transcripts that differ in their protein-coding exons and untranslated regions,

which in turn influences mRNA stability, efficiency of translation and/or RUNX1 protein structure (Bee et al., 2010; Levanon and Groner, 2004; Levanon et al., 2001; Pozner et al., 2000). While the P2 promoter is nested within a CpG island that is conserved down to Fugu, the P1 promoter region lacks such a CpG-rich region (Eggers et al., 2002; Levanon et al., 2001). A highly conserved haematopoietic-specific enhancer located in the first intron 23.5 kb downstream of the P1 promoter (+23) is conserved across vertebrate species down to frog and confers activity to both *RUNX1* promoters (Bee et al., 2009; Nottingham et al., 2007).

Further mechanisms that regulate RUNX1 activity involve post-translational modifications (Bae and Lee, 2006; Goyama et al., 2015; Levanon et al., 1996; Wang et al., 2009a). Among these, phosphorylation, acetylation and methylation promote transcriptional activity of RUNX1 (Aho et al., 2006; Bae and Lee, 2006; Goyama et al., 2015; Tanaka et al., 1996; Wang et al., 2009a; Yamaguchi et al., 2004; Zhang et al., 2004; Zhao et al., 2008). Phosphorylation and ubiquitination further influence RUNX1 protein stability (Bae and Lee, 2006; Biggs et al., 2006; Kurokawa et al., 2004; Wang et al., 2009a). Similarly, binding by CBFb protects RUNX proteins from proteasomal degradation and enhances its activity (Huang et al., 2001).

1.6.1 Role of RUNX1 during haematopoiesis

Runx1 expression is found in several tissues including neuronal cell types, muscle cells and throughout the haematopoietic system including HSCs (Chen et al., 2006; Kramer et al., 2006; Theriault, 2005; Wang et al., 2005; Zhu et al., 1994). All adult blood cells except erythrocytes express *Runx1* (Lorsbach et al., 2004; North and de Bruijn, 2004). Its critical association with the regulation of haematopoietic processes can be emphasised by its frequent role in haematopoietic malignancies. Together with its binding partner CBFb, RUNX1 is a frequent target for chromosomal translocations, inversions and point mutations in AML and other forms of leukaemia, that together account for about 30% of human leukaemia (Ito, 1999; Look, 1997; Lowenberg et al., 1999; Rowley, 1998; Speck, 1995; Speck et al., 1999). Furthermore, *Runx1* haploinsufficiency is associated with a familial platelet disorder with a propensity to develop into acute myelogenous leukaemia (FDP/AML) (Song et al., 1999).

Mice deficient for RUNX1 die at E12.5 and show substantial haemorrhaging, mainly in the central nervous system, closely phenocopied by the *Cbfb*$^{-/-}$ mutants (Okuda et al., 1996; Sasaki et al., 1996; Wang et al., 1996a; Wang et al., 1996b). The blood of *Runx1*$^{-/-}$ mutants consists solely of primitive erythrocytes, while definitive blood cells are completely lacking at all sites of haematopoieis, including the AGM region where most HSCs are initally formed (Mukouyama et al., 2000; North et al., 1999; Okuda et al., 1996; Rhodes et al., 2008; Sasaki et al., 1996; Wang et al., 1996b).

During murine primitive haematopoiesis, *Runx1* is already expressed in the mesodermal fraction of the presumptive blood islands that generate primitive blood from E7.5 onwards (North et al., 1999). Later, haematopoietic expression of *Runx1* precedes the emergence of definitive haematopoietic progenitors and HSCs in a subset of endothelial cells that can be identified as HE (Ottersbach and Dzierzak, 2005; Rhodes et al., 2008; Speck et al., 1999). The expression of *Runx1* in the aortic HE in the ventral aspects of the DA is conserved down to zebrafish (Bollerot et al., 2005; Ciau-Uitz et al., 2000; Patterson et al., 2005; Speck et al., 1999). *Runx1* expression is also found in sub-aortic mesenchymal cells that are co-expressing smooth muscle actin, suggesting that they are not part of the haematopoietic lineage (Swiers et al., 2013a).

One key finding was that *Runx1*$^{-/-}$ mutant mouse embryos lack IACs. Further studies utilizing the zebrafish embryo or the ESC system revealed that RUNX1 orchestrates the EHT process through which HE develops into HSC (Kissa et al., 2008; Lancrin et al., 2012). Additionally, EMPs are lost in *Runx1*$^{-/-}$ mutant YS, consistent with the idea that EMPs, like HSCs, emerge from HE (Yokomizo et al., 2001). The requirement for RUNX1 during HSPC development becomes obsolete during the FL stage as deletion with conditional *Vav1:Cre*, active in haematopoietic progenitors and HSCs in the FL from E11.5 onwards, does not block HSC formation (Chen et al., 2009). Follow-up studies based on tamoxifen-inducible deletions determined the time point at which HSC development becomes RUNX1-independent as ~E11.5 (Tober et al., 2013). Strikingly, between E10.5 and E11.5 most HE cells have already undergone EHT, suggesting that RUNX1 not only initiates EHT but further programmes the haematopoietic fate (Tober et al., 2013). *In vitro* studies utilizing *Runx1*$^{-/-}$ ES cultures similarly showed the crucial role of RUNX1 for EHT. Here, cells in a haematopoietic differentiation culture develop to an HE-like stage at which they form flat, adherent layers. While in the Wt situation

such HE-like cells undergo EHT, leading to the appearance of floating haematopoietic precursor cells, this is absent in the *Runx1*$^{-/-}$ condition. However, when EHT was rescued by introducing the RUNX1 target *Gfi1,* the full haematopoietic programme was not re-established indicating additional roles for RUNX1 in programming HSCs (Lancrin et al., 2012).

In parallel to the defects observed for definitive haematopoiesis, the loss of RUNX1 affects primitive blood development in that the generation of primitive macrophages is inhibited (Chen et al., 2006; Lacaud et al., 2002) and the differentiation of the megakaryocyte and platelet lineage is impaired (Okuda et al., 1996; Wang et al., 1996a). Formation of primitive erythrocytes however occurs in both *Runx1*$^{-/-}$ and *Cbfb*$^{-/-}$ mutants (Lacaud et al., 2002; Okuda et al., 1996; Sasaki et al., 1996; Wang et al., 1996b), despite exhibiting abnormal morphology, delayed maturation and altered expression of the erythroid marker Ter119 and of the TFs like EKLF and GATA1 (Yokomizo et al., 2008).

Compared to its crucial role during embryonic haematopoiesis, the effects of conditional deletion of *Runx1* from adult HSCs is less dramatic. Here, loss of RUNX1 expands the number of LSK cells that futhermore display decreased apoptosis (Cai et al., 2011; Chen et al., 2009; Growney et al., 2005; Ichikawa et al., 2004; Ichikawa et al., 2008; Motoda et al., 2007). Expansion of the LSK pool occurs despite the finding that RUNX1 deficiency results in a slow growth, accompanied by markedly decreased ribosome biogenesis (Cai et al., 2015). Loss of RUNX1 was also found to increase the stress resistance and to result in superior survival in the presence of genotoxic stress (Cai et al., 2015). HSCs from *Runx1*$^{-/-}$ mutant mice are sensitized to leukaemia caused by experimentally induced secondary mutations (Jacob et al., 2010).

1.6.2 Mechanisms of Runx1 action

RUNX proteins have been shown to act as either activators or repressors of transcription and domains mediating those activities have been mapped to sequences N- and C-terminal to the RD (Kanno et al., 1998). RUNX proteins themselves are not strong transcriptional activators/repressors but evoke their capacity through interaction with several other transcriptional regulatory proteins.

The RUNX1 activation domain (AD), mapped between AA 291-371 (Kanno et al., 1998), interacts with co-activators including MOZ, P300, YAP and CBP (Kitabayashi et al., 1998; Kitabayashi et al., 2002; Yagi et al., 1999) (see also Figure 1-5). Other TFs, including PAX5, SOX9, STAT5 and AP-1 interact with RUNX1 via the RD (D'Alonzo et al., 2002; Libermann et al., 1999; Ogawa et al., 2008). RUNX1 further interacts with GATA1, GATA2, ETS factors, C/EBPα, HES-1, ALY and SMAD proteins (Elagib et al., 2003; Hanai et al., 1999; Hannah et al., 2011; Kim et al., 1999; McLarren et al., 2000; Sauntharajah et al., 2006; Wotton et al., 1994). Recent studies further indicate that the DNA binding ability of RUNX proteins is also subjected to interactions with DNA binding factors other than and independent of its co-factor CBFb (Kim et al., 1999). A well-known example is the synergistic interaction between RUNX1 proteins and ETS1 (Jonsen et al., 1996) that cooperate in DNA binding on common target enhancer elements (Goetz et al., 2000; Gu et al., 2000; Wotton et al., 1994). In this case, the RUNX1 auto-inhibitory domain (ID) and the one possessed by ETS1 engage in mutual binding, thereby exposing the DNA binding surfaces and potentially the proteins' transactivation domains too (Kim et al., 1999).

Repressive function can be conducted by interaction with the TLE1 (Groucho) class of co-repressors recruited through a conserved VWRPY motif at the C-terminal end of RUNX (Levanon et al., 1998). Groucho-independent mechanisms are also described that involve interactions between a region of RUNX1 C-terminal to the RD and the SIN3A co-repressor (Lutterbach et al., 2000). RUNX1 further associates with HDACs and interacts with the histone-lysin N-methyltransferase SUV39H1 to evoke transcriptional repression (Guo and Friedman, 2011; Reed-Inderbitzin et al., 2006). RUNX1 was also shown to interact directly with BMI1, a component of the Polycomb-Repressive Complex 1 (PRC1), in megakaryocytic and T-lymhocytic cells and to recruit PRC1 to chromatin independent of previous PRC2 activity (Simon and Kingston, 2009; Yu et al., 2012).

Even though a large number of RUNX1 target genes have been reported, including essential components of lineage-specific differentiation programmes as well as regulators of basic cellular processes and cell-cycle control (Hug et al., 2004; Michaud et al., 2008; Okada et al., 1998; Otto et al., 2003; Sakai et al., 2009; Wotton et al., 2008), the number of targets for embryonic haematopoiesis

or HSC *de novo* generation remain limited. The critical role of RUNX1 for EHT was shown by its direct regulation of the GFI1 and GFI1b genes, which in turn repress endothelial genes through the recruitment of the LSD1-CoREST complex (Lancrin et al., 2012; Thambyrajah et al., 2015). In this regard, it is not surprising that earlier studies found expression of endothelial markers, including *KDR* and *VE-cadherin*, to be down-regulated after *Runx1* activation in the HE (Hirai et al., 2003; Hirai et al., 2005; Jaffredo et al., 2005).

Since the repression of endothelial genes and the facilitation of EHT do not explain the complete function of RUNX1 during HSC development (Lancrin et al., 2012; Tober et al., 2013), it is further suggested that RUNX1 activates expression of genes that are blood-specific and/or are involved in establishing full haematopoietic competence. RUNX1 was shown to bind and facilitate expression of the haematopoietic commitment marker CD41 (Tanaka et al., 2012). However, in the absence of RUNX1, aortic HE cells are still capable of developing into phenotypical pre-HSC type-I cells defined by the expression of CD41 (Liakhovitskaia et al., 2014). A recent screen utilising the DNA adenine methyltransferase identification (DamID) technology to map RUNX1 binding sites using *in vitro* generated HE-like cells found RUNX1 as a positive regulator of the cell adhesion and migration programme, including components of the integrin signalling pathway (Lie-A-Ling et al., 2014). This was further exemplified by the RUNX1-dependent expression of CD61 that can heterodimerise with CD41 (Van der Flier and Sonnenberg, 2001). A microarray screen, performed with similarly *in vitro* generated HE-like cells, identified the uncharacterized gene *AI467606* as a *bona fide* RUNX1 target, that is expressed during the development of the haematopoietic system *in vivo* and *in vitro*, as well as in all haematopoietic cell lineages of adult mice, except for mature erythrocytes (Ferreras et al., 2011). However, whether the *AI467606* gene plays a role during haematopoiesis is not known yet. Recently, the zebrafish DNA methyltransferase *dnmt3bb.1* was found to be positively regulated by Runx1 in the HE where it is involved in maintaining expression of *cmyb* (Gore et al., 2016).

The *Spi1* (*Pu.1*) gene is another *bona fide* RUNX1 target, which is directly bound at its promoter and an additional 3' upstream regulatory element (URE) (Hoogenkamp et al., 2009; Huang et al., 2008a). *In vivo* analysis demonstrated that RUNX1 regulates *Spi1* expression lineage-dependently

either positively or negatively (Huang et al., 2008a). Using the mouse ESC differentiation model Hoogenkamp et al. (2009) showed that RUNX1 can bind the *Spi1* regulatory elements in a CBFb-independent manner at the onset of haematopoietic development. This binding was weak and unstable but induced local chromatin unfolding that became manifested even after RUNX1 was withdrawn (Hoogenkamp et al., 2009). The mechanistic details by which RUNX1 promotes this early chromatin unfolding are not known. Additionally, RUNX1 induction was shown to re-organize the binding pattern of haematopoietic TFs including FLI1, TAL1, and C/EBPα during haematopoietic differentiation and RUNX1 binding strongly increases the active chromatin mark H3K9Ac at both promoter and distal elements (Lichtinger et al., 2012).

Interestingly, the Groucho-dependent repressive function of RUNX1 for HSC development seems to be replaceable as deletion of the VWRPY domain does not affect definitive progenitor cell emergence (Goyama et al., 2004; Nishimura et al., 2004). By contrast, the AD domain is required for haematopoiesis, both *in vitro* and *in vivo* (Goyama et al., 2004; Nishimura et al., 2004; Okuda et al., 2000), implying that gene activation is here the predominant function, even though Groucho-independent repressive mechanisms have been described (Guo and Friedman, 2011; Lutterbach et al., 2000; Reed-Inderbitzin et al., 2006; Yu et al., 2012). As exemplified by the regulation of GFI1 and GFI1b, gene repression controlled by RUNX1 can also be indirectly achieved via the up-regulation of transcriptional repressors (Lancrin et al., 2012).

Overall, the current understanding of RUNX1 protein function is that of a context-dependent regulator which activates or represses transcription depending upon the organization of a particular promoter/enhancer in a specific cell type at a certain time (Wheeler et al., 2000). Furthermore, different transcriptional outcomes can be generated through combinatorial protein binding with various transcriptional regulators and epigenetic modulators. These characteristics are challenging for the understanding of RUNX1 function and will always demand a highly context-specific analysis.

Objectives and general experimental strategy

The importance of RUNX1 for the generation of HSPCs has been shown across most vertebrate model species (Kissa and Herbomel, 2010; North et al., 1999). To our knowledge, no global gene expression screen for the identification of RUNX1 regulated genes during definitive haematopoiesis *in vivo* has been performed yet. We aim to do so using the zebrafish model system (Figure 1-6 A). Here, the cellular processes leading to the formation of the aortic HE and the subsequent EHT are well-described and experiments can be precisely timed (Kissa and Herbomel, 2010; Murayama et al., 2006). Additionally, a simple combination of two marker genes, namely *kdrl* and *runx1*, define the HE population with high specificity from 24 hpf onwards (Figure 1-6 B) (Gering and Patient, 2005).

The overall experimental strategy is to generate a *runx1* reporter line that is based on a bacterial artificial chromosome (BAC) to drive fluorescent transgene expression that reliably recapitulates endogenous expression (Figure 1-6 C). When crossed with an endothelial reporter line *Tg(kdrl:mCherry),* it will allow the isolation of double-positive HE-cells via fluorescence activated cell sorting (FACS). The remaining endothelial population will be used as a reference tissue. Next-generation sequencing analysis will reveal the HE-specific set of genes.

Functional expression data can be generated by using a well-validated *runx1* morpholino (MO). Newly identified potential target genes will be cross-compared to equivalent murine gene-expression data-sets or ChIPseq studies for RUNX1 performed in haematopoietic cell lines (Kartalaei et al., 2015; Li et al., 2014; Li et al., 2016; Ptasinska et al., 2012; Wilson et al., 2010a). Gene expression analysis via ISH and qPCR in Wt and *runx1*$^{-/-}$ mutant embryos will finally validate their regulation by Runx1.

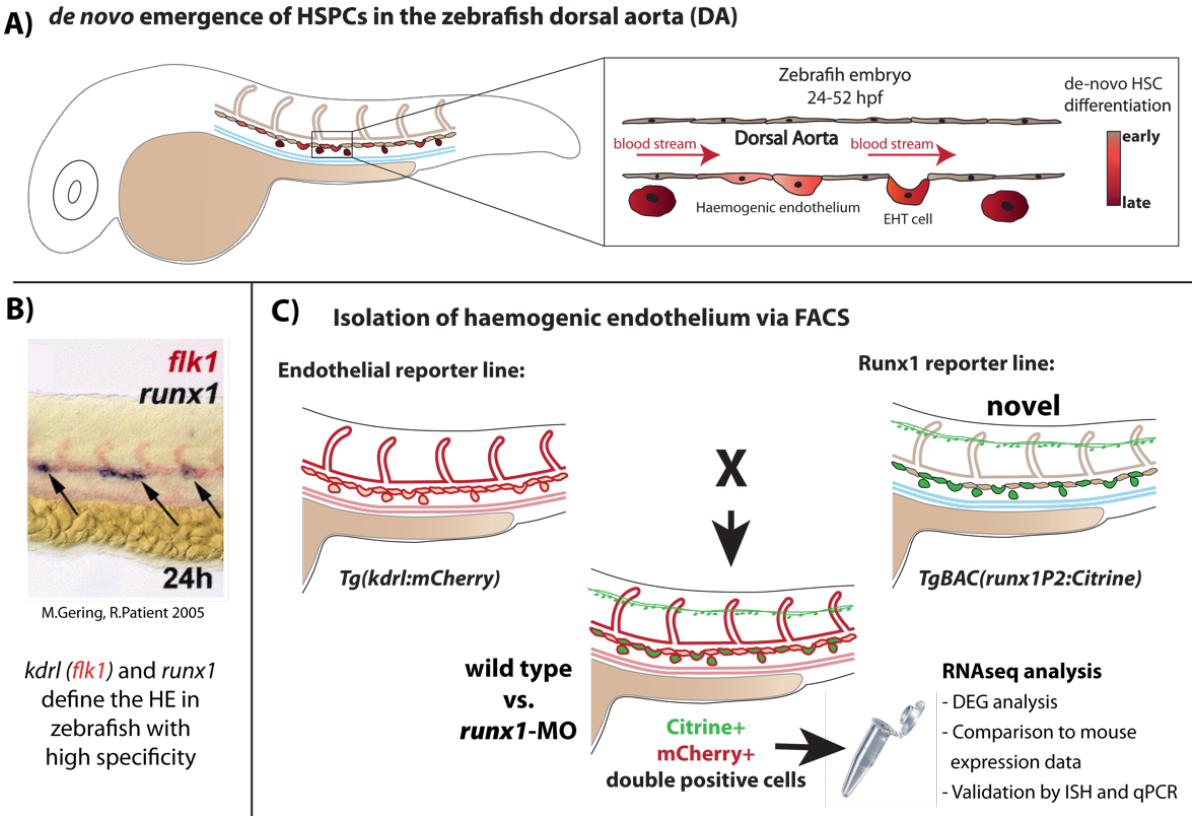


Figure 1-6: Experimental strategy to study global gene expression of the HE in vivo.

A) The HE is located in the ventral wall of the developing DA. These cells will eventually undergo a differentiation process in which they undergo EHT and bud into the sub-aortic space. B) Gene expression of *kdrl* and *runx1* define the HE with high specificity. Image adapted from Gering and Patient (2005). C) It is planned to isolate cells of the HE via FACS. For this, a well-described endothelial reporter line *Tg(kdrl:mCherry)* will be crossed to a runx1 reporter line. Gene expression of double-positive cells will be compared to the "general" endothelium as a reference. Functional data will be generated with a well-described *runx1* MO. The *runx1* reporter line has to be generated. DEG: differentially expressed genes.

Thesis aims

1. Generation of a novel *runx1* reporter line that allows for the reliable isolation of HE cells (Chapter 3).
2. Genome wide expression analysis of the HE and identification of potential Runx1 target genes (Chapter 4).
3. Detailed analysis of the consequences of the loss of Runx1 on the HE *in vivo* (Chapter 5).

- Chapter 2 -

Materials and Methods

2.1 Materials

2.1.1 Solutions

Solutions were made up with Milli-Q water and autoclaved as required. Their composition and storage conditions as used in this thesis are detailed in the table below.

Table 2-1: Solutions.

solution	composition	storage conditions
0.1M glycine	0.1 M glycine, p. H 2.2, 0.1% Tween	RT
Agarose gel	1-3% Agarose, 0.5x TEA	RT
AP Buffer	100 mM Tris-HCl pH 9. 5, 100 mM NaCl, 50 mM MgCl2, 0.1% Tween	prepared fresh
Bleaching solution	10% hydrogen peroxide, 5% formamide in 0. 5% SSC, 0.1% Tween	prepared fresh
Collagenase/ Trypsin	2% Collagenase in Trypsin solution	prepared fresh
Developing solution	50% BM Purple (Roche), 50% AP Buffer	prepared fresh,
DMSO	Dimethyl sulfoxide (Sigma-Aldrich): $(CH_3)_2SO$	RT
E3 medium (60x)	1.75% NaCl, 0.08% KCl, 0.3% CaCl2•2H2O, 0.5% MgCl2•6H2O, pH 7.2 + methylene blue	autoclave, 4°C
EDTA	0. 5M EDTA, pH8	RT
gel loading dye	2.5% Ficoll-400, 11mM EDTA, 3.3mM Tris-HCl, 0.017% SDS, 0.015% bromophenol blue, pH 8. 0	RT
HBSS++	1x HBSS (-Ca, -Mg, -phenol red), 0.01M Hepes pH8. 0.25% BSA	prepared fresh, filter steralize
HBSS++ with Hoechst^{DEAD}	0.1% Hoechst^{DEAD} (33258; Invitrogen) in HBSS++	RT
HotSHOT neutralising reagent	40 mM Tris-HCl, pH 5.0	RT
HotSHOT alkaline lysis reagent	25 mM NaOH, 0.2 mM disodium EDTA, pH 12.0	RT
Hybe$^{-/-}$	50% Formamide, 5x SSC, 9.2 mM Citric Acid, 0.1% Tween	4°C
Hybe$^{+/+}$	50% Formamide, 5x SSC, 9.2 mM Citric Acid, 0.1% Tween, 0.05% tRNA, 0.005% Heparin	4°C
LB (Agar)	1% Bactotryptone, 0.5% Bactoyeast, 1% NaCl, pH 7.0 (1. 5% Bactoagar)	4°C
LB + Ampicillin	50 µg/ml Ampicillin in LB	4°C
LB + Chloramphenicol	12.5 µg/ml Chlormaphenicol in LB	4°C
LB + Kanamycin	50 µg/ml Kanamycin in LB	4°C
MAB	0.1 M maleic Acid, 0.15 M NaCl, pH 7.5	4°C
MAB Block	0.1 M maleic Acid, 0.15 M NaCl, pH 7.5, 2% Blocking Reagent	-20°C
MABtw	0.1 M maleic Acid, 0.15 M NaCl, 0. 1% Tween, pH 7.5	4°C
Methylene blue	$C_{16}H_{18}N_3SCl$ in solution	RT
MS222	0.4% Ethyl-3-amino benzoate methanesulfonate salt, pH 7.0	4°C
PBS	0.8% NaCl, 0.15% Na2HPO4, 0.025% KH2PO4, 0.02% KCl, pH 7.2	RT
PBStw	0.8% NaCl, 0.15% Na2HPO4, 0.025% KH2PO4, 0.02% KCl, pH 7.2 + 0.1% tween	RT
PFA	4% PFA, 96% PBStw	RT

Proteinase K	20 mg/ml proteinase K (Sigma-Aldrich) in dH_2O	-20°C
Ringer's solution (-Ca^+)	116 mM NaCl, 2.9 mM KCl, 5 mM HEPES, add fresh 1 mM EDTA	Filter sterilize, RT
PTU (50x)	0.15% 1-phenyl-2-thiourea	4°C
SSC (20x)	saline-sodium citrate buffer (20x): 3 M NaCl, 0. 3 M $Na_3C_6H_5O_7$, pH 7. 0	autoclave, RT
TAE	40 mM Tris, 20 mM CH_3COOH, 1 mM EDTA.	RT
TE	10 mM Tris, 0.1 mM EDTA, pH 7.6	autoclave, RT
TRIS	tris(hydroxymethyl)aminomethane: 0.1 M Tris-HCl, pH 9. 5	autoclave, RT
Trypsin solution	0.05% Trypsin (with EDTA) in 1xHBSS (-Ca, -Mg, -phenol red)	-20°C

2.1.2 Morpholinos

All MOs used in this thesis were obtained from Gene Tools, LLC (Philomath, USA). MOs were dissolved in RNase free H_20 to make a 50 µM stock solution and stored at RT with parafilm sealed lids. Where not previously published, MOs were designed using Gene Tools' Oligo Design Service to bind over a splice junction to cause aberrant splicing. Just before use, MOs were heated to 65 °C for 5 minutes and kept at 4 °C. Morpholino information is given in the table below.

Table 2.2 List of morpholinos used in this thesis

MO name	Target region	MO sequence (5' – 3')	ng	Reference
gata2b-MO	intron 3 / exon 4	TTCACGTCCTATTGGCACACGATGC	25 ng	(Butko et al., 2015)
gfi1aa-MO	exon 3 / intron 3	TTTAAGCTAGAACAGACCCGGCGAG	10 ng	This thesis
gfi1ab-MO	exon 1 / intron 1	ATTTAATTTAGACATACCTGCACAT	14 ng	This thesis
gfi1ab-MO2	exon 2 / intron 2	ATATAGCTACCCGTACCCGGTGATG	15 ng	This thesis
runx1P2-MO	exon 1 / intron 1	AGCGCTCTTACCGTATTTGGCGTCC	4 ng	(Gering and Patient, 2005)
tal1b-MO	ATG	GCGGACTCAACTGCACCATTCGAGT	6 ng	(Zhen et al., 2013)

2.1.3 Chemical inhibitors

Chemical inhibitors used in this thesis, their targets, solubility, and concentration are detailed below.

Table 2-2: Chemical inhibitors.

Inhibitor	Name	Primary targets	Solubility	Conc.	Source
Tranylcypromine (TC)	Trans-2-phenylcyclopropanamine, SKF trans-355, 2-PCPA; $C_9H_{11}N$	Demethylase inhibitor; Lsd1	Water	0.12M	Enzo Life Science

2.1.4 Primers

All primers were designed using the Primer3 web service (Koressaar and Remm, 2007). Primers were ordered from Sigma-Aldrich (DST purified), re-suspended to a stock solution of 100 µM (stored at -20°C) and used at 10 pmol per reaction.

2.1.5 Primers for quantitative real-time PCR

Primer pairs for qPCR designed in this thesis amplify a product around 100 to 150 bp. If qPCR primers for a given gene have been previously published, these were used. Melting curves and efficiencies were tested for all qPCR primers used in this thesis.

Table 2-3: qPCR primer pairs

gene	forward primer sequence (5' → 3')	reverse primer sequence (5' → 3')	source
cmyb	AGAGGGTGAAAGAAATCGAG	ACTGAAACAACAATGCCAAC	(Pouget et al., 2014)
dll4	CTTCACCGGACCCCTCTGT	TGGAAGCGGTCTTGAGTTTCTC	(Geudens et al., 2010)
eef1a1l1	GAGAAGTTCGAGAAGGAAGC	CGTAGTATTTGCTGGTCTCG	(Bertrand et al., 2008)
efnb2a	CCCATTTCCCCCAAAGACTA	CTTCCCCATGAGGAGATGC	(Lee et al., 2014)
gfi1aa	AGCGGGTCCTACAAATGCAT	ACAGATTTCGCAGGCAAAGG	This thesis
tal1	TCGAATGGTGCAGTTGAGTC	CCATACTGTTCCGCATCTCC	This thesis
tbx20	CCAAGATTTCCTGCAGCCTC	ACTCCGGAAAATGACACACG	This thesis
ubc	AAGAGACTCCCATACACCGC	ATTCTCAATGGTGTCGCTGG	This thesis

2.1.6 Primers used for making gRNA, MO and CRISPR/Cas9 validation, genotyping and characterisation of BACs

Splice MO efficiency were tested by PCR across the splice-junction. The *runx1*^{W84X} mutant was genotyped by PCR following a restriction digest with *HaeII*. BACs were characterised for the presence of certain *runx1* exons.

Table 2-4: Primers for testing MO efficiency and genotyping.

Gene	gRNA name	sequence (5' → 3') if not already present the first 2 bp were converted into "g"	
irf1b	Guide 438	gGCTGACGGCGGGGAGCATG	
pik3cd	Guide 443	gGACGTTACCGTGGGACAAG	
Gene	**CRISPR/MO/mutant**	**Forward primer sequence (5' → 3')**	**Reverse primer sequence (5' → 3')**
irf1b	Guide 438	TCCTTCACTGCAACTAACACT	AAATGGGATGCTTGAGGTGC
pik3cd	Guide 443	CAGCCCAAGTTGACAACACA	CCCTCGCATGCAAAACAATA

gfi1aa	*gfi1aa*-MO	GCAATCAGTCCCCTAAAGCC	GTTGGATGAGGAGATGAGCG
gfi1ab	*gfi1ab*-MO / *gfi1ab*-MO2	GGAGGACACAATGCCTAGGT	GGTCACGCGCTCATTGTAAT
runx1	*runx1P2*-MO	CCGGAGATTCATTCACAGCG	GCCCACGAAACGCAGGTC
runx1	runx1^{W84X}	GCTCTGGTGGGCAAACTG	CATGTGTTTGGACTGTGGGG
Gene	**Exon**	**Forward primer sequence (5' → 3')**	**Reverse primer sequence (5' → 3')**
runx1	exon 1	CTGCGTTTTGGGGTGTTCAG	GGTACGACTGGAAGCTCTCG
runx1	exon 2a	CGGAGATTCATTCACAGCGC	GGTGTGAAGTGGGCTGAACT
runx1	exon 2b	CCTGAGCTCGGGGAAGAT	GGCAGGGTCTTGTTGCAG
runx1	exon 6	TCCCATCACTCCCATCTCTC	GACGGCCTCTACAGACATGC

2.1.7 Primers used for BAC cloning

Primers used for the BAC cloning were ordered from Sigma-Aldrich (RP1 purified).

Table 2-5: BAC recombineering primers

Gene	Forward primer sequence (5' → 3')	Reverse primer sequence (5' → 3')
Runx1P2-Citrine	AAGAGCTGTGGATCCCCATCCCGAAACTCACG GAGACAAAAACATACCAAATGGTGAGCAAGG GCGAGGAG	GTCCCGATCTGAGAAGCGCTCTTACCGTATTT GGCGTCCCAAAGAAAAACTCCAGAAGTAGTG AGGAGGCTTT
pindigobac_itol2	TTCTCTGTTTTTGTCCGTGGAATGAACAATGGA AGTCCGAGCTCATCGCTCCCTGCTCGAGCCGG GCCCAAGTG	CCCGCCAACACCCGCTGACGCGAACCCCTTGC GGCCGCATATTATGATCCTCTAGATCAGATCT

2.1.8 Primers for ISH

For all novel *in situ* hybridisation (ISH) probes created, primer pairs were designed against the Ensembl cDNA sequence using Primer3 (Koressaar and Remm, 2007). For genes with several protein coding splice isoforms, a probe was designed to bind to all isoforms, where possible. ISH probes were designed to be 1 kb or shorter in length, as dictated by gene size.

Table 2-6: Primer pairs for general ISH probes.

Gene	Forward primer sequence (5' → 3')	Reverse primer sequence (5' → 3')
agap2	TGCCGTCTTCTCCAATGACT	AGGGGTTACGAGACTCTCCT
angpt1	AACCAGAAGATCCAGCAGCT	TCGATGTGGAAGCTGTCGTA
bcl6a	AAAGTCCAGCGATTCCTCCA	GAGAGAAGAGTGGCTGTCGA
dnmt1	CATAAAGCCCGTTGAGCCTC	GAATCAAAGAGGGAGAGACGC
dnmt3aa	TTCCTGCCATACCTCAGCAA	ATGAGCACTTGTCTCCCTCC
dnmt3ab	CAGGTGTGTATGGGCAGAGA	GGACATCGAACACAAGCCTC
dnmt3ba	ATCGTGCCGGTAGATGATGT	TCCTTTTGACAAGCACGCAA
dnmt3bb.1	CATCTGTGGGCGATGGATTG	AACTCTGGAGGGCTGGATTC
dnmt3bb.2	AGGGAAGCAGCATATCCGAT	TTTGCCTGGACTCGATCTGT
dnmt3bb.3	GGGGAAGAGGAAGACGTTGA	GACTGCAAGGGTCTCAGAC

gata2a	GGCGTGAATGTGTGAACTGT	GAGTTTTGAGGTGGCCGTTT
gfi1ab	TATGAGGCAGAGGAGGACAC	CGCCATTTAACAGGAGACGAC
gfi1b	AGCACTTTTCATCAGCTCGA	TGGAAAACACCTTGTCGCAG
irf1b	TGGACTCATGTGGGTCAACA	GGTCTCCAGGTGTCCAGAAA
itga4	GACAACGTGGCTGCTCTTAC	TGCCTTTATCCTGCTGACCA
itm2cd	GCTTGATCATGGCCTCAGTG	CAAACCGAGGCTTGACACAA
map2k6	AGCCTACGTTACCATTGGAGA	ACAGGAACACTCTCAGTCCC
mknk2b	TCCCTTTCTTCTGCCCACTT	TCATTCCTCTAGCCACACCG
mpl	TGGCCAGAAGTCCAATGTCT	GCCCCAAGATCGGATTGTTC
pbrm1l	AAGCAGCTGACCTTCCTGAT	ATTCCTCCATTCTGTGCCCA
pik3cd	TGGTCGATTCAGGAGCCTTT	TAGAACCTGAACCAGCTGCA
pik3r1	AATGGAGTGACAGTCAGGCA	CCTCATCGTCCTCCACCATT
ppm1g	CTTGTCTCAACCCAACACGG	GTCTGAACCTGGCTCCTCTT
prkab1a	AGGGTTGTACAGAGATGCACT	TATCAGAGCGCGTGACAGAT
prkcbb	CTCGCACTTTGGCTGTCTTT	CATCAACGCGTCCCAACTAG
Runx1(short)	AGAAGCCGGATGAAGCAGTA	GGTTGGGTGAGATTGGTGTT
Runx1P1	ATCTCCCGTCAACAATGCAC	AGCAAGACTGGTACGACTGG
Runx1P2	CCGGAGATTCATTCACAGCG	TGAGTTTCGGGATGGGGATC
skap1	TTGCAGCTCGAGAGTCTAGG	TCGCAACAAAGCAAACCTCA
sox17	CAAAACCCAGACCTGCACAA	ATTAAGTCAGACCCCGGCAA
syk	ATGGCAGCTGGTAGAACACT	GTACATCTCAGGTGGGCAGT
tfr1b	TCCAGGCCAATGAAGTGAGT	TTGGTCATGAACGAGACGGA
uhrf1	ATAAAGCAGGACCCCGACAA	CACCTTCATCTTCTTCGGCG
wdhd1	CTGGTGAAGATGGTGGAGGT	GACGATGCCCACAGAATTCC
xbp1	TGGTCGTAGTTACAGCAGGG	AACCGTTTGATCCTCCACCT

2.1.9 TaqMan probes

TaqMan probes were ordered from Thermofisher.

Table 2-7: TaqMan probes.

Gene	Probe	Gene	Probe
abcg2a	Dr03142535_m1	jag1a	Dr03093484_m1
angpt1	Dr03125214_m1	lsm12b	Dr03139893_m1
cbfb	Dr03093780_m1	map2k6	Dr03138399_m1
cmyb	Dr03432754_m1	mdm2	Dr03074092_m1
csf1ra	Dr03125169_m1	mpl	Dr03102183_m1
dll4	Dr03428642_m	mrpl28	Dr03424514_m1
dnmt1	Dr03131383_m1	mtf2	Dr03072836_m1
dnmt3aa	Dr03096717_m1	pbx2	Dr03093359_m1
dnmt3ab	Dr03141501_m1	pik3cd	Dr03088274_m1
dnmt3ba	Dr03083770_m1	ppm1g	Dr03113930_m1
dnmt3bb.1	Dr03090840_m1	prkcbb	Dr03081363_m1
dnmt3bb.2	Dr03105760_m1	pygl	Dr03128068_m1
eef1a1l	Dr03432748_m1	rpl13a	Dr03101115_g1
efnb2a	Dr03073975_m1	rpl22l1	Dr03427679_m1

ephb4b	Dr03125078_m1	runx1	Dr03074179_m1
foxo3b	Dr03112018_m1	sox17	Dr03105677_m1
gata2a	Dr03086716_m1	sox32	Dr03105985_m1
gata2b	Dr03140570_m1	sox7	Dr03092284_m1
gfi1aa	Dr03090706_m1	syk	Dr03088634_m1
gfi1ab	Dr03438442_m1	tal1	Dr03133428_m1
ifng1-2	Dr03081923_m1	tbx20	Dr03086930_m1
ikaros	Dr03105499_m1	tp53	Dr03112085_m1
irf1b	Dr03151900_g1	uhrf1	Dr03133361_m1
itga2b	Dr03089673_m1	wdhd1	Dr03127427_m1

2.1.10 Plasmids with antisense probe constructs for ISH

For all novel *in situ* hybridisation (ISH) probes created, PCR products were sub-cloned into the pGEM-Teasy vector.

Table 2-8: ISH probe plasmids.

Gene / probe	Plasmid	Restriction enzyme	Polymerase	Source
agap2	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
angpt1	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
bcl6a	pGEM-Teasy	SacII	Sp6	This thesis
cmyb	pBK-CMV	EcoRI	T7	(Thompson et al., 1998)
dnmt1	pGEM-Teasy	SacII	Sp6	This thesis
dnmt3aa	pGEM-Teasy	SacII or NcoI	Sp6	This thesis
dnmt3ab	pGEM-Teasy	SpeI or NdeI	T7	This thesis
dnmt3ba	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
dnmt3bb.1	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
dnmt3bb.2	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
dnmt3bb.3	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
gata2a	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
gfi1aa	pCS2	ClaI	T7	(Cooney et al., 2013)
gfi1ab	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
gfi1b	pGEM-Teasy	SpeI	T7	This thesis
irf1b	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
itga4	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
itm2cd	pGEM-Teasy	SpeI or NdeI	T7	This thesis
map2k6	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
mknk2b	pGEM-Teasy	SpeI or NdeI	T7	This thesis
mpl	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
pbrm1l	pGEM-Teasy	SacII	Sp6	This thesis
pik3cd	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
pik3r1	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
ppm1g	pGEM-Teasy	SacII	Sp6	This thesis
prkab1a	pGEM-Teasy	SpeI or NdeI	T7	This thesis
prkcbb	pGEM-Teasy	SpeI or NdeI	T7	This thesis

runx1	pBluescript SK(+)	HindIII	T7	(Kalev-Zylinska et al., 2002)
runx1(short)	pGEM-Teasy	NcoI	Sp6	This thesis
runx1P1	pGEM-Teasy	SacII	Sp6	This thesis
runx1P2	pGEM-Teasy	SacII	Sp6	This thesis
skap1	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
sox17	pGEM-Teasy	SpeI	T7	This thesis
syk	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
tfr1b	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
uhrf1	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
wdhd1	pGEM-Teasy	NcoI or SacII	Sp6	This thesis
xbp1	pGEM-Teasy	SpeI or NdeI	T7	This thesis

2.1.11 BACs and Plasmids for BAC recombineering

BACs that included the *runx1* locus were previously identified by Southern blotting to the BAC Zebrafish High-Density Filter Library (Incyte Genomics, Wilmington, DE) (Lam et al., 2009). These BACs were kindly given to us by the Crosier lab since the Filter Library has been discontinued.

Table 2-9: BAC cloning plasmids.

BAC	**Resistence**	**Source**
97a02	Chlor	(Lam et al., 2009)
14i20	Chlor	(Lam et al., 2009)
47j23	Chlor	(Lam et al., 2009)
135g16	Chlor	(Lam et al., 2009)
Plasmid		
BirA-2A-Citrine-SV40pAFKF	Amp and Kan	Tatjana Sauka-Spengler lab
ITol2	Amp and Spec	(Suster et al., 2011)
Recombineered BAC		
97a02-P2Citrine-tol2(BAC)	Amp and Chlor	This thesis
14i20-P2Citrine-tol2(BAC)	Amp and Chlor	This thesis

2.2 Zebrafish husbandry and manipulation of embryos

2.2.1 Zebrafish lines

Zebrafish lines employed in this thesis were Wt^{KCL} (originated from Nigel Holder's lab), *Tg(kdrl:Hsa.HRAS-mCherry)*, here called *Tg(kdrl:mCherry)* (Chi et al., 2008) and $runx1^{W84X}$ (Jin et al., 2012).

2.2.2 Zebrafish maintenance

Danio rerio (zebrafish) were maintained in flowing system water at 28.5°C, 450-550 µS conductivity range, a pH of 7.1-7.2 and a light/dark cycle of 14 hours of light and 10 hours of darkness. Fish were fed on a mixed diet of high protein pellets, flake food and enriched brine shrimp three times a day (BMS Staff, University of Oxford). All zebrafish work was approved by the Research Ethics Committee of the University of Oxford.

2.2.3 Breeding, embryo collection and maintenance

Marbles boxes were placed in tanks about 15 hours before the lights came on. Alternatively, breeding pairs were set up as and when necessary (eg. crossing different lines). Eggs were collected from natural spawning and washed in system water. Embryos were grown between 22 °C and 33 °C in system water or E3 medium. Methylene blue was added to inhibit fungal growth. Embryos were staged according to morphology under a dissection microscope (Kimmel et al., 1995).

2.2.4 Micro-injections of zebrafish embryos

Embryos were aligned against a glass slide in a petri dish and injected with 1 nl of either DNA (180 ng/µl) plus *tol2* mRNA (120 ng/µl), gRNA (200 ng/µl) plus Cas9 protein (500 ng/µl), or morpholino (1-20 ng) as required. DNA plus *tol2* mRNA and gRNA plus Cas9 protein were injected into the cell of early 1-cell stage embryos. Morpholinos were injected into the yolk of 1- to 2- cell stage embryos.

2.2.5 Pharmacological treatment of embryos

Tranylcypromine Table 2-2) was diluted in E3 media and added to embryos at the desired stage until collection. Tranylcypromine was dissolved in H2O to a stock concentration of 0.12M. Control embryos were exposed to normal E3 medium.

2.2.6 Dechorionation and fixation of zebrafish embryos

Embryos fixed or imaged till 3 dpf were manually dechorionated with INOX number 5 forceps. Embryos fixed at 4 and 5 dpf were allowed to hatch naturally. Embryos were fixed in 4% paraformaldehyde in PBStw for at least 3 hours at room temperature or at 4 °C o/n, then moved through 25%, 50% and 75% ethanol washes to 100% ethanol and stored at -20 °C.

2.2.7 Fin-clipping of zebrafish embryos

Three days old embryos were anaesthetised in 10% 5 g/l MS222 (pH 7) in E3 medium for the minimum period required, transferred to a petri-dish, and a small section of the caudal fin was removed with a sterile scalpel. Fish were placed into a 24 well plate with fresh E3 medium. The dissected tissue was collected in a PCR strip for extraction of genomic DNA via the HotSHOT method (Truett et al., 2000). Fish were kept in isolation after the procedure until results were obtained. All fin-clipping was carried out in accordance to Animals (Scientific Procedures) Act 1986 (ASPA) regulations and fish were monitored for adverse effects during and after the procedure.

2.2.8 Fluorescence-activated cell sorting (FACS)

Embryos were staged accordingly, pre-sorted for the presence of fluorescent reporters using the Olympus stereo microscope MVX10 and collected in low binding microcentrifuge tubes (SafeSeal Microcentriguge Tubes, Sorenson; Cat#39640T). Using 1 ml of deyolking buffer (Ringer -Ca/-Mg: 116 mMNaCl, 2.9 mM KCl, 5 mM HEPES with freshly added 1mM EDTA) the embryos were rinsed through a p200 tip several times and briefly spun down in a table-top centrifuge (500g) for 2 min. The

supernatant was removed and 300 µl per 100 embryo of a collagenase/trypsin solution (20 mg collagenase in 0.05% Trypsin with EDTA in 1x HBSS solution) were added. For the cell dissociation, the embryos were incubated at 30°C on a rocker for 15 min. Every 3 min, the embryos were pipetted through a p200 tip coated with $HBSS^{+/+}$ solution against the wall of the tube for better tissue disruption. To stop the reaction, the cell solution was transferred into 5 ml of $HBSS^{+/+}$ (1x HBSS, 10 mM Hepes, 0.25% BSA) solution provided in a 50 µl falcon tube. Glass serological pipettes coated with $HBSS^{+/+}$ were used to further dissociate the cells. Cells were spun down at 500g for 10 min at RT. The supernatant was replaced with 5 ml of fresh $HBSS^{+/+}$ medium and the pellet was re-suspended by pipetting it up and down using a glass serological pipettes coated with $HBSS^{+/+}$. Dissociated cells were passed through a 40 µm cell strainer placed into a fresh 50 ml falcon tube by centrifugation (750g for 10 min at RT). The supernatant was removed. Cells were re-suspended in appropriated volume (~10 µl per embryo; ~3-7x10^6 cells per ml) of $HBSS^{+/+}$ with $Hoechst^{DEAD}$ (33258; Invitrogen) (1:4000 dilution) and transferred to a 5 ml polystyrene round bottom tube for FAC-sorting. The FAC-sorting was carried out by the WIMM Flow Cytometry Facility.

2.3 Extraction, preparation and manipulation of DNA

2.3.1 Bacterial culture

Bacteria were cultured at 37 °C for 9–18 hr as required either in shaking liquid culture (LB) or solid culture (streaking on LB agar). As required, cultures were supplemented with 50 µg/ml of ampicillin or kanamycin. Cultures were stored at 4 °C.

2.3.2 Transformation of chemically competent *Escherichia coli*

Approximately 20 µl of DH5α competent cells (NEB) were thawed per transformation, and incubated with DNA at 4 °C for 30 min. Cells were heat shocked for 30 s at 42 °C and allowed to recover for 30 min at 37 °C in 200 µl SOC, before plating onto selective LB agar.

2.3.3 Isolation and extraction of plasmid DNA

Bacterial cells from liquid culture were pelleted and plasmid DNA isolated by the principle of alkaline lysis on an immobilised column with either the Quicklyse mini- or HiSpeed midi-kit (QIAGEN) according to the manufacturer's instructions. DNA was eluted in QLE elution buffer and stored at -20 °C. Plasmids were sequenced by semi-automated Sanger sequencing (WIMM sequencing service) with the appropriate primers.

2.3.4 Genomic DNA extraction from zebrafish embryos

Genomic DNA was extracted from a batch of 10-50 embryos using the DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's instructions.

2.3.5 Genomic DNA extraction from fin-clips

Genomic DNA was extracted using the HotSHOT method (Truett et al., 2000). In short, collected tissue samples were immersed in 75 µl alkaline lysis reagent and incubated at 95°C for 1 hr. Samples were cooled to 4°C before adding 75 µl neutralising reagent. Samples were well mixed and stored over night at 4°C.

2.3.6 Determining nucleic acid concentration

Nucleic acid concentration was determined by its A260 with an NanoDrop (ND)-1000 (Thermofischer). An extinction coefficient of 50 was used for double-stranded DNA, 40 for RNA. Purity was determined by the profile of the sample on an agarose gel and/or the A230:260:280, with an idealised ratio of 1:2:1 for DNA and 1:1.8:1 for RNA.

2.3.7 Polymerase Chain Reaction (PCR)

PCR was conducted using JumpStart™ REDTaq® ReadyMix™ (Sigma-Aldrich) reaction mix with 10 µM of each primer per reaction, in a total volume of 20 µl. The standard programme used has an

initial denaturation step at 95°C for 4 min. The following PCR cycles have a denaturation step at 95°C for 15 sec, followed by an annealing step at 60°C for 45 sec. Extension at 68°C was 1 min per kb, depending on the length of the PCR product. The number of PCR cycles depended on the concentration of the PCR product and the length, followed by a 7 min final extension step.

2.3.8 Purification of PCR amplified DNA

As required, DNA was purified by selective binding to a silica membrane using the PCR purification kit (QIAGEN).

2.3.9 Restriction enzyme digest

DNA was digested at 37 °C for 1-12 h with restriction enzymes (NEB) in the provided buffers, and according to the manufacturer's instructions. The cut and uncut fragments were analysed on an agarose gel. Typically, 2 units of enzyme was used per 1 μg of DNA. DNA fragments were ligated as required with 3:1 molar ratio of fragment to vector using T4 DNA ligase (Promega). Reactions were typically left for 6 hours at room temperature before being used to transform bacteria.

2.3.10 Agarose gel electrophoresis

Nucleic acid or DNA ladders (NEB, 1 kB or 100 bp) were mixed 1:5 with loading dye, to be run on a 1-3% agarose gel made up with 0.5% TAE and 0.01% GelRed™ (VWR). Typically, the gel was run at 135V in 0.5% TAE for the required amount of time and visualised with trans-UV light, then photographed at 300 dpi.

2.3.11 Purification of gel extracted DNA

As required, DNA bands were cut out of an agarose gel under UV light and purified using the QIAEX II Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions.

2.3.12 Cloning of PCR products

PCR products were cloned into expression vectors using the pGEM-T easy Vector System (Promega), according to the manufacturer's suggestions. In short, up to 5.5 µl PCR product was incubated with 1 µl T4 ligase and 1 µl pGEM-T easy vector in 2x ligation Buffer in a total volume of 15 µl for 1 hr at RT or o/N at 4°C.

2.3.13 BAC cloning

The here performed BAC cloning followed the Suster et al. protocol (Suster et al., 2011). Alterations to the published protocol were that the Citrine-FKF (containing *kanamycin* flanked by two FRT sites) transgene cassette was introduced first and the iTol2-*amp* cassette second. BAC recombineering was performed in SW105 cells (Warming et al., 2005), that can be used for *galK* positive/negative selection and that contain an ara-inducible *Flpe* gene. The *BirA-2A-Citrine-SV40pAFKF* vector (Table 2-9), that was used as a PCR template for the transgene cassette, was a gift from the Tatjana Sauka-Spengler lab. The used primers containing 50 bp homology flanks can be found in Table 2-5. For PCR, the Phusion High-Fidelity PCR Master Mix (NEB) was used.

PCR of the recombineering cassettes was performed the following:

98°C	5 min	x1
98°C	30 sec	x5
56°C	30 sec	
72°C	2 min 15 sec	
98°C	30 sec	x23
65°C	30 sec	
72°C	2 min 15 sec	
72°C	7 min	x1
10°C	10 min	x1

All bacterial incubation steps were performed at 32°C. BACs were purified using the Invitrogen PureLink HiPure Plasmid Midiprep Kit (cat no K2100-04).

2.3.14 Sanger DNA sequencing

All sequencing of plasmids, BACs or PCR products was carried out using the Sanger DNA Sequencing facility at the WIMM.

2.4 Extraction, preparation and manipulation of RNA

2.4.1 RNA isolation from tissue samples

Total RNA was isolated from staged zebrafish using 500 µl TRIzol (Invitrogen). Embryos were homogenised by aspiration through a 23 G syringe. A further 500 µl TRIzol was added and the sample vortexed. Nucleic acid was isolated by the addition of 200 µl chloroform, mixing and centrifugation at the highest speed for 15 min. The top aqueous phase was transferred into a fresh tube and nucleic acids were precipitated with 500 µl isopropanol, pelleted and re-suspended. DNA was digested with 1 µl DNase I (Ambion). For RT-PCR, resultant RNA was precipitated with 3.75 x volume of ethanol, pelleted and re-suspended in 20 µl water. For qRT-PCR, RNA was isolated by selective binding to a silica membrane (RNeasy micro kit, QIAGEN) according to the instructions provided. RNA quality was always assessed on an agarose gel and by UV spectroscopy. Isolated RNA was stored at -20 °C.

2.4.2 RNA Sample Preparation from FAC-sorted cells for RNA sequencing and quality/quantity determination

For the isolation of RNA from FAC-sorted cells, cells were directly sorted into 350 µl RLT buffer (QIAGEN) containing β-mercaptoethanol (BME) (20 µl BME per 1 ml RLT buffer). To adjust for different volumes that come with various cell numbers, RNA free H_2O was added accordingly to the table below.

1 cell	~0.00325 ul	add 100 ul H_2O
3,000 cells	~9.6 ul	add 90.4 ul H_2O
5,000 cells	~16 ul	add 84 ul H_2O
10,000 cells	~32 ul	add 68 ul H_2O
30,000 cells	~96 ul	add 4 ul H_2O

The cell lysate was vortexed 3x for 20 sec. and run through a QIAshredder (QIAGEN) column. RNA was cleaned up using the RNeasy Plus Micro Kit (QIAGEN) according to the manufacturer's protocol. RNA was eluted twice with 10 µl of RNA free H_2O. Quality and quantity was determined using the 2100 Bioanalyzer Pico kit (Agilent technologies).

2.4.3 Sample preparation from FAC-sorted cells for Fluidigm BioMark gene expression analysis

For gene expression analysis of small cell numbers the Fluidigm BioMark system was used. All pipetting steps were performed in the clean-room of the WIMM single-cell facility. A TaqMan Assay master mix (0.2x) was prepared by combining 1 µl of each TaqMan probes (ThermoFisher) (Table 2-7) and topping it up to 100 µl with TE buffer. This 0.2x Assay mix was used for a specific target pre-amplification. cDNA generation was based on the Superscript III 1 Step Kit (ThermoFisher) supplemented with SUPERase-In RNase Inhibitor. On the day of the FAC-sort a cDNA master mix was freshly prepared (10 µl for each sample):

Reagent	**µl**
2x Reaction Buffer	5
Ambion SUPERase	0.1
Invitrigen TE buffer	1.2
0.2x Assay mix (Pool of TaqMAN for cDNA)	2.5
RT/Taq mix	1.2
total	10

The mix was strictly kept on ice. FAC-sorted cells (100 cells) were directly sorted into the cDNA master mix in PCR tubes. Cells were centrifuged at 300g for 1 min. RT-PCR and pre-amplification was performed in a normal PCR machine.

RT	50°C	15 min	x1
Inactivation of RTase/activation of Taq enzyme	95°C	2 min	x1
Specific target pre-amplification	95°C	15 sec	x20
	60°C	4 min	

After pre-amplification 40 µl TE buffer was added to the reaction mix. Gene expression analysis was performed using a 48.48 IFC chip (Fluidigm) according to the manufacturer's instructions.

2.4.4 Reverse Transcription of RNA

RNA was reverse transcribed to cDNA using the SuperScript III or SuperScript IV enzyme (Invitrogen), according to the manufacturer's instructions, in a total volume of 20 µl. Up to 1 µg of RNA, random hexamer primers (250 ng) and 10 mM dNTPs were mixed, adjusted to 13 µl and heated to 65 °C for 5 min. The reaction mix was then placed on ice for 1 min. 4 µl 5x First Strand Buffer (Thermo Scientific), 1 µl 0.1M DTT (Thermo Scientific) and 1 µl RNaseOUT (Thermo Scientific) 1 µl Superscript enzyme (Thermo Scientific) were added, mixed and let stand at RT for 5 min before incubation at 50°C for a minimum of 60 min. Enzyme activity was then heat-inactivated at 70°C for 15 min. The cDNA was usually diluted 1 to 10.

2.4.5 *In vitro* synthesis of capped mRNA

Capped mRNA in appropriate expression vectors were in vitro transcribed from 1 µg purified linearised DNA plasmid template using the SP6, T3 or T7 mMessage mMachine kit (Ambion) according to the manufacturer's instructions. Reactions containing 0.5 mM Cap analogue were incubated at 37 °C for 1 – 3 hours. DNA was digested with the Turbo Dnase provided (Ambion) and RNA cleaned and purified on RNeasy micro columns (QIAGEN). Purity and concentration of the mRNA was determined by agarose gel electrophoresis and UV spectroscopy. Synthesised mRNA was aliquoted and stored at -80°C.

2.4.6 *In vitro* synthesis of antisense RNA-probes

For RNA-probe synthesis 1 µg linearized plasmid was incubated for 2-18 hrs at 37°C with 10x concentrated NTP Labeling Mixture (DIG), 10x concentrated Transcription Buffer, 10U RNase inhibitor (RNAseOut, Invitrogen) and 20U SP6, T7 or T3 RNA Polymerase (Roche). Template DNA was digested with DNaseI (Roche) for 15 min at 37°C. The reaction was stopped with 1 µl EDTA,

the volume was then made up to 100 µl with dH_2O. Probes were precipitated with 1/3 volume of 10M NH_4Acetate and 2.5 volumes of 100% EtOH on ice for a minimum of 10 min. Probes were then pelleted by centrifuging for 10 min at 4°C, washed with 70% EtOH and air-dried. Probes were re-suspended in 100 µl $Hybe^{+/+}$. Typically, probes were diluted 1:200 in $Hybe^{+/+}$ for use in WISH.

2.4.7 In vitro synthesis of guide RNA (gRNA)

The online tool CCtop (Stemmer et al., 2015) was used for the design of gRNAs, followed by a cloning free synthesis protocol (Varshney et al., 2015). The used primer sequences are shown below.

Forward primer:

5'-GAAATTAATACGACTCACTATAGGX_{18}XXXXXXXXXXXXXXXXX$_1$GTTTTAGAGCTAGAAATAGC-3'

Reverse primer:

5'
AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3'

An 18 bp guide sequence (without PAM sequence) replaced the 18x X in the forward primer sequence. PCR was performed using the 2x Phusion MasterMix (NEB) with 10 µM of each primer.

98°C	30 sec	x1
98°C	10 sec	X35
60°C	30 sec	
72°C	15 sec	
72°C	7 min	x1
10°C	10 min	x1

Success of the PCR was validated on an agarose gel using 2 µl of the PCR product. The PCR amplicon was cleaned up with the PCR Purification Kit (e.g. QIAGEN) and eluted in 40µl EB.

RNA was *in vitro* transcribed using the NEB High Scribe T7 Quick High Yield Kit. A 15 µl reaction mix containing 250 ng of template DNA, 5 µl of NTP buffer and 1 µl T7 RNA Polymerase was incubated at 37°C for 4 hours. Template DAN was removed by adding 2 µl of DNAseI for another 15 min. RNA was purified with the MEGA Clear kit (Ambion) and eluted in 50 µl. RNA was quality checked on an agarose gel and quantified on the Nanodrop.

2.4.8 Whole mount *in situ* hybridisation (ISH)

Fixed embryos stored in 100 % EtOH were re-hydrated through a series of 5 min washes of 75%, 50% and 25% EtOH, followed by 3x 5 min washes in PBStw. Embryos older than 24 hpf were treated with 10 µg/ml proteinase K (Sigma) for 3 – 20 minutes as required. Embryos older than 4 dpf were treated with 20 µg/ml proteinase K for 20-25 minutes. Proteinase K activity was stopped by 2x 5 min washes in 0.1 M glycine, re-fixed with 4 % PFA for 20 min and washed 5 x with PBSTw. Embryos were transferred to 50 % Hybe$^{+/+}$ for 5 min and then pre-hybridised in 100 % Hybe$^{+/+}$ for at least 2 hours at 65 °C before hybridisation with probes in Hybe$^{+/+}$ (usually at a dilution of 1:200) at 65°C o/N.

Probes were replaced with pre-heated Hybe$^{-/-}$ and re-used multiple times. The following steps of the protocol were performed in a BioLaneTM HTI-16V (intavis AG). Embryos were washed through a series of 10 min washes in 75%, 50% and 25% Hybe-/- in 2x SSC, followed by a wash in 2x SSC, all at 65°C. 4 washes in pre-heated 0.2x SSC were followed by a series of 25%, 50%, 75% and 100% MABtw in 0.2x SSC before the blocking step in MAB-Block for 1 hr. Anti-Digoxigenin (Roche) antibody was diluted in MAB-Block and incubated o/N at 4°C. Embryos were then washed 8x at RT in MABtw.

Embryos were washed 3x for 5 min in AP buffer before addition of developing solution. Staining was developed in the dark at RT or 4°C, as appropriate. To stop the staining process, embryos were washed in PBStw and fixed in PFA for at least 20 min. For storage, embryos were washed 3x in PBS for 5 min, transferred to 80% glycerol and stored in an air- and light-tight environment at RT. Pictures were taken on a dissecting microscope in 100% glycerol.

2.4.9 Double fluorescent ISH

Double fluorescent in situ hybridization was performed as previously described (Brend and Holley, 2009). Briefly, 12-somite stage embryos from were incubated at 65°C for 18 hours in 2 ng/µl DIG- and fluorescein-labeled mRNA probes. After washes, the embryos were sequentially incubated with horseradish peroxidase-conjugated anti-FLUO (1:500; Roche, Basel, Switzerland) and anti-DIG

(1:500; Roche) antibodies. The RNA probes were sequentially revealed with custom-made FLUO (FITC; 1:200) and tetramethylrhodamine (TAMRA; 1:100) (Zhou and Vize, 2004) conjugated tyramide in PBST containing 0.003% (v/v) H_2O_2 for 35 min at RT.

Fluorescent images were taken on a Zeiss LSM880 confocal microscope using 10× air or 63× oil immersion objectives. Images were processed using the ZEN software (Zeiss).

2.4.10 fluorescent ISH with immunohistochemistry staining

First fluorescent ISH was performed which followed the general whole mount *in situ* hybridisation protocol. The signal was developed with SIGMAFAST Fast red/TR Naphthol (Sigma). Embryos were rinsed in phosphate-buffered saline with tween20 (PBT) and directly processed for immunohistochemistry.

After ISH, embryos were blocked in blocking buffer (5 % goat serum, 0.3 % Triton X-100 in PBT) for 1 hour at RT. They were incubated with the primary antibody against GFP (rabbit, 1:500, Molecular Probes), diluted in blocking buffer overnight at 4 °C. Secondary antibody raised in goat coupled to AlexaFluor488 (Invitrogen) was used in 1:500 dilutions for 3 h at RT. Hoechst 33342 was used as a nuclear counterstain.

Fluorescent images were taken on a Zeiss LSM880 confocal microscope using 10× air or 63× oil immersion objectives. Images were processed using the ZEN software (Zeiss).

2.4.11 Quantitative Polymerase Chain Reaction (qPCR)

Quantitative real-time PCR (q-RT-PCR) was carried out using SYBR green on an Applied Biosystems 7500 real-time cycler, according to the manufacturer's instructions. If possible reactions were carried out in triplicate, otherwise in duplicates. Reactions were adjusted to 20 µl total volume and cycled 40 times before melting curve analysis. Only primers and reactions with a single-peaked melting curve and a similar efficiency around 100 % were used in this thesis. Thresholds were automatically determined by the 7500 fast software (Applied Biosystems). Data were analysed using the ΔΔCT method as described previously (Dussault and Pouliot, 2006). The arithmetic mean of Ct

values from replicates was determined, resulting in a mean Ct value for the gene of interest (GOI). A combined value for the two housekeeping genes (HkG) was determined from the geometric mean of the two arithmetic means of the individual genes. The genes *eef1a1l1* and *ubc* were selected as housekeepers due to their stable expression during the developmental time points analysed. The ΔΔCt value between the control and test sample was calculated using equation (1) and the fold increase approximated by equation (2).

$$(1)\ \Delta\Delta Ct = [(Ct_{GOI\text{-}control} - Ct_{HkG\text{-}control}) - (Ct_{GOI\text{-}test} - Ct_{HkG\text{-}test})]$$

$$(2)\ \text{Fold increase} = 2\exp(\Delta\Delta Ct)$$

Results from biological replicates were combined and error bars are representative of the standard error about the mean (SEM). Significance was assessed using a two-tailed Student's t-test or an ANOVA analysis.

2.5 Microscopy

2.5.1 Bright-field microscopy

Bright-field microscopy was performed for embryos that underwent ISH. The number of embryos showing a particular phenotype was scored and images captured with a Nikon DS-Ri2 camera attached to a Nikon SMZ1000 dissecting microscope and the NIS-Elements BR 4.50.00 software. Exposure and white balance was set automatically and embryos were illuminated from the top and bottom on a glass depression slide, or from the top on a white depression tile as appropriate. Figures were prepared with Adobe Photoshop CS5.

2.5.2 Wide-field fluorescence imaging

Live embryos were anaesthetised with 160 μg/ml MS222 and mounted in 3% methylcellulose. Embryos were assessed and imaged on a SteREO Lumar V12 microscope (Zeiss). Exposure was

manually fixed for each experiment. Image acquisition and analysis was conducted with the Axio Vision software (Zeiss).

2.5.3 (Live) confocal imaging

Embryos were grown in E3 medium supplemented with PTU to prevent pigment cell formation. For imaging, embryos were anaesthetised with 160 µg/ml MS222 and orientated at the bottom of 35 mm glass bottomed dishes (Iwaki) in 1 % low melt agarose (Renaud et al., 2011). Embryos were covered with E3 media containing MS222 and imaged on an LSM 780 or 880 confocal (Zeiss) at 28.5 °C. Maximal projections were generated with the LSM image browser software (Zeiss).

2.5.4 Image analysis and representation

Acquired images were adjusted and assembled for figures with Adobe Photoshop CS5. Images were adjusted for levels. Adjustments were always conducted on control images; images of test embryos were adjusted identically to control embryos. All images are represented from a lateral view, with anterior to the left and dorsal to the top, unless otherwise stated.

2.6 Bioinformatic tools

2.6.1 Genomic, transcript, and protein sequence information

All sequence information (genomic, transcript, and protein) for zebrafish, mouse and human referred to in this thesis have been obtained from the Ensembl database (Flicek et al., 2012). Reference genomes are Zv.10 (*D. rerio*), GRCh38.p10 (*H. sapiens*) and GRCm38.p5 (*M. musculus*).

2.6.2 Analysis of homology

All multiple sequence alignments were carried out using Clustal Omega (Sievers and Higgins, 2014).

2.6.3 Intersection of gene lists.

Gene lists were converted into respective species gene names using the Ensembl BioMart tool. The gene list intersection was performed with Venny2.1 (Oliveros, 2007). Calculation of the probability of gene list overlap between two lists was performed using hypergeometric probability statistics. The number of background genes of the respective reference genome was adjusted: human genome assembly GRCh38.p10 (20310 coding genes); mouse genome assembly GRCm38.p5 (22597 coding genes).

2.6.4 RNAseq analysis

Sequenced reads were checked for base qualities, trimmed where 20% of the bases were below quality score 20, and filtered to exclude adapters using *Trimmomatic* (Version 0.32). Sequences were aligned to the *Zebrafish Genome Zv10* with *STAR* with default parameters. Aligned read features were counted using Subread tool: featureCounts method (version 1.4.5-p1). Differential gene expression (DEG) analysis was carried out using EdgR (Bioconductor Package).

For ANOVA-like test analysis, the generalized linear model (GLM) analysis was used as followed: Estimates “Dispersion”, fitted with “negative binomial model” and estimates “Generalized linear model likelihood ratio”. Genes with a False Discovery Rate (FDR) less than 0.05 were filtered out. Further, genes were filtered for log2 Fold Change <-0.7 and >0.7. Reads per kilobase per million (RPKM) values were used as input for Clustering analysis. Prior a z-score transformation of the log2 replicates mean for each sample was performed. Clustering analysis was performed using the Consensus Clustering Plus algorithm (Wilkerson and Hayes, 2010) with maximum cluster number to evaluate equal to 15. The algorithm generates a CDF plot to find the k at which the distribution reaches an approximate maximum stability. The number of 8 cluster was used as best stability of the clustering analysis. Using the genes of each cluster, heat-maps were generated using the same z-score transformation used as clustering analysis input.

To identify genes that were differentially expressed between two populations, we used GLM analysis as follows: Estimates “Dispersion”, fitted with “negative binomial model” and estimates “Generalized linear model likelihood ratio”. For the ++LOW vs. ++HIGH DEG analysis all 4 endothelial populations (mCherry^{+}, ++LOW, ++MEDIUM and ++HIGH) were used for normalisations. For the control vs. MO DEG analysis (for both NEGATIVE and ++HIGH) all 4 populations (NEGATIVE-Wt, NEGATIVE-MO, ++HIGH-Wt and ++HIGH-MO) were used for normalisations. All Genes with less than 3 samples >0 were removed before performing DEG analysis. All genes with FDR greater than 0.05 and all genes with at least 3 replicates =0 were removed from further analysis.

2.6.5 GO term and KEGG pathway analysis.

Enrichment of GO terms and KEGG pathways were calculated with DAVID6.8 (Huang et al., 2008b). Representative GO terms were taken from the BP_DIRECT class.

- Chapter 3 -

Generation of a novel *runx1* reporter line that allows for the reliable isolation of HE cells

3.1 Introduction

All previous screens for RUNX1 targets during haematopoiesis were performed on immortalized progenitor cell lines, or on cells derived from either *in vitro* differentiation cultures or the YS of mouse embryos (Lancrin et al., 2012; Lie-A-Ling et al., 2014; Michaud et al., 2008; Tanaka et al., 2012; Wilson et al., 2010a). An *in vivo* screen for RUNX1 targets focusing on the haemogenic endothelium (HE) in the embryonic dorsal aorta (DA) has not been conducted yet. For that purpose, the zebrafish model system provides several advantages: the cellular events during *de novo* generation of haematopoietic stem and progenitor cells (HSPC) in the embryonic DA have been described in detail (Bertrand et al., 2010a; Kissa and Herbomel, 2010; Murayama et al., 2006), zebrafish embryos are easily accessible (external development), they can be precisely staged and they can be genetically modified (Suster et al., 2009).

Several zebrafish transgenic reporter lines for *runx1* and other haematopoietic genes have been generated in different laboratories. However, none of them are an ideal tool to perform a screen for Runx1 targets. Haemogenic expression of *runx1* is found in the ventral aspects of the DA in both zebrafish and mouse (Gering and Patient, 2005; North et al., 1999). A *Tg(runx1P2:EGFP)* line using a short *runx1* promoter sequence shows ectopic expression in other vasculature tissues (Lam et al., 2009; Lam et al., 2010). Transgene expression of zebrafish reporter lines using the mouse *Runx1* +23 enhancer is dependent on Runx1 itself, even though it is not in mouse, making it impractical for functional experiments on *runx1* (He et al., 2015; Tamplin et al., 2015; Zhang et al., 2015). Additional reporter lines based on regulatory regions of other haematopoietic genes including *tal1b*, *cmyb* and *itga2b* (*CD41)* either show too broad expression, are dependent on Runx1 expression themselves, or their expression is only initiated in cells undergoing endothelial to haematopoietic transition (EHT), but not earlier in the HE (Bertrand et al., 2010a; Lin et al., 2005; Zhen et al., 2013).

Reporter lines utilizing short promoter sequences often do not contain all cis-regulatory elements since enhancers can be found at large distances from the promoter region, both up- and downstream (Stadhouders et al., 2012). As a solution to this issue, bacterial artificial chromosomes (BACs) have been used. BACs can hold genomic fragments as large as 300 kb that often include the complete structure of a gene including long-range cis-regulatory elements (Giraldo and Montoliu, 2001; Heintz,

2001). Hence, BAC transgenes often direct gene expression at physiological levels, faithfully recapitulating the developmental timing and expression patterns of endogenous genes (van Keuren et al., 2009). BAC reporter transgenes can be created by recombineering fluorescent reporters in bacteria and are a well established technique to generate transgenic mouse models (van Keuren et al., 2009). In addition, optimised protocols for zebrafish transgenesis based on Tol2 transposase mediated BAC integration have been established recently, allowing for the efficient generation of new BAC reporter lines (Fuentes et al., 2016; Kraus et al., 2014; Suster et al., 2011)

3.1.1 Chapter aims

There are 3 main aims in this chapter:

1. The generation of a novel *runx1* bacterial artificial chromosome (BAC) reporter line.
2. The thorough characterization of the novel *TgBAC(runx1P2:Citrine)* reporter line.
3. The establishment of a reliable protocol for the isolation of HE cells.

First, I will gather all possible information about the zebrafish *runx1* locus and its regulation. Second, BACs containing the *runx1* locus will be characterised and analysed for their utility to generate a novel *runx1* reporter line. Third, transgene expression during the primitive and definitive haematopoietic development of a *TgBAC(runx1P2:Citrine)* reporter line will be characterised. Fourth, a FACS protocol that allows the isolation of a pure population of HE cells will be established. Last, I will present data to explain the presence of weak Citrine fluorescence detected in the DA roof of the *TgBAC(runx1P2:Citrine)* reporter line around 24-36 hpf.

3.2 Results

3.2.1 The zebrafish *runx1* locus

Similar to other model organisms, the zebrafish *runx1* gene is controlled via two alternative promoters, P1 and P2 (Lam et al., 2009; Levanon and Groner, 2004). To generate a new zebrafish *runx1* reporter line we first wanted to collect all available information about the *runx1* locus and its regulation. Unfortunately, the annotation of the zebrafish *runx1* gene locus has not been resolved completely and remains so for the most recent version of the zebrafish genome assembly GRCz10/danRer10 (GRCz10) released in May 2015. For example, the Ensemble genome browser describes only the first three exons of the *runx1P2* transcript on chromosome 1. By contrast, the UCSC genome browser describes only the last four *runx1* exons placed on a pseudo-chromosome. For the previous Zv9/danRer7 (Zv9) genome annotation, the UCSC genome browser depicts the *runx1* locus as split into two genes facing in opposite directions (Figure 3-1 A).

The zebrafish *runx1* cDNA sequence had been sub-cloned by two labs independently (Kalev-Zylinska et al., 2002; Kataoka et al., 2000). The corresponding GeneBank files are AB043787 (2018 bp) and AF391125 (2543 bp) and both contain the full coding region of the P2-promoter specific isoform (*runx1b*). A Clustal Omega sequence alignment revealed that the two sub-cloned cDNAs differ in the length of their 3' and 5' UTRs (Appendix 1). The coding sequences are identical except for eight silent single nucleotide polymorphisms (SNPs). The protein sequence for Runx1b consists of 451 amino acids (AA) (Appendix 2).

The P1-promoter specific coding region was first identified by Lam et al. (2009), and was later refined by Marsman et al. (2014) who identified the *runx1P1* transcriptional start site (TSS) plus additional unique 5' sequences for the *runx1P1* transcript (Figure 3-1 B). The P1 promoter-derived transcript is called *runx1c*.

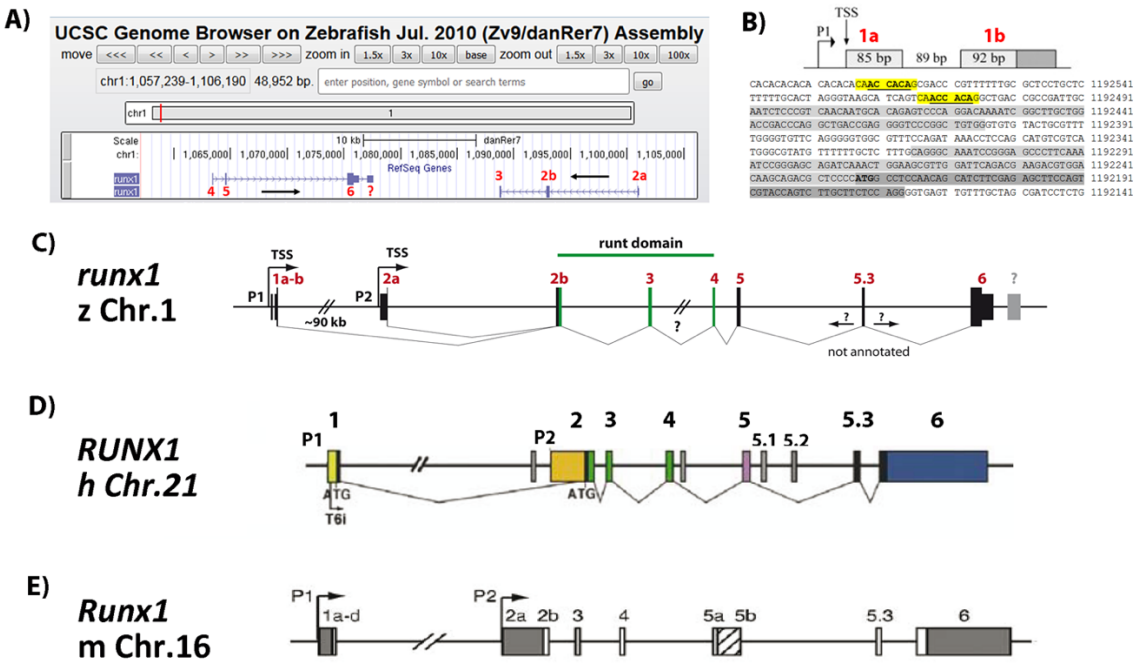


Figure 3-1: Annotation of the zebrafish *runx1* gene locus.

A) The Zv9/danRer7 genome assembly in the UCSC genome browser shows the *runx1* gene locus as split into two genes facing in opposite directions (black arrows). The P1 promoter region is not annotated. B) The P1 specific promoter region was identified by Lam et al. (2009) and Marsman et al. (2014). Image modified from Marsman et al. (2014). Two perfectly conserved Runx1-binding sites (5'-CAACCACAG-3') are highlighted in yellow. In humans, these binding sites are found in a short 18 bp sequence 5'-CAACCACAGAACCACAAG-3' that is conserved across all 3 RUNX genes (Levanon and Groner, 2004). C) Improved annotation of the zebrafish *runx1* gene locus. Exon sizes and distance are to scale. The naming of the exons was adjusted according to human *RUNX1* (D) and mouse *Runx1* (E) annotations. D) The human *RUNX1* gene locus. Image modified from Levanon and Groner (2004). E) The mouse *Runx1* gene locus. Image modified from Bee et al. (2009). h: human; m: mouse; z: zebrafish.

3.2.1.1 Splicing of the zebrafish P1 and P2 derived *runx1* isoforms involves a non-species conserved intron

To get better insights into the zebrafish *runx1* gene structure, the coding sequence of the zebrafish *runx1b* transcript (RefSeq:NM_131603; identical to the GenBank:AF391125 sequence) was compared to the coding regions of the human *RUNX1b* (Ensembl transcript ID: ENST00000344691.8) (Figure 3-1 D) and mouse *Runx1b* (Ensembl transcript ID: ENSMUST00000168195.7) (Figure 3-1 E) transcripts by using Clustal Omega sequence alignments (Appendix 3). The zebrafish sequence showed an overall 69.9 % similarity to the human sequence (69.7 % to mouse). The degree of conservation was even higher at the level of the AA sequence (Appendix 4). Here, the similarity

between the human and the zebrafish protein sequence was 78.6 % (77.2 % to mouse). Focusing on the highly-conserved Runt domain (RD), the zebrafish transcript showed an 84.4 % similarity to human (81.8 % to mouse). At the protein level, the similarity between the human and zebrafish RD was 96.7 % (96.9 % to mouse), with overall only four AA substitutions. All four substitutions involved only AAs with strongly similar properties.

Overall, the zebrafish *runx1b* transcript has the same number of exons as the human *RUNX1b* and mouse *Runx1b* transcripts with one exception (Figure 3-1 C-E): exon 2 (the first exon of the *runx1b* transcript) is split into two, separated by a non-species conserved intron. In view of the high conservation of the gene structure across the species, the nomination of the zebrafish exons was adjusted to that of the mouse and human nomenclature (Bee et al., 2009; Levanon and Groner, 2004). The two separated halves of the zebrafish exon 2 are now called 2a and 2b, respectively (Figure 3-1 C).

Next, a multiple sequence alignment was performed that incorporated the P1 and P2 specific transcripts of both human and zebrafish (Figure 3-2 A). In humans, the P1 specific coding exon (exon 1) splices into the exon 2 via an in-exon splice site, thereby skipping 16 bp that are specific for the P2 transcript (Levanon and Groner, 2004). The multiple sequence alignment highlighted that the position at which the zebrafish exon 2 is split into 2a and 2b is at a similar position as the in-exon splice site of the human (and mouse) transcript. Consequently, both the zebrafish P1 specific exon 1 and the P2 specific exon 2a splice into exon 2b (Figure 3-1 C). As for the two human major transcripts (*RUNX1C* and *RUNX1B*), the two zebrafish isoforms only differ in the 5' coding nucleotides (58 bp for P1 and 28 bp for P2) and in the 5' UTRs. The remaining 1328 coding nucleotides are shared between both isoforms. In conclusion, human (mouse) and zebrafish use two different strategies to generate the promoter-specific isoforms, but both strategies result in a similar outcome.

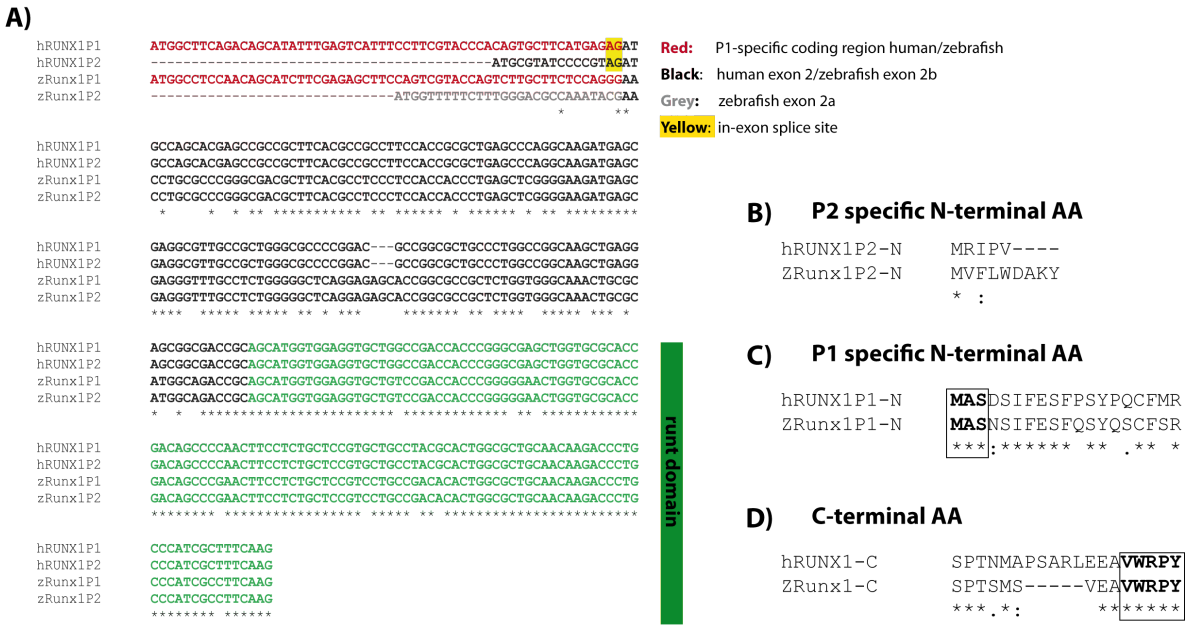


Figure 3-2: Species specific differences between the human *RUNX1* and zebrafish *runx1* gene.
A) Multiple sequence alignment between the first coding exon of the P1 and P2 specific transcripts in human and zebrafish. The human exon 2 and the zebrafish exon 2b sequences are shown in black with the runt domain highlighted in green. The P1 specific sequences are shown in red. The zebrafish exon 2a specific sequence is shown in grey. The in-exon splice site for the human transcript is highlighted in yellow. B) Sequence alignment between the P2 specific AAs. The peptide MRIPV is conserved across all 3 human *RUNX* and mouse *Runx* genes. C) Sequence alignment between the P1 specific AAs. The highlighted peptide MAS is conserved across all 3 human *RUNX* and mouse *Runx* genes. D) Sequence alignment between the most C-terminal AAs. The highlighted peptide motif VWRPY (for the interaction with the co-repressor Groucho is conserved across all 3 human *RUNX* and mouse *Runx* genes.

The multiple sequence alignment highlighted additional differences regarding the composition of the P2-specific coding regions. In humans, the P2-specific coding sequence is 16 bp long and encodes a peptide sequence (MRIPV) that is conserved across all three human *RUNX* and mouse *Runx* genes (Levanon and Groner, 2004). The zebrafish P2-specific coding sequence consists of 28 bp that lacks any conservation (Figure 3-2 B). In contrast, the P1-specific coding regions for both human/mouse and zebrafish are 58 bp long and the AA sequence shows a high level of conservation (73.7 % similarity), especially for the first 10 AA (Figure 3-2 C). Finally, we also analysed the conservation of the C-terminal sequence that harbours the VWRPY motif which allows for interaction with the RUNX1 co-repressor Groucho (Ito et al., 2015). The zebrafish Runx1 sequence shows a high degree of conservation despite the absence of a 5 AA stretch 3 AAs upstream of the VWRPY motif (Figure 3-2 D).

3.2.1.2 *Runx1* expression in the zebrafish HE is initiated via the P2 promoter

Studies in mouse have shown that promoter usage of the two alternative *Runx1* promoters is highly regulated during haematopoiesis. It follows a specific pattern in which the P2 promoter initiates primitive haematopoiesis, while the P1 promoter is active during definitive haematopoiesis (Pozner et al., 2007; Sroczynska et al., 2009). A closer assessment of *Runx1* expression in the mouse HE revealed that it is predominantly initiated via the P2 promoter, which seems to be most critical for the emergence of HSCs, with initiation of P1 activity shortly afterwards (Bee et al., 2010; Sroczynska et al., 2009).

The usage of the zebrafish P1 and P2 *runx1* promoters during the definitive wave of haematopoiesis has only been studied with the help of reporter lines based on short promoter sequences (~12,000 bp for the P1 promoter and ~8,000 bp for the P2 promoter) (Lam et al., 2009). Reporter activity in these transgenic lines suggested that, like in other organisms, the P2 promoter drives *runx1* expression in the HE. To clarify the endogenous activity of the alternative *runx1* promoters during embryonic haematopoiesis, isoform-specific *in situ* hybridisation (ISH) probes were generated (Figure 3-3 B). These ISH probes had to be relatively short to be isoform-specific (222 bp for P1 and 287 bp for P2). In addition, long and short "pan-*runx1*" probes were used as controls. Since we were mostly interested in the *runx1* promoter usage within the HE, embryos at 30 hpf (and 48 hpf) were investigated, a time point when haemogenic expression of *runx1* is well established. Additionally, embryos of the 10-somite stage (14 hpf) were analysed to investigate the *runx1* promoter usage during the primitive wave of haematopoiesis in the posterior lateral plate mesoderm (PLM). Only the P2 specific transcript was detectable by ISH in both the PLM at 10-somites and in the HE around 30/48 hpf. In agreement, exclusive activity of the P2 promoter in the PLM has also been shown recently by others (Marsman et al., 2014). More importantly, we could show that only the *runx1P2* driven transcript is expressed in the zebrafish HE (Figure 3-3 C).

We also found that *runx1* expression in the olfactory placodes (OP) is initiated via the P2 promoter in 10 somite stage embryos but is later driven via both promoters (Figure 3-3 C; green arrow). The detection of the P1 derived transcript in the OP simultaneously serves as a positive control for the functionality of the P1-specific ISH probe.

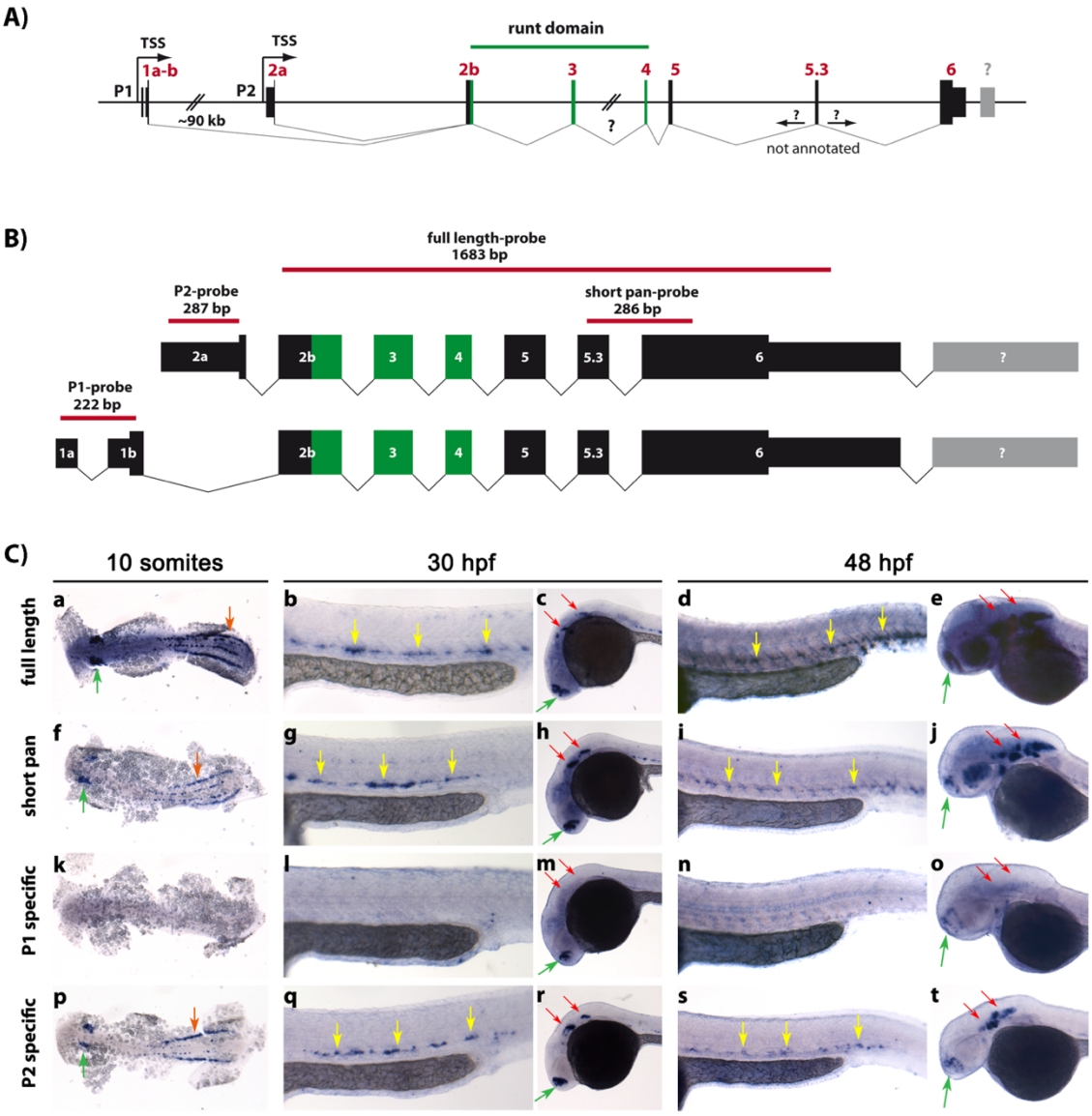


Figure 3-3: Analysis of the zebrafish *runx1* promoter usage.

A) Schematic of the zebrafish *runx1* gene locus. B) Schematic of the zebrafish *runx1b* and *runx1c* transcripts generated by the P2 and P1 promoter, respectively. The binding sites and length of transcript specific and pan-binding ISH probes are indicated. C) ISH results highlighting the different *runx1* promoter usages. Fish of three different stages (10 somites, 30 hpf and 48 hpf) have been analysed. Green arrows: olfactory placodes (OP). Orange arrows: posterior lateral plate mesoderm (PLM). Yellow arrows: haemogenic endothelium (HE). Red arrows: specific *runx1*$^{+}$ brain regions.

3.2.2 Generation of a novel *runx1* reporter line based on a zebrafish BAC

Because it has been reported that *runx1* reporter lines based on short promoter sequences show ectopic expression (Lam et al., 2009), we wanted to make use of a BAC transgenesis strategy to generate a novel *runx1* reporter line. BACs can often closely recapitulate the expression patterns of endogenous genes (Heintz, 2001; Suster et al., 2011). The analysis of the *runx1* promoter usage directed us to identify BACs that would allow the generation of a *runx1P2* reporter line.

3.2.2.1 Characterisation of zebrafish *runx1*-containing BACs

Since the *runx1* gene locus is poorly annotated, it was not possible to identify annotated *runx1* BACs by simply screening the genome browsers (Bussmann and Schulte-Merker, 2011; Suster et al., 2011). Fortunately, the Crosier lab had identified 4 BACs containing different regions of the *runx1* gene (Lam et al., 2009) by Southern blotting to the BAC Zebrafish High-Density Filter Library (Incyte Genomics, Wilmington, DE). This library has been discontinued, but the Crosier lab was kindly willing to send us these BACs as a gift. The BAC IDs are 14i20, 74j23, 97a02 and 135g16 (Figure 3-4 A).

To identify suitable BACs containing the P2 promoter, all BACs had first to be characterized in greater depth. The characterisation was performed via genotyping PCR for certain *runx1* exons (exon 1, 2a, 2b and 6) and via BAC-end sequencing. The genotyping results can be seen in Figure 3-4 B.

The BAC characterization highlighted that the 135g16 BAC does not contain the P2 promoter regions. Hence, this BAC was dismissed from further analysis/investigation.

Similarly, the 74j23 BAC was dismissed since it contains the P2 promoter with only ~4,500 bp of its upstream region. This was far less than the ~8,000 bp that has been used previously for a short promoter reporter line which shows ectopic expression in the vasculature (Lam et al., 2009).

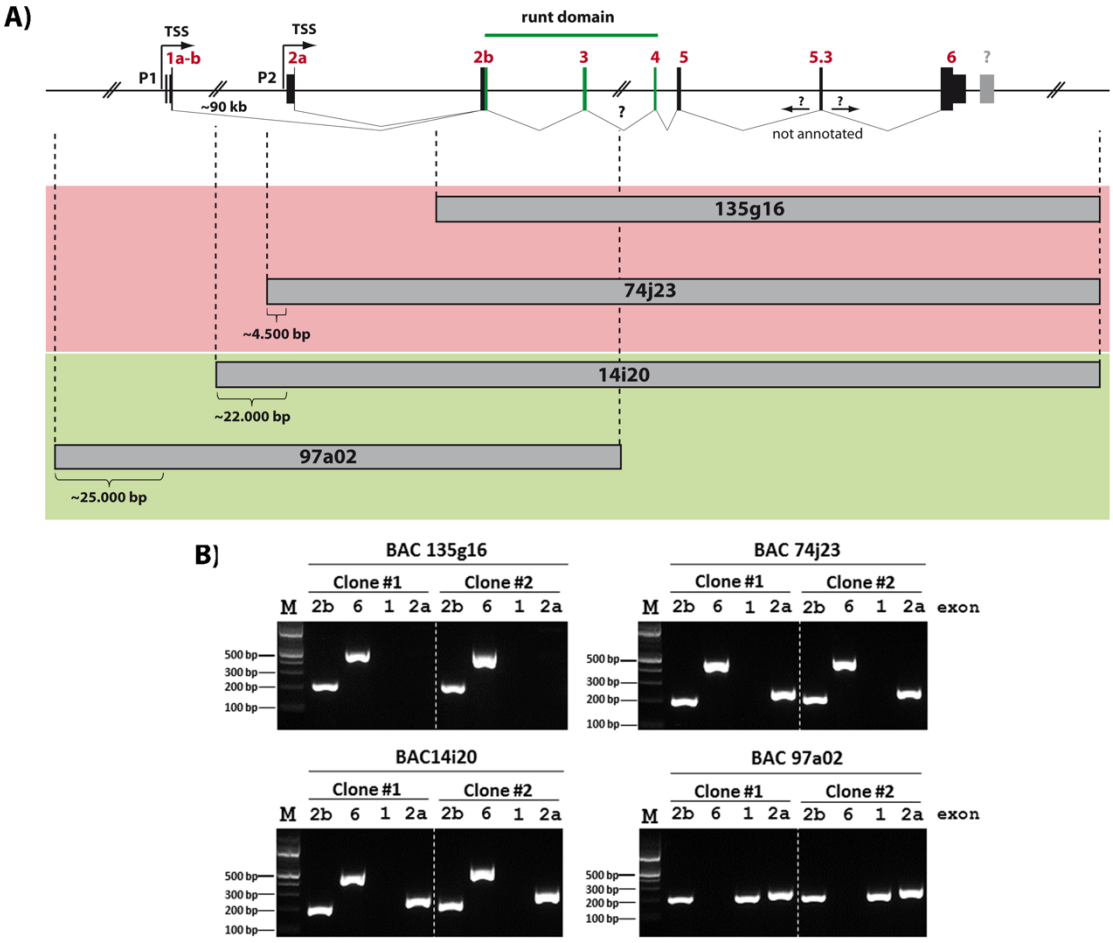


Figure 3-4: Characterization of zebrafish BACs containing different regions of the runx1 locus.

A) Schematic illustration of the zebrafish runx1 locus and the region covered by each of the four BACs (14i20, 74j23, 97a02 and 135g16). BACs on red background are not suitable for the generation of reporter constructs. BACs on green background cover significant regions of the P2 promoter. B) Genotyping examples for each of the 4 analysed BACs showing the presence/absence of the exons 1, 2a, 2b and 6.

The 14i20 BAC lacks the P1 promoter region but does contain the P2 promoter with additional ~22,000 bp of upstream sequence. Therefore, we considered this BAC as potentially useful. Nevertheless, this BAC contains a very large region downstream of the *runx1* gene. The determined position of the BAC end on the UCSC Zv9 annotation was chr1:684,867 (Table 3-1). Based on the annotation, this BAC does contain additional genes including *wu:fa56d06*, *setb2*, *creg1*, *n6amt1* and *gart*.

Table 3-1: Genomic location of BAC ends as determined by BAC-end sequencing.
Sequencing results were mapped against the Zv9 annotation using the BLAT tool in the UCSC browser. The sequencing for the upstream end of the 74j23 BAC failed.

BAC	Genomic location upstream end	Genomic location downstream end
14i29	chr1:684,867	chr1:1,123,988
74j23	/	chr1: 1,105,655
97a02	chr1:1,086,776	chr1: 1,217,212
135g16	chr1:691,962	chr1: 1,098,226

Lastly, the 97a02 BAC is the only BAC containing the P1 promoter region with additional ~25,000 bp of upstream sequence. This BAC also contains the P2 promoter and the ~90,000 bp large intron between P1 and P2. Moreover, it contains exon 2b and 3, but none of the other exons further downstream. We also considered this BAC as potentially useful.

At this point it was not clear if the lack of downstream regions in the 97a02 BAC might be a potential disadvantage, since it cannot be ruled out that these regions contain important regulatory sequences. However, studies in mouse have shown that the most important haematopoietic *Runx1* enhancer (+23) lies ~23,000 bp downstream of the P1 promoter within the first exon (Nottingham et al., 2007). Even though a conservation of this sequence is not detectable in zebrafish (Lam et al., 2009), the presence of the complete first intron was considered promising since it might harbour important regulatory sequences.

On the other hand, the discontinuance of the 97a02 BAC within the RD generated also an advantage: this BAC will not allow for the introduction of additional functional copies of *runx1* by either the P1 or P2 promoter.

In conclusion, two potentially useful BACs could be identified (14i20 and 97a02) that covered the P2 promoter region plus a substantial amount of upstream sequence. None of them covered the whole *runx1* locus. Both BACs were taken forward for the generation of a *runx1P2* reporter construct.

3.2.2.2 The 97a02 BAC allows for the generation of a novel *runx1P2* reporter line.

Protocols for BAC transgenesis have been successfully transferred to zebrafish (Fuentes et al., 2016; Kraus et al., 2014; Suster et al., 2011). These protocols are based on a transposon mediated transgenesis approach (Tol2), whereby the transposon cassette is introduced into the BAC backbone. The protocol and the plasmids containing the required recombineering cassettes were kindly provided by the Sauka-Spengler lab. The BAC cloning followed the Suster et al. (2011) protocol. For the *runx1* BAC constructs, a *Citrine* reporter sequence was placed directly downstream of the P2 ATG for both BACs, 14i20 and 97a02, thereby replacing the coding sequence of the 2a exon (Figure 3-5 A).

To test for the functionality of the generated reporter constructs, purified BACs were microinjected into early 1-cell stage embryos and analysed for transient expression. Citrine fluorescence could be detected for both constructs (*BAC^{97a02}(runx1P2:Citrine)* and *BAC^{14i20}(runx1P2:Citrine)*), suggesting that both transgenic constructs were functional and capable of driving *in vivo* reporter gene expression. However, injection of the *BAC^{14i20}(runx1P2:Citrine)* resulted in a very low survival rate (13.8 %) (Figure 3-5 B) accompanied by a high rate of severe deformations in surviving fish (Figure 3-5 C). In contrast, injection of the *BAC^{97a02}(runx1P2:Citrine)* at similar quantities resulted in a much higher survival rate (46.2 %) (Figure 3-5 B) and most embryos were healthy (Figure 3-5 C).

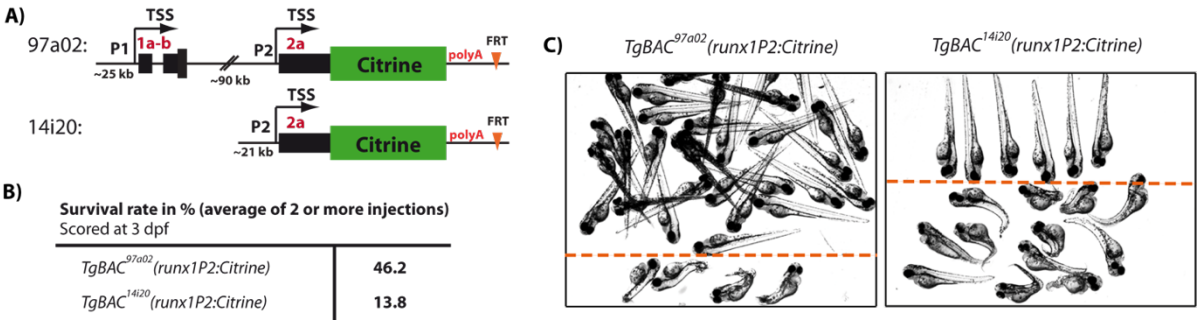


Figure 3-5: Generation of runx1P2 reporter constructs based on the two BACs 97a02 and 14i20.
A) Schematic of recombineered BACs. The Citrine reporter cassette was placed directly downstream of the P2 ATG replacing the P2 specific coding regions. B) Survival rates of embryos injected with the recombineered BACs. C) Exemplary pictures showing the health state of embryos after BAC injection.

The severe adverse effects on embryos after injection with the *BAC^{l4i20}(runx1P2:Citrine)* is most likely caused by the presence of the additional genes on that BAC. The severity of these adverse effects made this vector unsuitable for the generation of a stable transgenic line.

To allow for further analysis, the *BAC^{97a02}(runx1P2:Citrine)* was re-injected to perform a more comprehensive characterisation of the transient reporter expression. Injected fish were scored based on the presence of Citrine fluorescence in various tissues at around 48, 72, and 96 hpf. This thorough characterisation included *runx1P2:Citrine* expression in neuronal and haematopoietic tissues, as well as venous endothelium as a negative control. Prominent fluorescence was observed in the OP (Figure 3-6 B) for ~75 % of injected embryos at all three time points analysed (Figure 3-6 H). Citrine fluorescence was further detected in a particularly long axon growing alongside the ventral line (VL-A) (Figure 3-6 C-E).

Most importantly, the presence of Citrine fluorescence in the HE and subsequent haematopoietic tissues including the CHT and the thymus was detectable. Here, ~25 % of the injected fish showed Citrine^{+} cells in the region of the HE and in the sub-aortic space at around 48 hpf (Figure 3-6 C, H), in agreement with a previous report that cells of the zebrafish HE bud out into this sub-aortic space before entering circulation (Kissa and Herbomel, 2010; Kissa et al., 2008).

At around 72 hpf, Citrine^{+} cells were detected in the CHT (Fig D, H). At around 96 hpf the number of Citrine^{+} cells in the CHT increased (Figure 3-6 E) and Citrine^{+} cells could also be detected in the thymus (Figure 3-6 F,G,H). No Citrine fluorescence was detected in venous endothelium (Figure 3-6 H) or in further arterial endothelium but the HE.

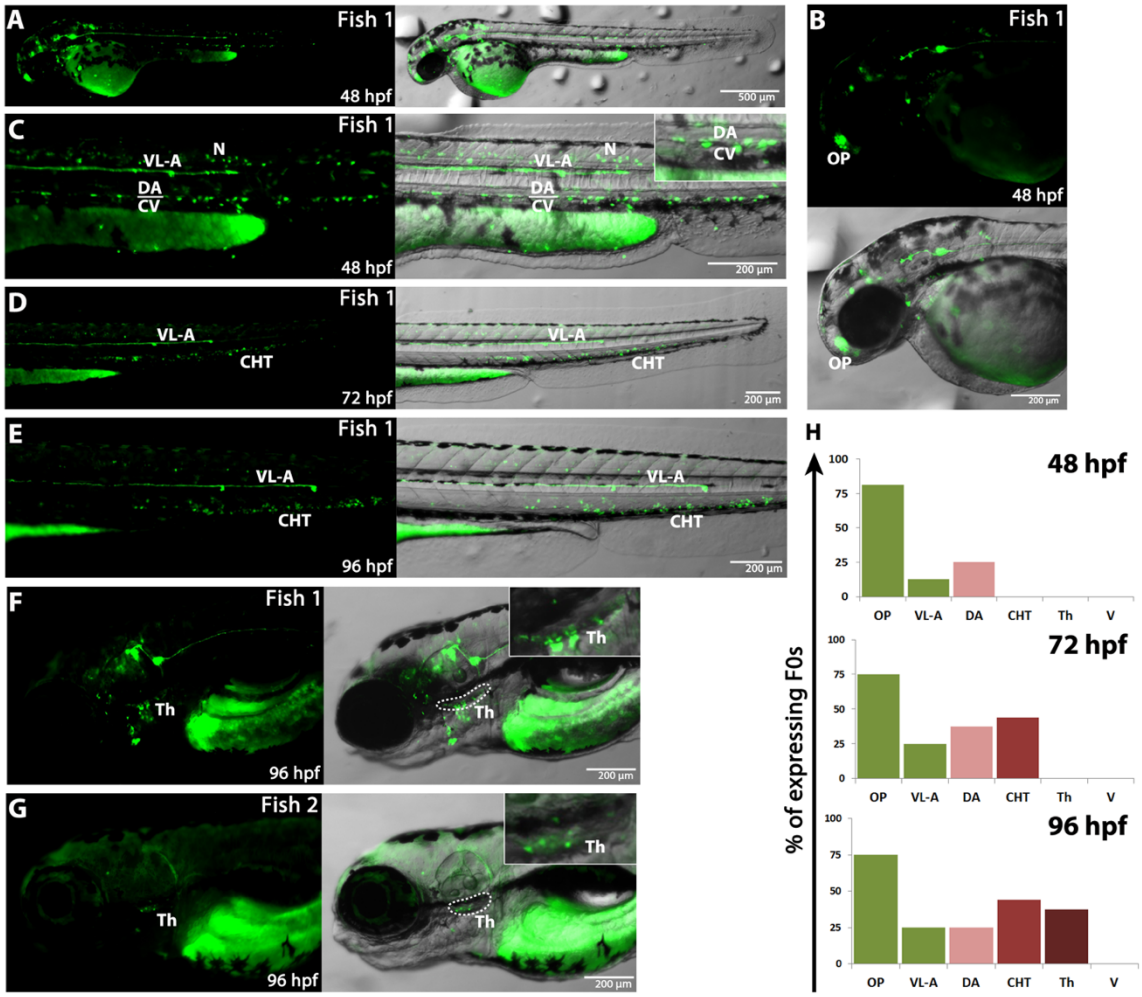


Figure 3-6: Transient expression analysis of embryos after injection with the recombineered *97a02* BAC.
A-G) Example of transgene expression in the same embryo (fish 1) followed over a time-course of 48 to 96 hpf highlighting different tissues. G) Example of a second embryos (fish 2) at 96 hpf. H) Quantification of injected embryos showing transgene expression in indicated tissues over a time-course of 48 to 96 hpf. CHT: caudal hematopoietic tissue; DA: dorsal aorta; OP: olfactory placodes; Th: thymus; V: vein; VL-A: ventral line axon.

3.2.2.3 Generation of a stable *TgBAC(runx1P2:Citrine)* reporter line

To generate a novel *runx1* reporter line, the recombineered 97a02 BAC containing the Citrine reporter cassette downstream of the P2 ATG was injected together with *Tol2* mRNA, and embryos were raised to adulthood. The injected fish were crossed with Wt fish and the F1 embryos were screened for stable BAC integration. In total, 3 founder fish were identified and 2 stable transgenic lines were established. These lines were subsequently called line VI ($TgBAC^{97a02}(runx1P2:Citrine)^{VI}$) and line XIII ($TgBAC^{97a02}(runx1P2:Citrine)^{XIII}$).

Both reporter lines exhibit a comparable *Citrine* expression pattern apart from two differences: First, the level of expression is much stronger in line XIII (Figure 3-7). Second, line XIII shows a strong Citrine fluorescence in neurons located in the spinal cord that is absent in line VI. Despite these differences, both lines show Citrine^{+} cells in all haematopoietic organs including the CHT, the thymus, and additionally in the kidney, which is the HSC niche equivalent to the bone marrow in higher vertebrates (Murayama et al., 2006).

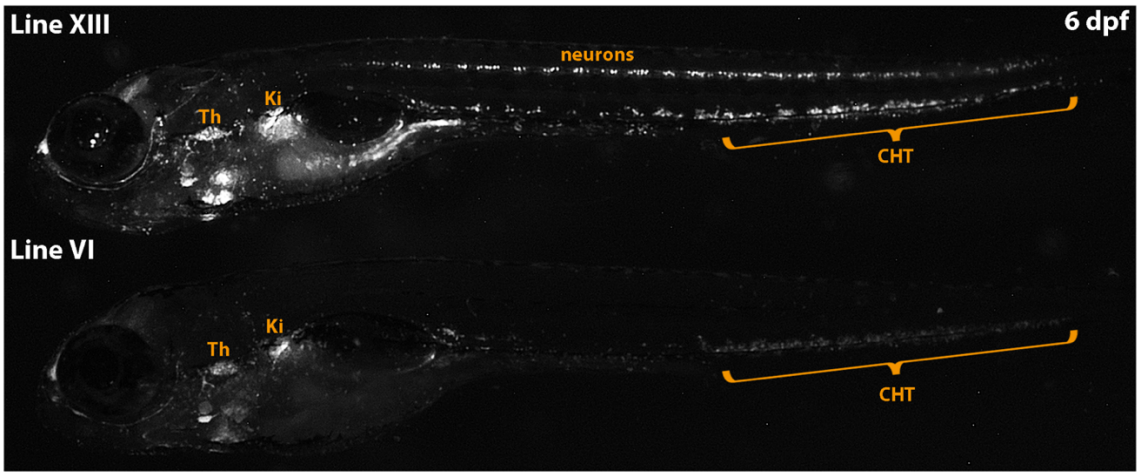


Figure 3-7: ***Citrine*** **expression in stable** ***TgBAC^{97a02}(runx1P2:Citrine)*** **reporter lines.**
Two stable lines (XIII and VI) from two different founders were generated. CHT: caudal hematopoietic tissue; Ki: kidney; Th: thymus.

Outcrossing F1 fish with Wt fish always resulted in a Mendelian distribution, indicating a single BAC integration event for both lines. Since the stronger level of fluorescence in line XIII was considered beneficial, only this line was used for further experiments. From now on the line XIII will be referred to as *TgBAC(runx1P2:Citrine)*.

3.2.3 In depth analysis of reporter expression in *TgBAC(runx1P2:Citrine)* embryos

To gain insights into the activity kinetics of the novel *TgBAC(runx1P2:Citrine)* line, Citrine reporter expression was analysed in greater depth. This analysis was centred around both primitive haematopoiesis (expression in the PLM) and definitive haematopoiesis (expression in the HE and subsequent haematopoietic niches).

3.2.3.1 Transgene expression during early haematopoiesis recapitulates *runx1* expression in the PLM

It has been reported that zygotic *runx1* expression in the PLM initiates at around 12 hpf (6-somite stage) (Burns et al., 2002; Kalev-Zylinska et al., 2002). ISH for *Citrine* mRNA in the *TgBAC(runx1P2:Citrine)* line showed detectable expression in the PLM at 11 hpf (3-somite stage), at least 1 hr earlier than endogenous *runx1* expression becomes detectable (Figure 3-8 A). We speculate that this is simply due to the shorter length of the *Citrine* transcript (717 bp) compared to the greater length of the endogenous *runx1b* transcript of at least ~26,000 bp including introns (a more detailed estimation is not possible due to the incomplete annotation of the *runx1* locus). The first Citrine fluorescence in the PLM is detectable at around 12 hpf (data not shown). The delay in protein fluorescence detection compared to *Citrine* mRNA detection is most likely caused by the time needed for the folding, maturation and accumulation of the Citrine protein. Overall Citrine fluorescence was detectable around the same time as endogenous *runx1* transcripts.

Fluorescent double ISH for *Citrine* and *runx1* mRNA at 15 hpf (12 somite) showed an almost 100 % overlap in the PLM, confirming the reliability of the *runx1P2:Citrine* reporter for haematopoietic tissue (Figure 3-8 B). In contrast, expression in the Rohon-Beard neurons (RBNs) is only detectable for endogenous *runx1*, but not for *Citrine*.

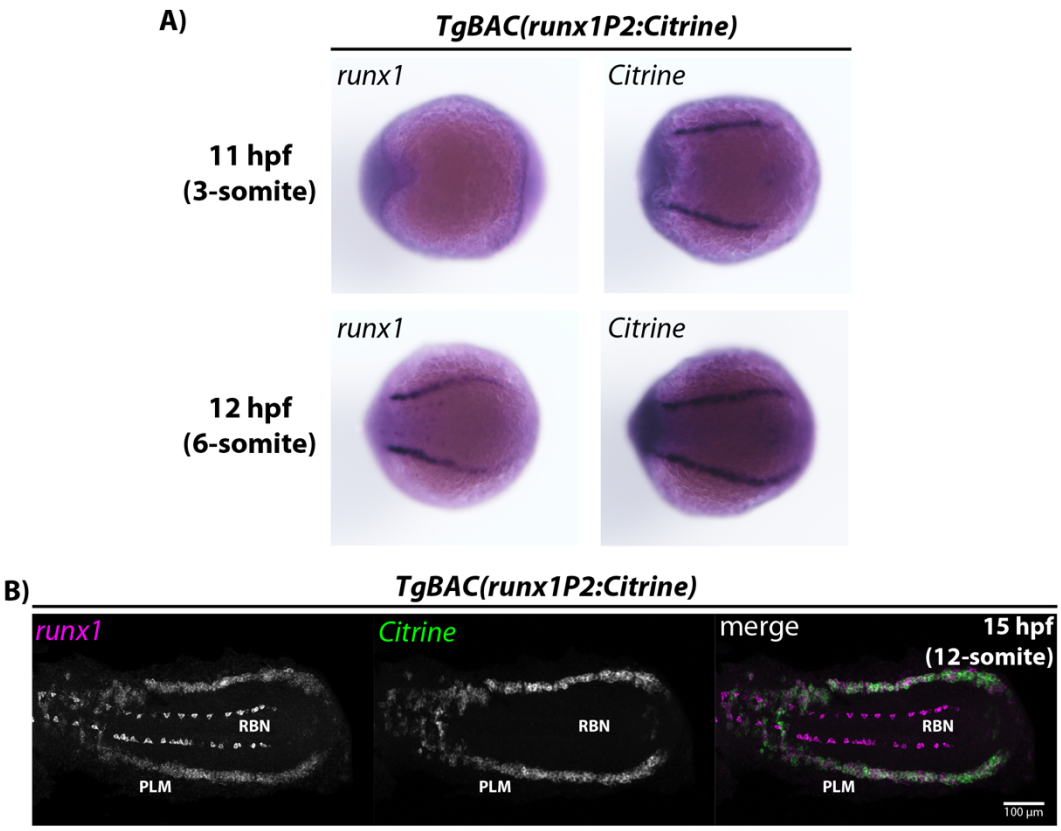


Figure 3-8: ***Citrine*** **transgene expression in the PLM of** ***TgBAC(runx1P2:Citrine)*** **embryos.**
A) ISH for *Citrine* and *runx1* revealed that *Citrine* mRNA can be detected in the PLM around 11 hpf, at least 1 h earlier then *runx1*. B) Double fluorescent ISH for *runx1* and *Citrine* in 15 hpf (12 somite stage) embryos proves a reliable expression of the *Citrine* reporter in hematopoietic tissue. The *Citrine* reporter lacks expression in Rohon-Beard neurons (RBNs) (double fluorescent ISH Experiment performed by Monika Krecsmarik).

From 13 hpf onwards, Citrine fluorescence can also be detected in the OP, and weaklier in cells of the ALM and in the Kupffer's vesicle (KV) (Figure 3-9). Expression of *runx1* in the KV has not been reported yet. Citrine^{+} cells of the PLM migrate towards the midline to form the ICM that harbours the primitive erythrocytes (Gering and Patient, 2005). This migration could easily be followed under the fluorescent microscope (Figure 3-9). Hereby, PLM cells at the anterior end start to migrate first and the more posterior cells follow in a zipper like manner.

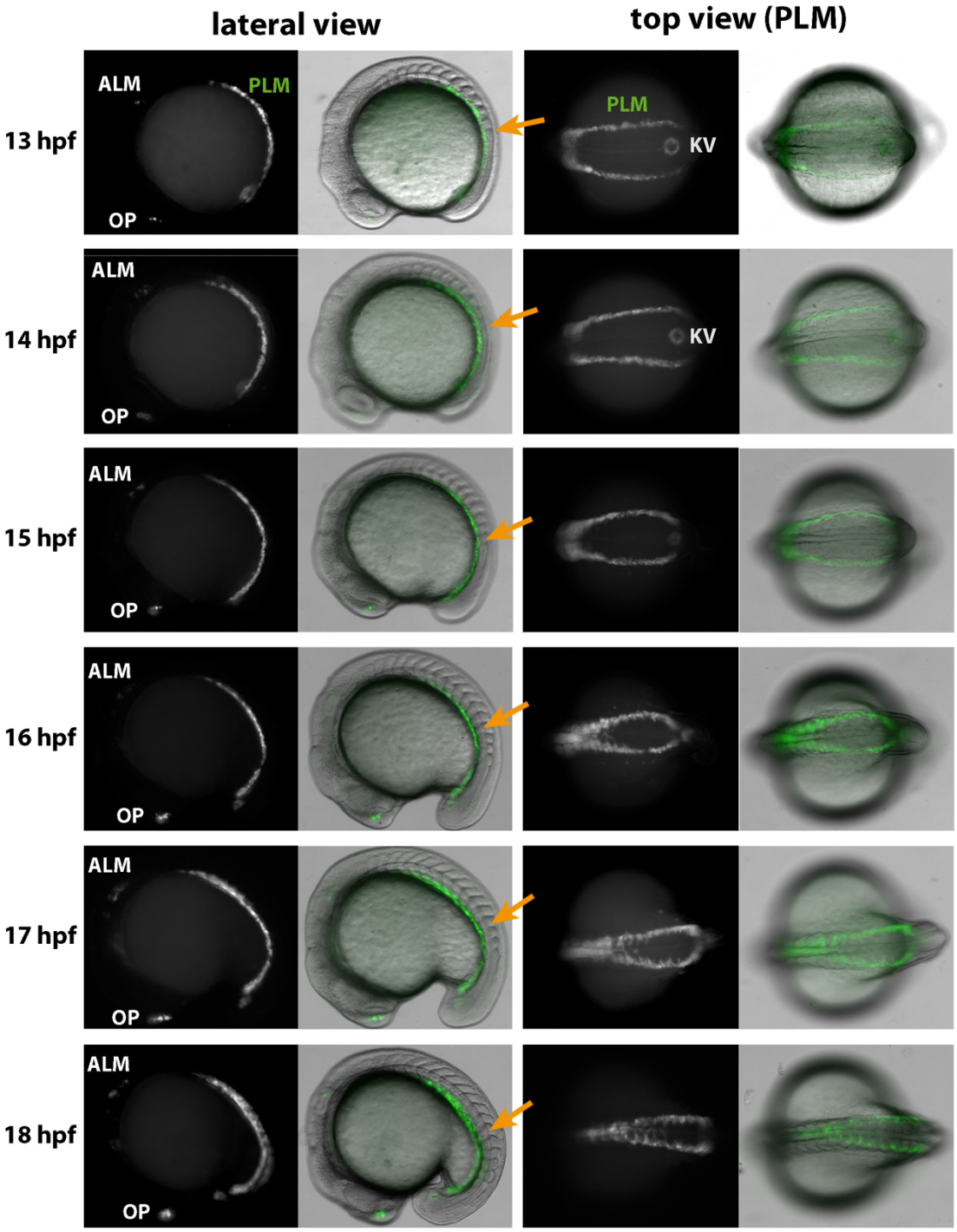


Figure 3-9: Citrine fluorescence analysis during primitive haematopoiesis of *TgBAC(runx1P2:Citrine)* embryos. Citrine fluorescence can be detected in *runx1* expressing tissues including the ALM, PLM and the OP. Citrine fluorescence in early embryos can also be detected in Kupffer's vesicle (KV). The migration of PLM cells towards the midline can be followed easily over the next 5 hr.

3.2.3.2 Transgene expression during definitive haematopoiesis recapitulates *runx1* expression in the HE

Runx1 expression in the primitive erythrocytes is down-regulated at around 18 hpf, when haematopoiesis is transitioning to the definitive wave (Kalev-Zylinska et al., 2002). Initiation of *runx1* expression in cells of the HE is detectable at around 23/24 hpf, and from 26 hpf onwards a relatively continuous population of *runx1* expressing HE cells is detectable (Kalev-Zylinska et al., 2002; Wilkinson et al., 2009). Further characterisation of the *TgBAC(runx1P2:Citrine)* line was centred around the HE and the CHT.

Like *runx1* expression, *Citrine* expression was detectable in the HE around 24 hpf (Figure 3-10 A). At around 30 hpf, when Citrine fluorescence in primitive erythrocytes has mostly disappeared, the detectable Citrine fluorescence faithfully mimics the expression pattern of endogenous *runx1* (Figure 3-10 B, C). Cells with high levels of Citrine fluorescence can be found all along the ventral wall of the DA. These cells also express the *mCherry* reporter marking endothelial cells in *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* double-transgenic embryos (Figure 3-10 C').

In addition, approximately 3-5 Citrine$^+$ cells with strong fluorescence can be detected in the region of the posterior blood island (PBI), close to the yolk-extension (Figure 3-10 C'; green arrows). Due to their location, it is very likely that these cells represent erythroid-myeloid progenitors (EMPs) (Bertrand et al., 2007).

Citrine expression in cells of the spinal cord was very prominent. A careful examination revealed that those spinal cord cells were also expressing endogenous *runx1* (Figure 3-10 B; orange arrow head), that however was very weak and not always detectable by ISH. This is most likely due to the sensitivity of the chromogenic ISH technique.

At around 50 hpf, spherical Citrine$^+$ cells in the trunk region can be detected mostly within the sub-aortic space (Figure 3-10 D). These cells are likely HSPCs that are going to enter circulation by passing through the CV wall (Kissa et al., 2008; Murayama et al., 2006). Subsequently, HSPCs in circulation will seed the CHT where they will undergo further maturation and proliferation (Murayama et al., 2006; Tamplin et al., 2015). Many Citrine$^+$ cells can be detected in the CHT, always near the vascular niche (Figure 3-10 E).

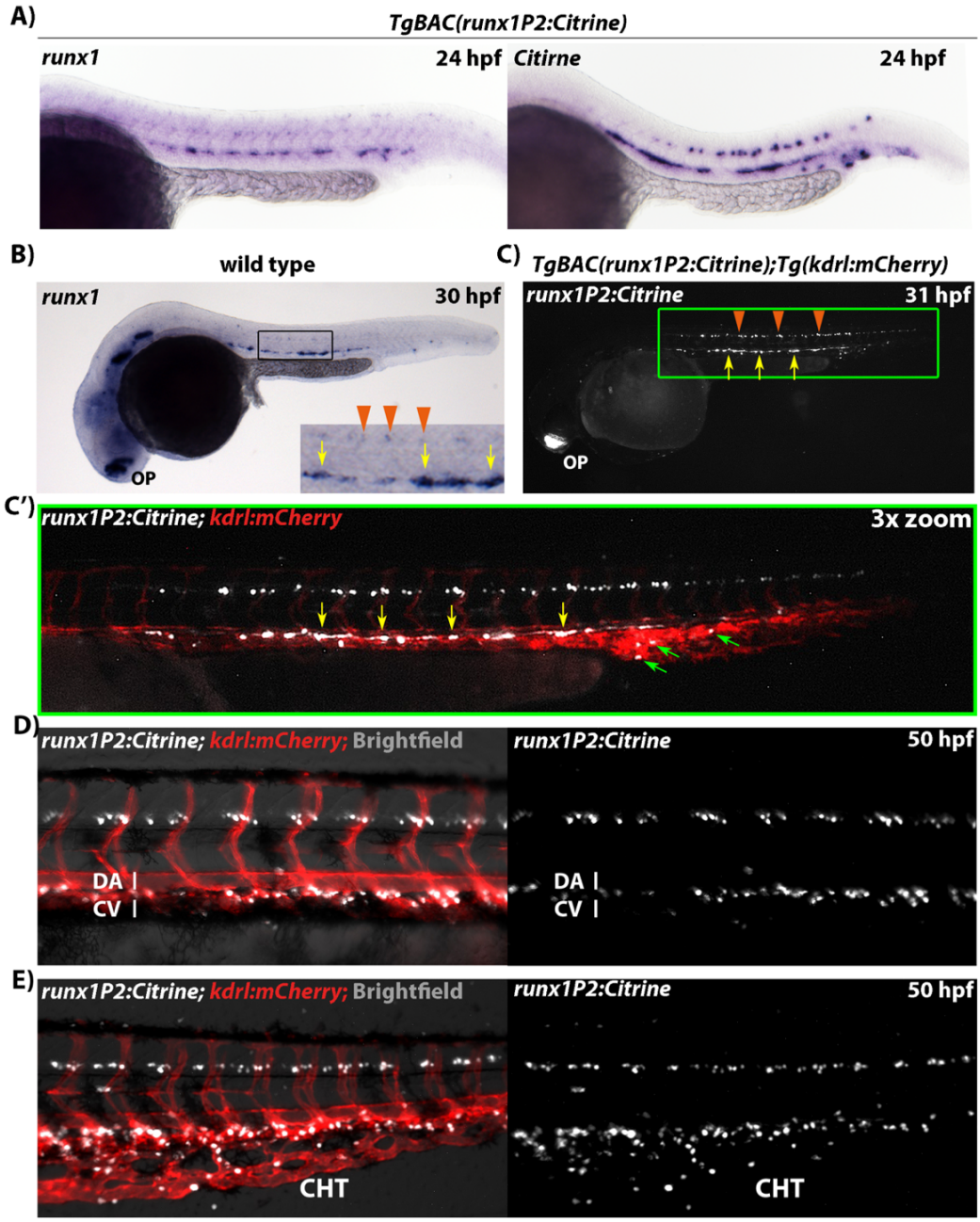


Figure 3-10: ***Citrine*** **expression analysis during definitive hematopoiesis in double transgenic** ***TgBAC(runx1P2:Citrine), Tg(kdrl:mCherry)*** **embryos.**

A) ISH for *runx1* and *Citirne* in *TgBAC(runx1P2:Citrine)* embryos at 24 hpf B) Endogenous *runx1* expression around 30 hpf. Yellow arrows indicate expression in the HE. Orange arrow-heads indicate expression in spinal cord neurons. C) Citrine fluorescence in double transgenic *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* embryos around 31 hpf mimics endogenous *runx1* expression. C') 3x zoom of the boxed area in (C) highlights Citrine^{+} cells in the region of the DA floor (yellow arrows) co-expressing *kdrl:mCherry*. Green arrows indicate Citrine^{+} cells in the posterior blood island (PBI) most likely representing EMPs. Citrine^{+} "neurons" in the spinal cord are indicated (orange arrow-heads). D-E) Citrine fluorescence in the trunk region (D) and the CHT (E) around 50 hpf. Spherical Citrine^{+} cells can be detected in the sub-aortic space (D). Spherical Citrine^{+} cells can also be detected in the CHT (E) always near the vasculature.

3.2.3.3 The *runx1P2:Citrine* transgene is Tal1b and Gata2b dependent but Runx1 independent

We conducted further experiments to investigate the upstream regulation of our *runx1P2:Citrine* reporter. The TFs *tal1b* and *gata2b* are well described haematopoietic genes and regulate the expression of *runx1* in the HE (Butko et al., 2015; Qian et al., 2007). Injection of morpholinos (MOs) against either *tal1b* (Qian et al., 2007; Zhen et al., 2013) or *gata2b* (Butko et al., 2015) not only led to a strong reduction of *runx1* but also of *Citrine* expression (Figure 3-11 A, B). Of note, *Citrine* expression in the spinal cord neurons is only affected by the *tal1b* MO. Integrity of the vasculature was not affected by either MO as assessed by ISH for *kdrl* (Figure 3-11 C).

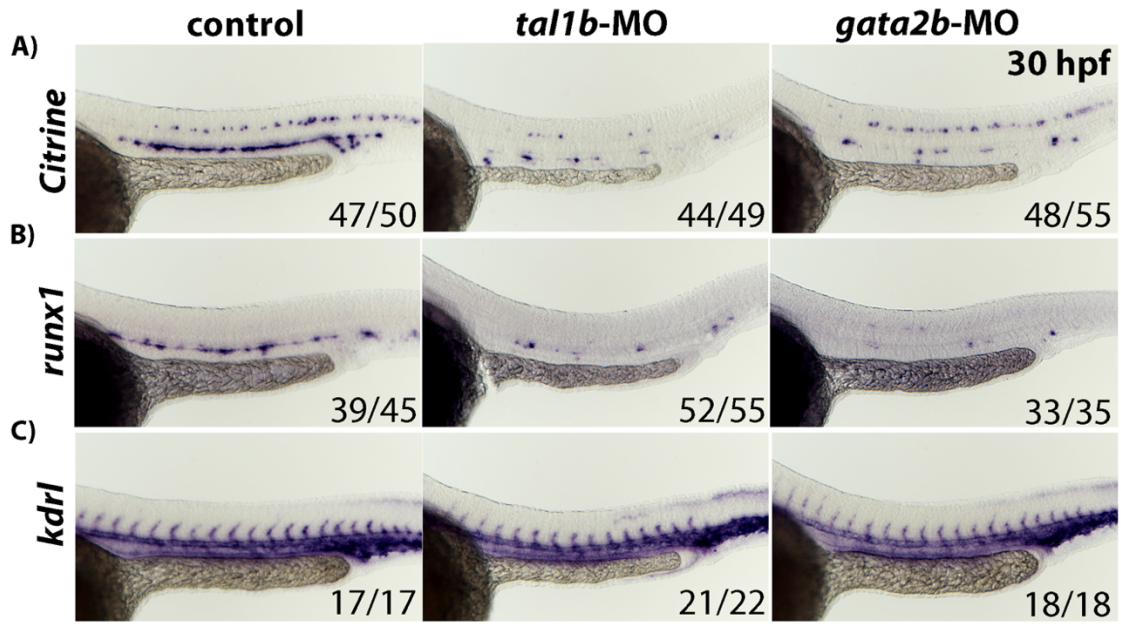


Figure 3-11: Analysis of *Citrine* transgene regulation in *TgBAC(runx1P2:Citrine)* embryos.
A-C) Tal1b and Gata2b are well known upstream regulators of *runx1* expression in the HE. *Citrine* (A) and *runx1* (B) expression are lost in the HE of *tal1b* MO and *gata2b* MO injected embryos. C) Vasculature integrity is not effected as indicated by *kdrl* expression.

Next, we wondered what the phenotype in the *TgBAC(runx1P2:Citrine)* line might look like on loss of Runx1 itself. In mouse, homozygous *Runx1*$^{-/-}$ mutants lack IACs (North et al., 1999) and further studies revealed that RUNX1 is required for the initiation of EHT (Lancrin et al., 2012; Thambyrajah et al., 2015). In zebrafish, live imaging of *runx1* morphants showed that lower numbers of HE cells initiated an EHT event and those that did burst (Kissa and Herbomel, 2010). A detailed analysis of

the zebrafish HE in a *runx1* mutant has not been reported yet. A zebrafish *runx1* mutant (*W84X*) that is thought to be a null allele has been identified in an ENU mutagenesis screen. This mutant harbours a G to A nucleotide substitution converting a Trp (amino acid 84) encoding triplet (UGG) to a premature stop codon (UGA) within the RD (Jin et al., 2009; Sood et al., 2010). Like *runx1* morphants, homozygous *runx1*$^{W84X/W84X}$ embryos lack definitive haematopoiesis.

We crossed our *TgBAC(runx1P2:Citrine)* reporter onto a *runx1*$^{W84X/W84X}$ background. At 3 dpf, *runx1*$^{W84X/W84X}$ embryos lack almost all Citrine^{+} cells in the CHT compared to *runx1*$^{+/+}$ or *runx1*$^{W84X/+}$ control fish (Figure 3-12 A, B). Whether the few remaining Citrine^{+} cells in the CHT are HSPCs or cells of other *runx1* expressing blood lineages, such as macrophages or EMPs (Bertrand et al., 2007) remains to be determined. Strikingly, Citrine^{+} cells of the HE in *runx1*$^{W84X/W84X}$ mutants retained an elongated endothelial cell shape and appeared to remain part of the DA lining (Figure 3-12 D). Additionally, endogenous *runx1* expression was still detectable in the region of the ventral DA around 3 dpf (Figure 3-12 F). By contrast, Citrine^{+} cells in the trunk region of *runx1*$^{+/+}$ embryos had a more spherical cell shape and appeared to be located within the sub-aortic space (Figure 3-12 C). Endogenous *runx1* expression in *runx1*$^{+/+}$ embryos could be detected in the CHT and in a few cells scattered along the trunk region (Figure 3-12 E). Overall, *runx1P2:Citrine* expressing cells in *runx1*$^{W84X/W84X}$ mutants embryos did not undergo cell death but remained in the DA.

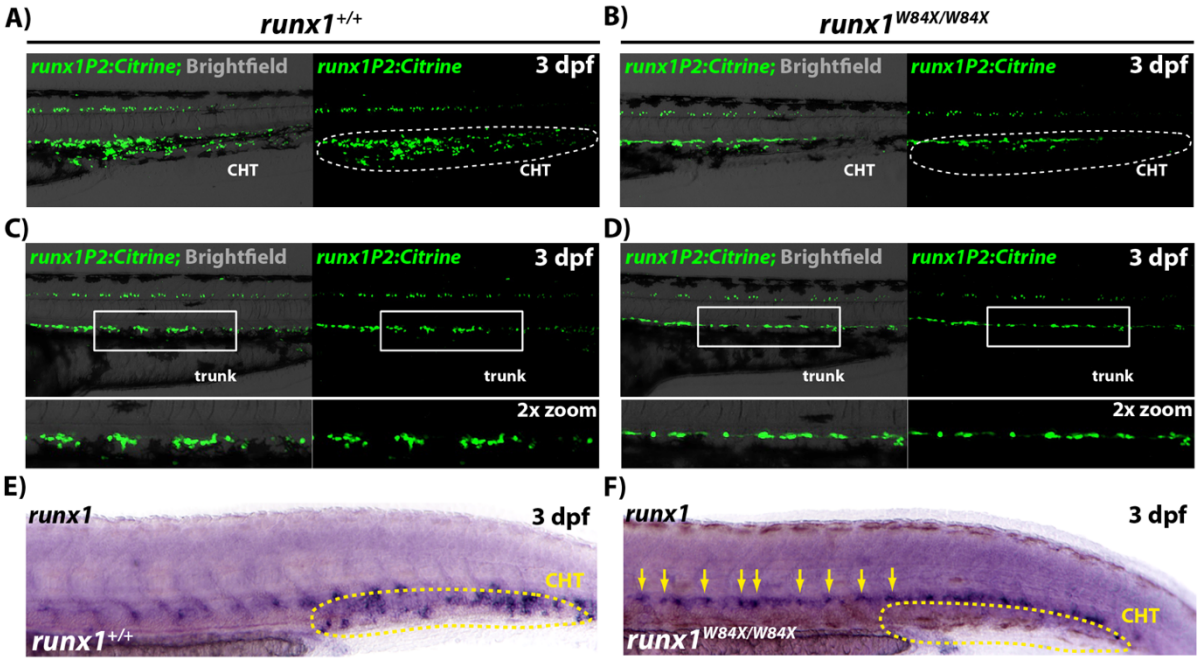


Figure 3-12: Analysis of Citrine fluorescence in *TgBAC(runx1P2:Citrine)* embryos on a *runx1*$^{W84X/W84X}$ mutants background.

A) In *runx1*$^{+/+}$ embryos, many Citrine^{+} cells can be detected in the CHT at 3 dpf. B) In *runx1*$^{W84X/W84X}$ embryos, the number of Citrine^{+} cells in the CHT is significantly decreased. C) In the trunk region of *runx1*$^{+/+}$ embryos, spherical Citrine^{+} cells can be detected in the sub-aortic space at 3 dpf. D) In the trunk region of *runx1*$^{W84X/W84X}$ embryos, Citrine^{+} cells maintain an elongated endothelial cell shape and are associated with the DA. E-F) Expression of endogenous runx1 in *runx1*$^{W84X/W84X}$ embryos can be detected in the region of the DA at 3 dpf but is missing from the CHT.

3.2.4 FACS based isolation of the HE

3.2.4.1 An optimised FACS gating strategy allows for the isolation of the HE and additionally of the DA roof

Fluorescent reporter lines are particularly useful to isolate specific cell types by FACS, especially in the absence of known surface markers or reliable antibodies. Since most genes are expressed in more than one cell type, higher sort purity can often be achieved by combining fluorescent reporter lines with overlapping expression in the cell type of interest. To isolate HE cells, we crossed the newly generated *TgBAC(runx1P2:Citrine)* line with a reporter expressing *mCherry* under the regulatory element of the pan-endothelial marker *kdrl* (*Tg(kdrl:mCherry)*). By doing so, Citrine^{+}mCherry^{+} double-positive HE cells can be separated from Citrine^{+}mCherry^{-} neurons and other blood cell-types (Figure 3-13 A). Simultaneously, the Citrine^{-}mCherry^{+} endothelium can be used as a reference.

Initial flow cytometry experiments showed that Citrine^{+}mCherry^{+} double-positive cells exhibit an unexpectedly broad range of Citrine fluorescence (Figure 3-13 C). We re-analysed Citrine expression under the confocal microscope using laser intensities that over-saturate Citrine fluorescence in the HE (Fig. 13 A, A'). Weak Citrine fluorescence could be detected in cells of the DA roof and the sprouting inter-somitic vessels (ISVs) (Fig 13 A'). No Citrine fluorescence could be detected in cells of the cardinal vein (CV).

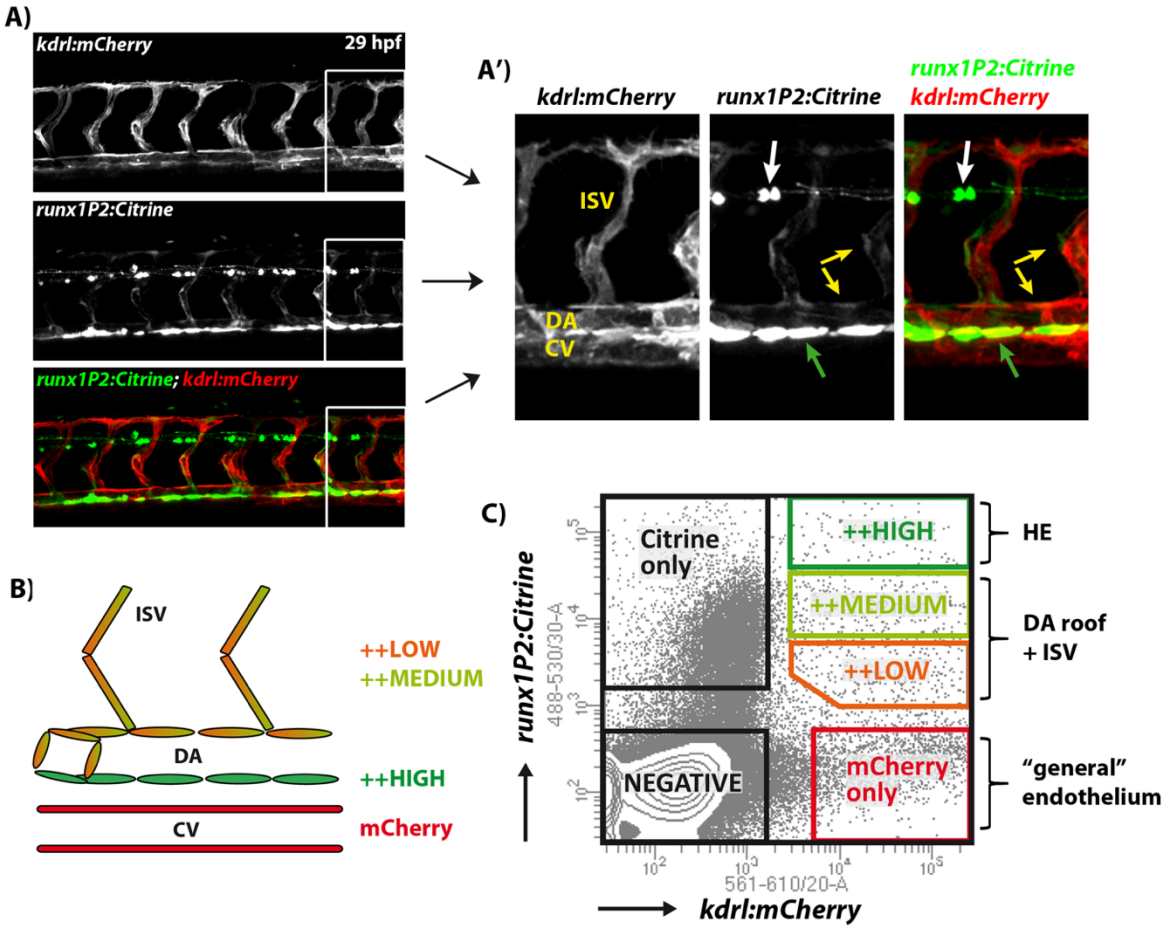


Figure 3-13: Isolation of HE cells from double transgenic *TgBAC(runx1P2:Citrine), Tg(kdrl:mCherry)* embryos via FACS.

A) Confocal image analysis of *runx1P2:Citrine* and *kdrl:mCherry* expression at 29 hpf. A') 2.5x zoom of the boxed area in (A). Laser intensities that over-saturate Citrine fluorescence in the HE (green arrows) reveal a weak Citrine fluorescence in the DA roof and ISVs (yellow arrows). No Citrine fluorescence can be detected in the cardinal vein (CV). Citrine^{+} neurons in the spinal cord are mCherry^{-} (white arrow). B-C) Citrine^{+}mCherry^{+} double-positive (++) cells of the DA can be isolated via FACS from mCherry^{+} single positive "general" endothelium. Based on the level of Citrine fluorescence intensity, ++ cells can be separated into ++HIGH ++MEDIUM and ++LOW populations.

Potential underlying reasons for the presence of Citrine fluorescence in the DA roof/ISV will be addressed further below. Here, we made use of this knowledge to adjust the gating strategy accordingly. We hypothesized that Citrine^{HIGH}mCherry^{+} double-positive cells (++HIGH) represent the HE whereas cells with lower levels of Citrine fluorescence (++MEDIUM, ++LOW) contain cells of the DA roof and the ISVs (Figure 3-13 B-C).

To validate the new gating strategy, qPCR analysis was performed for a selection of marker genes at 28-30 hpf. Here, gene expression within the ++LOW population (presumptive DA roof) was used as a reference to calculate fold changes. The expression levels of *tal1* were highest in the ++MEDIUM and ++HIGH populations (Figure 3-14 A). More strikingly, expression of the HE marker genes *cmyb* (Thompson et al., 1998) and *gfi1aa* (Cooney et al., 2013; Thambyrajah et al., 2016) were specifically up-regulated only in the ++HIGH fraction (Figure 3-14 B, C).Next, the expression of endothelial marker genes was analysed. *Tbx20* is a particularly useful marker, as its expression in the DA is restricted to the roof and the sprouting ISV (Ahn et al., 2000). As expected, *tbx20* had the lowest expression in the ++HIGH population (Figure 3-14 D). *Dll4* and *efnb2a* are widely used arterial markers (Siekmann and Lawson, 2007; Zhong et al., 2001). Notably, *dll4* showed highest expression in the ++LOW population (Figure 3-14 E). The expression of both *efnb2a* and *dll4* was gradually lost with increasing levels of detectable Citrine fluorescence in the other fractions, indicating that cells in the ++HIGH population have already downregulated expression of endothelial genes as they are becoming haemogenic (Figure 3-14 E, F).

Overall, the gene expression analysis showed that cells of the ++HIGH fraction highly express HE marker genes, whereas cells of the ++MEDIUM and ++LOW gates have higher expression of endothelial/arterial and DA roof marker genes. In conclusion, the gating strategy used here not only allows for the isolation of HE cells (++HIGH) but also of cells of the DA roof/ISV (++LOW, ++MEDIUM), separately from the remaining mCherry+ endothelial population (e.g. head vasculature, endocardium, venous cells).

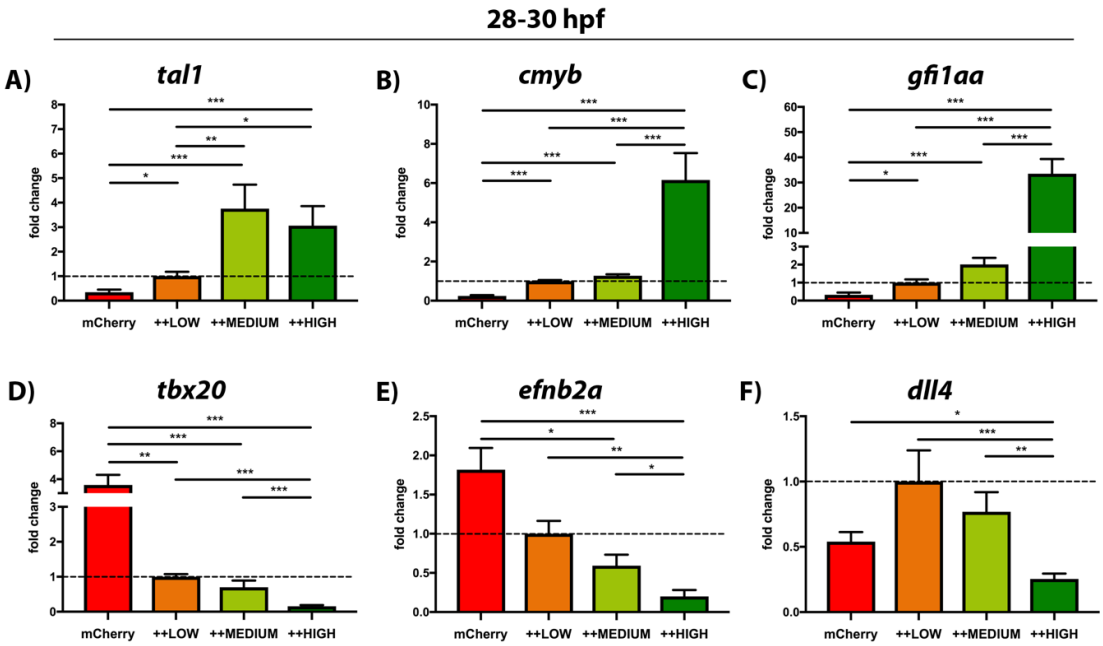


Figure 3-14: qPCR analysis of HE and endothelial marker genes in FACS isolated cell populations of double transgenic *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* embryos.

A-C) Haemogenic marker genes *tal1* (A), *cmyb* (B) and *gfi1aa* (C) were enriched in the ++HIGH fraction. D-F) DA roof marker *tbx20* (D) and arterial marker *efnb2a* (E) and *dll4* (F) were depleted from the ++HIGH fraction. Expression in the ++LOW population was used as a baseline. Graphs show the fold change relative to the ++LOW population (N=4). Statistical analysis (1way ANOVA) was performed between all populations: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean.

3.2.4.2 Quantification of FACS isolated DA and HE cells

The FACS analysis of *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* embryos at 28-30 hpf showed that, as expected, cells of the HE and the DA are only a minor fraction of the whole embryo (Figure 3-15 A). Approximately 0.872 % of the whole embryo are *kdrl:mcherry* expressing endothelial cells containing cells of the HE (++HIGH: 0.047 %) and of the DA roof/ISV (++MEDIUM: 0.094 %; ++LOW: 0.086 %) (Figure 3-15 A). In contrast, Citrine single positive ($Citrine^{+}mCherry^{-}$) cells represent more than 4 % of all cells. It is likely that these cells contain a mixed population of cells including different types of neurons and cells of the earlier waves of blood production.

On average, a single embryo contained a total of 86 ++ cells that split into 33.4 (+/- 12.3) ++LOW cells, 34.4 (+/- 9.4) ++MEDIUM cells and 18.2 (+/- 3.0) ++HIGH cells (Figure 3-15 B) indicating that the HE represents approximately 1/5 (21.2% +- 2.7) of the whole DA (including ISV cell population). A previous analysis of the DA at a similar time point showed that the DA circumference

consists of no more than four cells (not taking the sprouting ISVs into consideration) and that most if not all cells forming the DA floor at 35 hpf eventually undergo EHT (Kissa and Herbomel, 2010). This would imply that the HE represents approximately 1/4 of the DA. Further, ISV sprouts consist of 2 to 4 cells at around 28-30 hpf (Ellertsdóttir et al., 2010) and are usually spaced by ~2 endothelial cells. Taken together, the here detected ratio of 1/5 (++HIGH out of all ++ cells) seems to reflect the actual composition of the endothelial cell populations present in the DA (Figure 3-15 C).

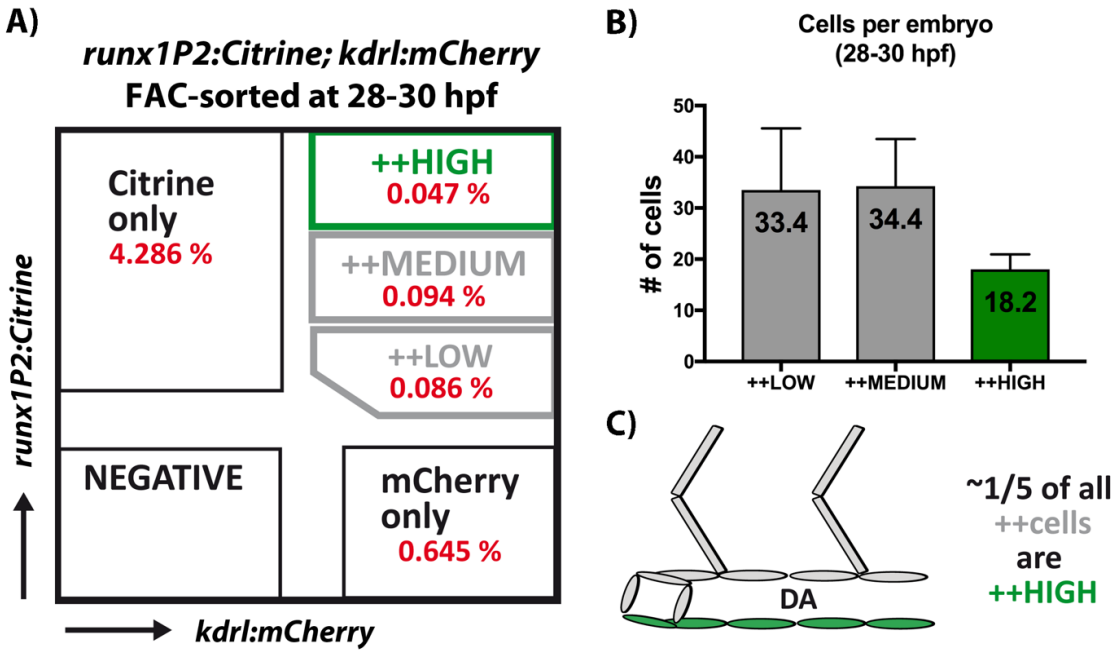


Figure 3-15: Quantification of FACS isolated cell populations of double-transgenic *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos at 28-30 hpf.

A) Numbers represent the % of the respective cell population in relation to the whole embryo. B) Total number of ++LOW, ++MEDIUM and ++HIGH cells that can be detected by FACS per embryo. C) Schematic of the zebrafish DA with sprouting ISVs. Grey cells represent cells from the ++LOW and ++MEDIUM gates combined. Green cells represent cells from the ++HIGH gate.

3.2.5 Analysis of the *runx1P2* reporter expression in the DA roof

3.2.5.1 Citrine fluorescence in the DA roof is caused by residual protein

The detection of Citrine fluorescence in the DA roof and the ISV, even though very weak, was unexpected. The simplest explanations are either that this weak fluorescence reflects the true endogenous expression of *runx1* (that is too weak to be detectable by chromogenic ISH) or that it represents ectopic expression of the *runx1P2:Citrine* transgene. However, analysis of Citrine fluorescence in the DA roof over several time points showed that this fluorescence was diminishing and became undetectable at around 50 hpf (Figure 3-16). This suggested that Citrine fluorescence in the DA roof might be caused by the presence of residual Citrine protein, rather than by weak transgene expression.

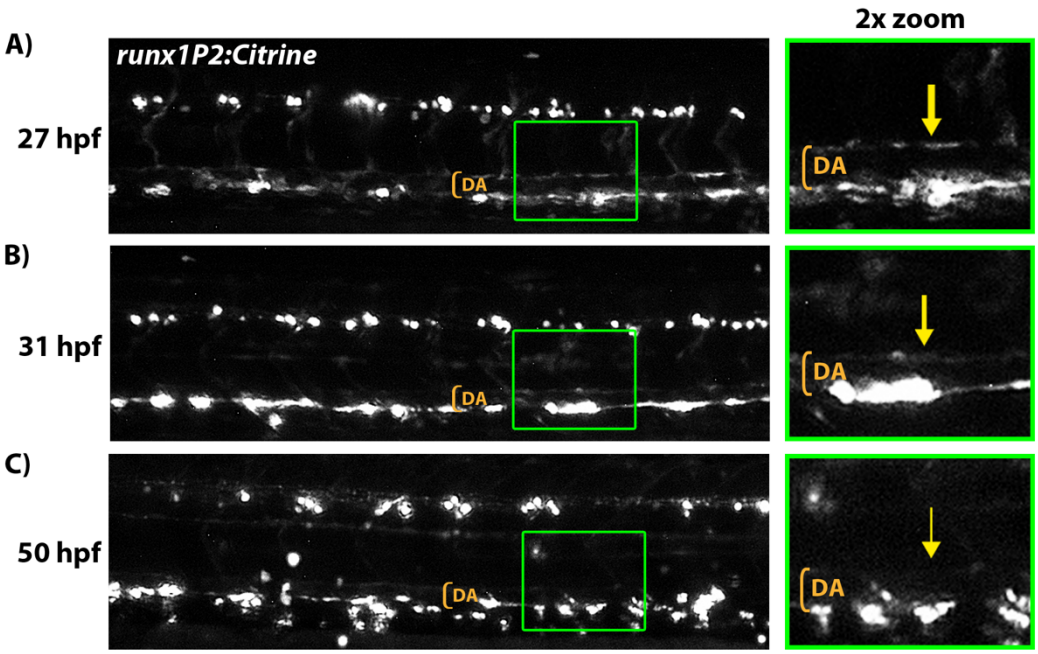


Figure 3-16: Analysis of Citrine fluorescence in the DA roof of *TgBAC(runx1P2:Citrine)* embryos.
Citrine fluorescence in the DA roof can be detected around 27 and 31 hpf (yellow arrow) but was absent in older embryos around 50 hpf.

To clarify this, we performed fluorescent ISH for *Citrine* mRNA in combination with an immunofluorescence staining against the Citrine protein at 25 hpf, when most Citrine$^+$ erythrocytes are in circulation. As detected for the native Citrine fluorescence, antibody staining for Citrine protein

revealed a strong signal in the DA floor and a weak signal in the DA roof (Figure 3-17). In contrast, the fluorescent ISH for *Citrine* mRNA only gave a strong signal in the DA floor, but no signal was detectable in the DA roof (Figure 3-17). These results further indicate that Citrine fluorescence in the DA roof is caused by residual protein from an earlier time point/progenitor population.

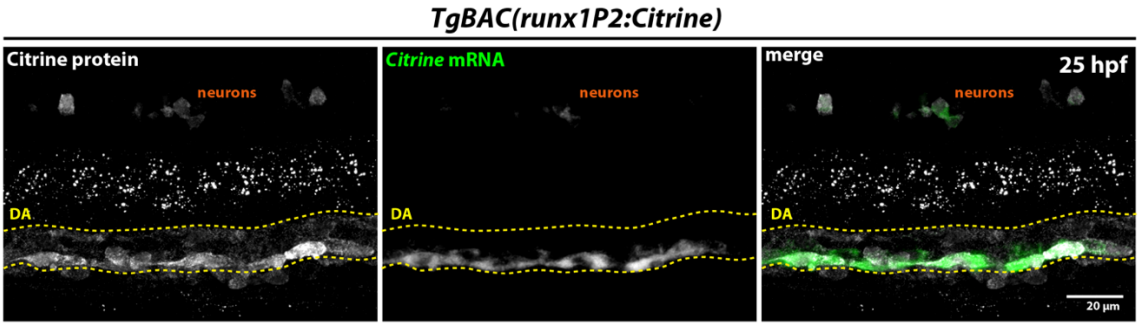


Figure 3-17: Combined immunofluorescence staining against Citrine protein and fluorescent ISH for *Citrine* mRNA. Citrine protein can be detected in the DA floor and to a lower extent in the DA roof. *Citrine* mRNA is only detected in the DA floor. Experiment performed by Monika Krecsmarik.

3.2.5.2 Residual Citrine protein in the DA roof is caused by transgene expression from around 18-22 hpf

To identify the source of the residual Citrine protein detectable in the DA roof, we examined *runx1P2:Citrine* expression prior to 25 hpf. At 18 hpf, when most Citrine$^+$ cells of the PLM have reached the midline and the DA is forming, a tubular structure that formed out of Citrine$^+$ cells was detectable right above the bulk of spherical Citrine$^+$ cells (likely primitive erythrocytes, Figure 3-18 A). It became clear that the tubular structure represents the forming DA, since its shape became more pronounced at 20 hpf and sprouting ISV were detectable at 22 hpf (Figure 3-18 B-C). Interestingly, during these early time points the levels of Citrine fluorescence in the DA floor and the DA roof were relatively comparable.

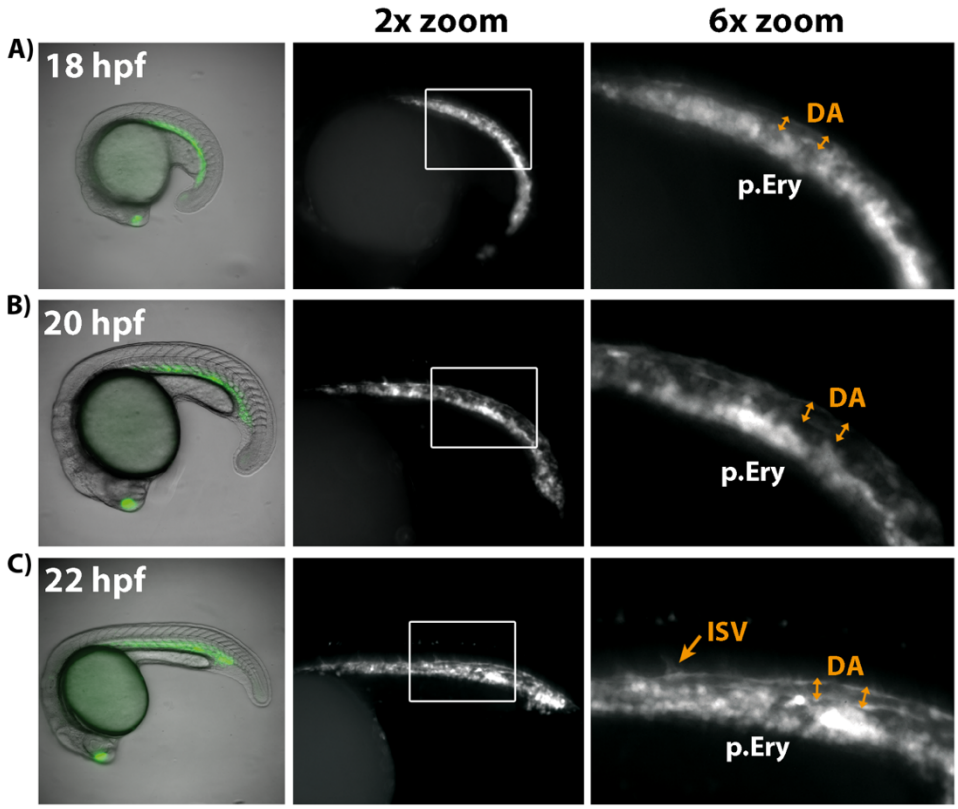


Figure 3-18: Citrine fluorescence analysis in 18-22 hpf *TgBAC(runx1P2:Citrine)* embryos.
A-C) Citrine fluorescence can be detected in primitive erythrocytes (p.Ery) and in cells of the early DA that is undergoing lumenisation.

To determine if the Citrine fluorescence in the DA between 18-22 hpf is also caused by Citrine protein carry-over or by active transgene expression, we performed ISH experiments on *TgBAC(runx1P2:Citrine)* embryos. As expected, *runx1* expression in primitive erythrocytes was down-regulated at around 18 hpf and was not detectable between 20-22 hpf before its re-activation in the HE around 24 hpf (Figure 3-19 A). Similarly, *Citrine* expression in primitive erythrocytes is down-regulated at 18 hpf (yellow arrow) (Figure 3-19 A). However, *Citrine* expression was also detectable right above primitive erythrocytes in the region of the forming DA (orange arrows) around 18 and 20 hpf. From 22 hpf onwards, *Citrine* expression becomes more restricted to the DA floor (red arrows). These data show that a weak transgene expression present in all cells of the DA around 18-20 hpf is the cause of residual Citrine protein that is still detectable in the DA roof at later time points. Since the ISH for *Citrine* was much more sensitive than for *runx1*, it remains a possibility that the detected *Citrine* transgene expression in the DA around 18-20 hpf is not ectopic but that *runx1* expression is too weak to be detectable by chromogenic ISH. Like the protein carry-over in the DA roof, Citrine

fluorescence was also clearly detectable in primitive erythrocytes at 24 hpf (Figure 3-19 B). Furthermore, strong Citrine fluorescence in cells of the HE became apparent at 24-25 hpf (Figure 3-19 C), coincide with endogenous *runx1* expression in the HE.

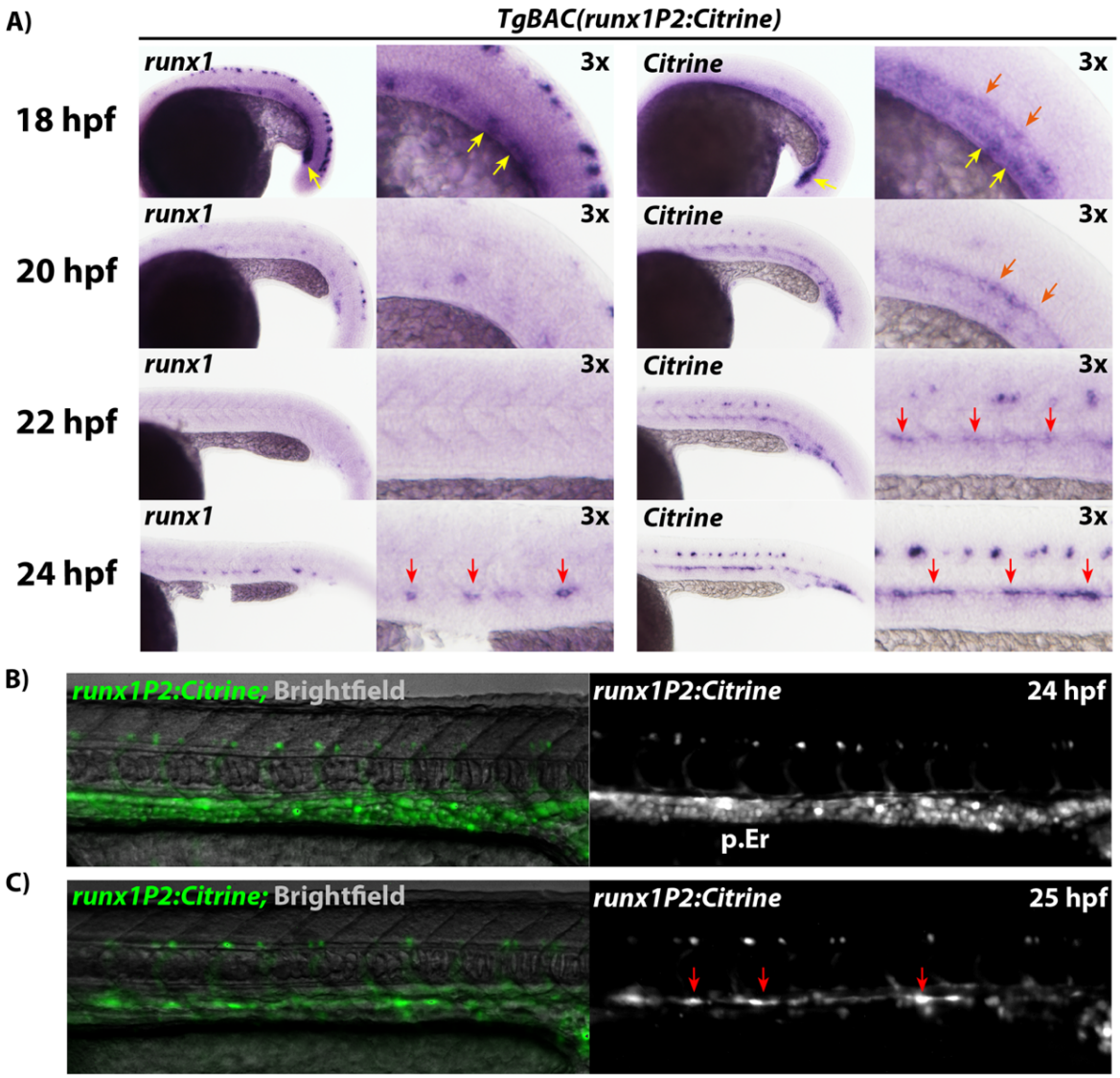


Figure 3-19: *Runx1* and *Citrine* mRNA expression analysis in 18-24 hpf *TgBAC(runx1P2:Citrine)* embryos.
A) Expression of *runx1* in the PLM (yellow arrows) is down-regulated around 18 hpf and is re-activated in cells of the HE around 24 hpf (red arrows). Similarly, *Citrine* expression in the PLM is down-regulated from 18 hpf onwards (yellow arrows) but it is also detectable in the region of the forming DA (orange arrows). *Citrine* expression in the DA-region is also detectable around 20 hpf and becomes more restricted to the DA floor from 22 hpf onwards (red arrows). B) At 24 hpf, primitive erythrocytes (p.Er) are located beneath the DA and show detectable Citrine fluorescence. C) At 25 hpf, Citrine^{+} erythrocytes are in circulation and HE cells with strong Citrine fluorescence are detectable in the DA floor (red arrows).

3.3 Discussion

Here we describe the generation of a novel *runx1* reporter line. To recapitulate endogenous *runx1* expression as faithfully as possible, we designed our reporter based on a BAC transgenesis strategy. We could demonstrat that the newly generated *TgBAC(runx1P2:Citrine)* reporter line recapitulates expression of *runx1* during both primitive and definitive haematopoiesis. In addition, based on the in-depth characterization of transgene expression, we devised a FACS gating strategy that allowed us to establish a protocol to reliably sort cells of the HE and of the DA roof.

3.3.1 Species-specific differences in the P2 N-terminus and the zebrafish *runx1* promoter usage

Runx1 expression is regulated via two alternative promoters, P1 and P2. In human and mouse, the P1-specific coding sequence splices into the P2-specific coding sequence via an in-exon splice site, thereby skipping 16 bp of the P2 isoform (Figure 3-1 and Figure 3-2) (Levanon and Groner, 2004). Interestingly, the corresponding promoter-specific isoforms in zebrafish arise differently: a non-conserved intron separates the P2 isoform-specific 28 bp (exon 2a) from the common region of exon 2 (now called exon 2b) at the same position where the in-exon splice site is found in human and mouse (Figure 3-2 A). However, this species-specific difference does not seem to have any significant relevance for the regulation of *runx1* expression.

The N-terminus of the P2 isoform is not conserved between human/mouse and zebrafish (Figure 3-2 B). The human/mouse sequence contains a conserved 5 AA sequence (MRIPV) that is also present in the RUNX2 and RUNX3 proteins, but absent in zebrafish (Cameron and Neil, 2004). In contrast, the P1-specific N-terminus is highly conserved across all species (Figure 3-2 C). The strong evolutionary constraint on the P1 N-terminus might be caused by its dominance during adult haematopoiesis. Initiation of P1 promoter activity is associated with the peak in HSC expansion in the FL, it defines fully committed definitive hematopoietic progenitors and is used almost exclusively in adult HSCs (Bee et al., 2009; Sroczynska et al., 2009; Telfer and Rothenberg, 2001). In contrast, the P2 promoter shows haematopoietic activity mostly in transient tissues, including the primitive blood

and in the HE during embryogenesis (Bee et al., 2010; Sroczynska et al., 2009). Here, species-specific differences in these processes might be reflected in the different P2 N-termini.

Despite these differences, the activity of the zebrafish *runx1* P2 promoter seems to recapitulate what is known from mouse. P2 activity was detectable during primitive haematopoiesis in the PLM and most importantly in the HE (Figure 3-3), where it drives *runx1* expression that is critical for the emergence of HSCs (Bee et al., 2010; Sroczynska et al., 2009).

3.3.2 The 97a02 BAC covers the complete first *runx1* intron and contains most likely all haematopoietic enhancers

During the course of this study, several other zebrafish *runx1* reporter lines have been generated, all based on the +23 mouse enhancer element (He et al., 2015; Tamplin et al., 2015; Zhang et al., 2015). The +23 element is conserved across most higher vertebrates and integrates the GATA, ETS, and TAL1 transcriptional networks (Nottingham et al., 2007). However, a similar enhancer element for zebrafish has not yet been identified (Lam et al., 2009). Whether the +23 enhancer is conserved in zebrafish remains an unresolved question due to the poor annotation of this genomic locus.

The +23 enhancer drives GFP expression in the HE of the mouse embryo from E8.5 onwards (Bee et al., 2009; Nottingham et al., 2007; Swiers et al., 2013b). Strikingly, it can also drive transgene expression in zebrafish HSPCs suggesting that a similar GATA, ETS and TAL1 transcriptional network governs HSPC generation. However, transgene expression across the various zebrafish +23 reporter lines shows some variability, including the numbers of labelled cells or the timing of transgene activation (He et al., 2015; Tamplin et al., 2015; Zhang et al., 2015). Also, +23 enhancer driven transgene expression in zebrafish seems to be dependent on Runx1 itself (Zhang et al., 2015). Since we wanted to isolate cells of the HE to identify Runx1 targets in HE, we decided to use a BAC transgenesis strategy instead.

We generated two stable *TgBAC(runx1P2:Citrine)* lines from two different founders (line VI and XIII). Both BAC reporter lines show consistent expression in haematopoietic tissues but differ strongly in the level of transgene expression. These differences are most likely caused by positional

effects, although BACs are thought to be generally more resistant to such effects (Giraldo and Montoliu, 2001; Heintz, 2001).

Other differences include expression in neurons of the spinal cord in line XIII (Figure 3-7). We could show that endogenous expression of *runx1* can be detected in these neurons (Figure 3-10 A) indicating that the transgene expression in these cells is not ectopic. However, *Citrine* expression in these neurons is much stronger than endogenous *runx1* expression, suggesting a misregulation in expression strength (Figure 3-10 and Figure 3-11). In addition, the XIII line lacks expression in the Rohan beard neurons (line VI not examined). We concluded that the 97a02 BAC contains most, if not all enhancer elements driving *runx1* expression in haematopoietic tissues but might lack other enhancer elements for proper regulation of transgene expression in neuronal tissues.

3.3.3 Expression kinetics of the *runx1* reporter places upstream signal activity at around 22 hpf or earlier

Our analysis showed that the P2 promoter drives *runx1* expression in both the PLM and the HE (Figure 3-3 C). Characterisation of these two tissues revealed that the novel *TgBAC(runx1P2:Citrine)* line (Figure 3-7 and Figure 3-9) is a reliable reporter for *runx1* expression in haematopoietic tissues. Citrine fluorescence in the PLM is detectable from 12 hpf onwards, the same time when *runx1* becomes detectable by ISH. The same is true for the HE, where Citrine fluorescence becomes prominent from 24 hpf onwards, when *runx1* expression becomes detectable by ISH.

Notably, *Citrine* mRNA was always detectable 1-2 h earlier than *runx1* expression (Figure 3-7 and Figure 3-19). We speculate that this is simply due to a difference in transcription rates. Not much is known about the general rate of transcription in zebrafish. For eukaryotes, transcription rates, as measured by several techniques and in different conditions, vary between 18–72 nucleotides per second (Swinburne and Silver, 2008). A recent *in vivo* approach reported an average polymerase velocity of only 6.3 nucleotides per second (including Pol-II pausing) (Darzacq et al., 2007). Moreover, the predicted transcription rate might also be affected by lower temperatures, as zebrafish embryos are raised at around 29°C. A 717 bp long gene like *Citrine* will be transcribed in a couple of minutes (Darzacq et al., 2007). However, a gene like *runx1,* containing large intronic sequences, will

need a significantly longer time for transcription (Swinburne and Silver, 2008). The *runx1P2* driven transcript is minimal ~26,000 bp in length, including introns (a more detailed estimation is not possible due to the incomplete annotation). Based on the measured *in vivo* transcription rate, this would account for at least 1.5 hours.

Endogenous *runx1* expression in the HE is detectable at 23/24 hpf by ISH. Up-regulated Citrine expression in the HE was detectable already at 22 hpf (Figure 3-19 A; red arrows). Since *Citrine* expression is also controlled by the endogenous zebrafish *runx1P2* promoter on the 97a02 BAC, this implies that the upstream signals that drive *runx1* expression in the HE are active at least around ~22 hpf. In agreement with this, expression of *gata2b*, an important upstream regulator of *runx1*, can be detected in the forming DA from 18-20 hpf onwards (Butko et al., 2015).

3.3.4 *Runx1P2* promoter activity is independent of Runx1

Data about the regulation of *runx1* expression upon Runx1 loss of function is limited. Of note, the *runx1* P1 promoter sequence contains two conserved Runx1 binding sites in zebrafish and higher vertebrates (Levanon and Groner, 2004), suggesting an auto-regulatory loop. The P2 promoter sequence however lacks Runx1 binding sites. In one mouse study, *Runx1* expression was visualized by LacZ staining in $Runx1^{lz/rd}$ mutant embryos in which exon 7 and 8 of one allele were replaced with the *LacZ* coding sequence (*lz*) and exon 5 of the other allele was lacking (*rd*) (North et al., 1999). Here, *Runx1* expression in the vitelline arteries and in the DA ventral wall of mutants and control embryos at E9.5 was comparable. However, at E10 the endothelial *Runx1* expression in both the vitelline arteries and the DA was strongly reduced in the mutants compared to control embryos. Interestingly, 3 dpf $runx1^{W84X/W84X}$ mutant zebrafish embryos showed strong *runx1* expression still detectable in the region of the DA floor (Figure 3-12). Moreover, the *runx1P2:Citrine* reporter was active in $runx1^{W84X/W84X}$ mutants, with $Citrine^{+}$ cells exhibiting an elongated cell shape still found in the DA floor, whereas in Wt embryos only spherical $Citrine^{+}$ cells were found in the sub-aortic space (Figure 3-12). How can this discrepancy be explained? In mouse, it was shown that *Runx1* expression in the HE is initially driven via the P2 promoter, but also that the P1 activity takes over shortly afterwards (Bee et al., 2010; Sroczynska et al., 2009). Since in $Runx1^{lz/rd}$ mutant embryos the *LacZ*

sequence is present in both promoter products, a weakening expression could indeed indicate that P1 activity is regulated by RUNX1 itself. In zebrafish however, HE expression of *runx1* is solely driven by the P2 promoter (Figure 3-3). Also at 48 hpf, mainly the P2 promoter product is detectable (Figure 3-3). It is not known when or if *runx1* expression in zebrafish HSPCs switches to the P1 promoter. Since we can still detect *runx1* expression in the DA region of *runx1*$^{W84X/W84X}$ mutant embryos at 3 dpf we speculate that it is solely P2 derived. This is supported by the detection of Citrine protein under the control of the *runx1* P2 promoter (Figure 3-12).

It is reported that in *runx1* morphants less cells of the HE initiated EHT and if so, these cells burst shortly after emergence (Kissa and Herbomel, 2010). However, on the *runx1*$^{W84X/W84X}$ mutant background Citrine^{+} cells were still detectable in the DA ventral wall at 3 dpf forming a continuous line, indicating that these cells did not undergo apoptosis. We speculate, that in r*unx1* morphants with a "sub-optimal" knock-down of *runx1,* certain differentiation processes, including EHT, might still be initiated but the full haematopoietic program is not. Consequently, some HE cells will initiate EHT but finally undergo apoptosis.

Furthermore, strong Citrine fluorescence was still detectable in DA ventral wall cells of *runx1*$^{W84X/W84X}$ embryos older then 5 dpf (see Chapter 5 Figure 5-21 B). This raises the question if upstream signals that drive *runx1* expression are still present or if the *runx1* locus is epigenetically locked in an active state (Allis and Jenuwein, 2016). A recent study has shown that the P1 promoter region gets hypo-methylated upon transition to definitive haematopoiesis and acquires active chromatin modifications (Webber et al., 2013). However, the P2 promoter contains a classic CpG island that was un-methylated in all cell types examined (Webber et al., 2013). This question may be resolved once more detailed data about the epigenetic regulation of the *Runx1* locus is obtained.

3.3.5 *Runx1P2* reporter activity in the early DA suggests a close relationship between HE and aortic endothelium

The detection of weak Citrine fluorescence in the DA roof of the *TgBAC(runx1P2:Citrine)* reporter line around 24-36 hpf was unexpected. At first glance, this feature resembled the ectopic expression seen in the *Tg(runx1P2:EGFP)* promoter line (Lam et al., 2009; Lam et al., 2010). However, a detailed

analysis of *Citrine* expression has shown otherwise. First, the Citrine fluorescence in the DA roof was extremely weak and restricted to that tissue; *Citrine* expression in further vasculature could not be detected. Second, Citrine fluorescence in the DA roof was only transient and caused by protein carry-over from the 18-20 hpf time point, when the DA is forming (Figure 3-16 and Figure 3-19).

In zebrafish, the DA precursor cells (angioblasts) are found in the PLM intermingled with or near precursor cells of the CV and primitive erythrocytes (Davidson and Zon, 2004; Zhong et al., 2001). The DA angioblast lineage is already separated from CV precursor cells at around 13–15 hpf (Zhong et al., 2001) and are the first to start migrating inwards at ~14 hpf. They reach the midline at around 16-18 hpf, shortly before the prospective primitive erythrocytes and CV precursors come to lie ventral to the DA. The DA cord starts to lumenize at ~18 hpf (Jin, 2005). This is also the time point when *Citrine* transgene expression and Citrine fluorescence became detectable in the forming DA (Figure 3-19). We were not able to clarify if this transgene expression is ectopic or if it might reflect the endogenous situation in which *runx1* expression is too weak to be detectable by chromogenic ISH. In either case, these observations do suggest a close relationship between the arterial and haemogenic lineages for the following reasons: Citrine fluorescence could only be detected in arterial cells of the DA, but not in the CV (Figure 3-13 A). Furthermore, *Citrine* expression levels were comparable in all DA cells at around 18-20 hpf (Figure 3-18). An up-regulation of *runx1/Citrine* expression and a true commitment to the haemogenic lineage would then be partly dependent on the positioning of the arterial cell within the DA scaffold and by specific local signals that these cells receive. For example, localized BMP signalling produced by the sub-aortic mesenchyme is required for *runx1* expression in the HE (Pouget et al., 2014; Wilkinson et al., 2009). Increased BMP signalling, accompanied by a reduction of somite derived BMP antagonists causes *runx1* expression to extend dorsally within the DA (Pouget et al., 2014).

A suggested favourable model (Figure 3-20) is one in which DA precursor cells that have just reached the midline are all exposed to the same signals that initiate *Citrine* expression (and presumably *runx1*) (Burns et al., 2005; Gering and Patient, 2005; Wilkinson et al., 2009). As the DA undergoes lumenization, cells in the DA roof move away from these signals (which are now restricted to the ventral wall of the DA) and down-regulate *Citrine* expression, while cells in the DA-floor

proceed to up-regulate *Citrine*. To answer the remaining questions, it is planned to perform fluorescent ISH experiments for *runx1* at these time points (18-20 hpf) which might be more sensitive than chromogenic ISH.

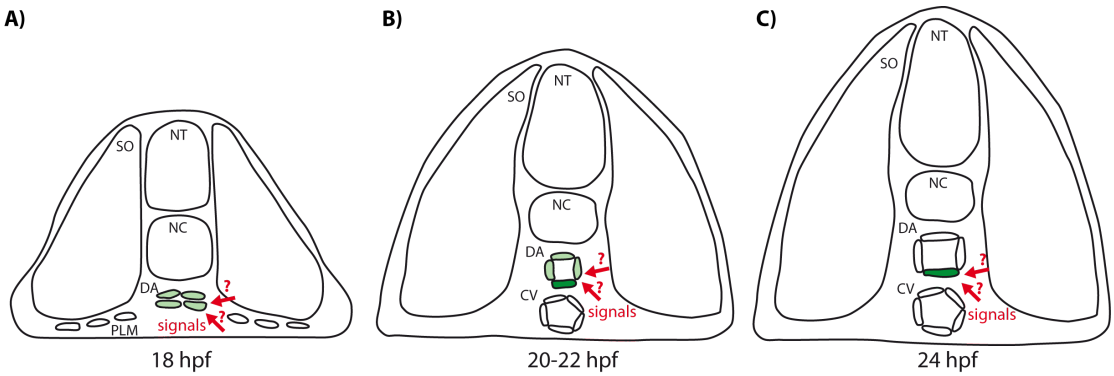


Figure 3-20: Suggested model about the initiation of *Citrine* expression in *TgBAC(runx1P2:Citrine)* embryos.
A) At around 16-18 hpf DA precursor cells have reached the midline and might already be exposed to external signals (red arrows) that initiate a weak Citrine expression (light green). B) From 18 hpf onwards the DA starts to lumenize and DA roof cells depart from ventral signalling cues around 20-22 hpf. In contrast, DA floor cells are still exposed to signals promoting haematopoiesis and continue to up-regulate *Citrine* expression (dark green). C) From 24 hpf onwards, strong *Citrine* expression is maintained in the HE.

3.3.6 The ++HIGH fraction contains purified HE

Since *runx1* expression is detectable in a plethora of cell types, including many different neurons and blood cells, we crossed the *TgBAC(runx1P2:Citrine)* reporter to the endothelial *Tg(kdrl:mCherry)* reporter line to achieve specificity for the HE. By doing so, we could isolate double positive cells with a strong Citrine fluorescence as early as 24-25 hpf (data not shown). The knowledge about the weak Citrine fluorescence in the DA roof/ISVs prompted us to distinguish these ++HIGH cells from groups of double positive cells with weaker Citrine fluorescence (++LOW and ++MEDIUM).

The presumptive HE cells (++HIGH) were very rare and on average only 18.2 (+- 3.0) cells per embryo could be detected (Figure 3-15 B). It is thought that the HE region extends over 12 somites in length, with ~2 endothelial cells bridging one somite (Murayama et al., 2006). Furthermore, most if not all cells forming the DA floor eventually undergo EHT (Kissa and Herbomel, 2010) which would account for ~24 HE cells. Thus, ~18.2 ++HIGH cells/embryo might be a slight under-representation. However, other sorting strategies that attempted to count HE cells by using *cmyb:eGFP;lmo2:dsRed*

transgenic embryos also identified around 18-20 double-positive cells per embryo (Harris et al., 2013). We analysed a single time point (28-30 hpf) only, and it could very well be that other HE cells still haven't upregulated *runx1*/*Citrine* expression yet.

Under stringent sorting conditions, only ~10 ++HIGH cells could be isolated per embryo, so we focused our initial molecular analysis on the expression of a few key marker genes. The high expression levels of the haematopoietic markers *tal1*, *cmyb* and *gfi1aa* in the ++HIGH population strengthened our assumption that these cells are indeed HE cells. Conversely, cells of the ++LOW population showed higher expression of *tbx20* (DA roof marker) and of the arterial gene *dll4*, suggesting that these cells are indeed derived from the DA roof/ISVs. Cells of the ++MEDIUM gate had a similar low expression of *cmyb* and *gfi1aa* when compared to the ++LOW population, suggesting that the gating strategy was well chosen.

3.3.7 How the refined gating strategy will influence further analysis.

To investigate the HE specific gene expression it was intended to compare the HE to an endothelial reference population. With the here established gating protocol it was possible to isolate the DA endothelium out of the whole population of "general" endothelium that itself contains a mix of cells of the arterial and venous lineages. Further, this endothelial population contains cells from a variety of different cellular locations. Indeed, it was shown that the head-vasculature is derived from the ALM while the DA is PLM derived (Zhang and Rodaway, 2007). Despite all being endothelial cells, these different endothelial sub-groups might have different molecular properties due to their respective origins and specifications.

Otherwise, in all vertebrate model organisms the HE is always embedded within the major arteries, mainly the DA, suggesting a close relationship between these two cell types (Ciau-Uitz et al., 2013; Gering and Patient, 2005; North et al., 1999; North et al., 2002; Richard et al., 2013). With this established protocol, a comparison of the gene expression of the zebrafish HE to its closest related tissue, namely the cells of the DA roof can now be done. We assume that this will significantly increase the specificity of gene expression differences and thus benefit all further analysis.

3.4 Conclusion and directions

Here, I presented the generation of a novel *TgBAC(runx1P2:Citrine)* reporter line that reliably recapitulates endogenous *runx1* expression in haematopoietic tissues. HE cells are marked by a strong Citrine fluorescence as early as 24 hpf. An unexpected weak Citrine fluorescence in the DA roof and sprouting ISVs was shown to be caused by Citrine protein carry-over from its earlier transgene expression in all DA-cells around 18-20 hpf. Incorporating this knowledge into our FACS gating strategy allowed us to specifically isolate cells of the HE as well as cells of the DA roof. Although this has not been anticipated, this has turned out to be a significant advantage. Originally it was planned to compare gene expression of HE cells with a general endothelial reference population. However, the strategy we developed allowed us to compare the HE directly to arterial cells of the DA, which is thought to be its closest related tissue and thus provides further enrichment compared to our original strategy.

- Chapter 4 -

Genome wide expression analysis of the HE and identification of potential Runx1 target genes

4.1 Introduction

The identification and isolation of defined haematopoietic progenitor sub-types during embryogenesis is laborious and often results in low cell numbers. Therefore, early attempts to study global gene expression changes during developmental haematopoiesis or to identify targets of key transcription factors (TFs) were mainly based on immortalised blood progenitor cell lines or on *in vitro* differentiation culture protocols (Lancrin et al., 2012; Lichtinger et al., 2012; Lie-A-Ling et al., 2014; Tanaka et al., 2012; Wilson et al., 2009; Wilson et al., 2010a). To a certain degree, these cultures resemble embryonic haematopoiesis as it is shown by the expression of key marker genes, including *Runx1*, and by the generation of various blood cell types in clonogenic assays (Irion et al., 2010; Lancrin et al., 2012; Murry and Keller, 2008). Notable progress has been made in identifying conditions that lead to the generation of blood cells with lymphoid potential for both human and mouse ESC/iPSC cultures (Irion et al., 2010; Kennedy et al., 2012). The potential to produce cells of the lymphoid lineage is found only during the production of definitive but not primitive blood (Böiers et al., 2013; Ditadi et al., 2015; Irion et al., 2010; Kennedy et al., 2012; Lin et al., 2014). However, in the embryo, lymphoid potential was also identified for the 2nd wave of haematopoiesis occurring in the YS (Böiers et al., 2013; Chen et al., 2011; Luc et al., 2012; Yokota et al., 2006; Yoshimoto et al., 2011; Yoshimoto et al., 2012). It became evident that most of the currently used culture protocols mainly resemble YS haematopoiesis (Goode et al., 2016).

The latest improvements in the field of next generation sequencing led to a significant reduction in the amount of required material. Consequently, researchers were now able to screen for novel factors involved in embryonic haematopoiesis using *in vivo* tissue (Kartalaei et al., 2015; Li et al., 2014). Among others, these efforts identified the importance of the inflammatory signalling pathway during both, mouse and zebrafish haematopoiesis (He et al., 2015; Li et al., 2014). However, so far, no functional genome wide *in vivo* study has been performed with the aim to identify targets of the key transcription factor RUNX1 during the transition from HE to HSPCs.

The previous chapter (Chapter 3) described the generation and characterisation of a novel *TgBAC(runx1P2:Citrine)* reporter line that labels cells of the HE. By using an optimised FACS gating strategy on *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* double-transgenic embryos we sorted for

Citrine^{LOW}*mCherry*$^{+}$ (++LOW) and *Citrine*^{HIGH}*mCherry*$^{+}$ (++HIGH) cells that were considered to represent cells of the DA roof and the HE, respectively. With this strategy in hand, cells of the whole DA can now not only be specifically isolated out of the whole endothelial population, but also further sub-divided into haemogenic and non-haemogenic cells. To identify novel targets of Runx1 within the HE, we aim to perform deep sequencing on the here described endothelial cell populations.

4.1.1 Chapter aims

There are three main aims in this chapter:

1. To perform next generation RNA sequencing of the HE and other endothelial populations.
2. To identify HE-specific and Runx1-regulated genes.
3. To identify relevant Runx1 target genes with an evolutionarily conserved expression.

First, I will present the detailed experimental design for the generation of the required RNAseq data-sets to study the zebrafish HE in both control and loss of *runx1* conditions. Second, the genome-wide gene expression data were analysed to characterise the sorted cell populations in greater depth. Third and fourth, I aimed to determine the set of HE-specific genes and to determine those that are potentially under the regulation of Runx1 respectively. This will be guided by cross-comparisons to related RNAseq and ChIPseq data-sets derived from mouse or human haematopoietic cell types. Last, a screening strategy is presented to identify zebrafish HE-specific Runx1 target genes that are evolutionarily conserved in their expression in haematopoietic progenitor cell types.

The bioinformatic analysis of the RNAseq data was performed in collaboration with Rossella Rispoli.

4.2 Results

4.2.1 Experimental design for the RNAseq analysis of the zebrafish HE

4.2.1.1 Using a *runx1* MO for the functional analysis of HE

RNAseq analysis of FACS isolated endothelial cells aimed to identify HE-specific genes. In order to perform functional experiments against *runx1,* a well-established splice morpholino (MO) was used (Gering and Patient, 2005) that binds to the 3' end of exon *2a* specific for the *P2* promoter product (Figure 4-1 A-B) and causes the loss of blood progenitors (Gering and Patient, 2005; Kissa and Herbomel, 2010). Indeed, injecting this MO into our *TgBAC(runx1P2:Citrine)* reporter line resulted in a significant loss of Citrine^{+} cells in the CHT at 56 hpf (Figure 4-1 C), phenocopying the *runx1*$^{W84X/W84X}$ mutant (Chapter 3 Figure 3-12). Importantly no detectable difference in Citrine fluorescence could be observed in the HE of MO injected and control embryos at around 35 hpf (Figure 4-1 C), thus allowing for the FACS based isolation of the HE under both conditions.

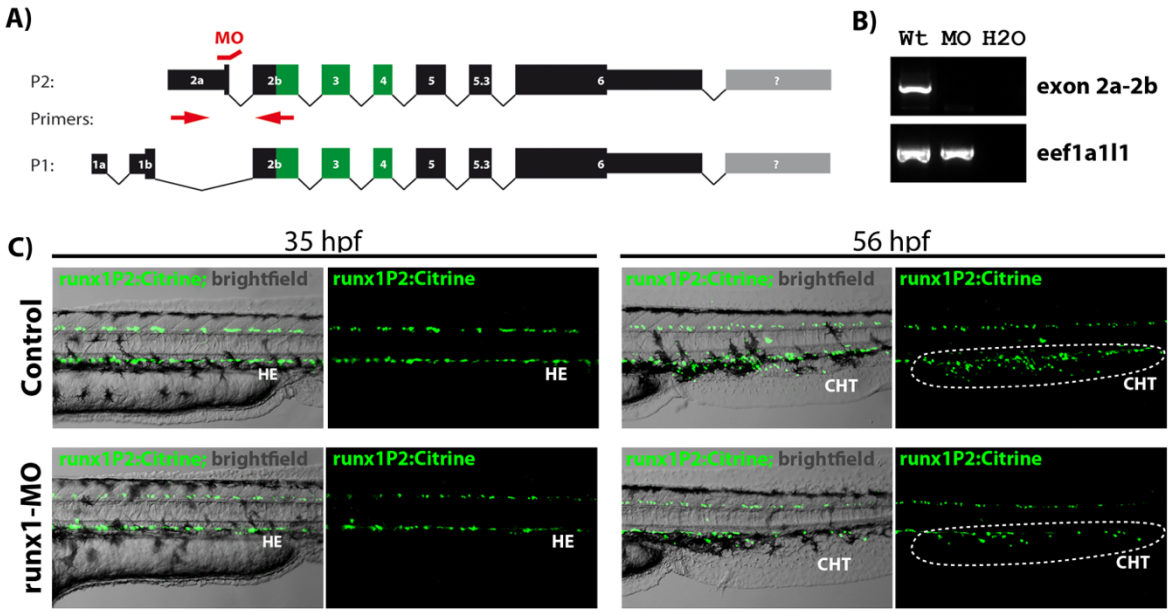


Figure 4-1: Validation of a *runx1*-MO for functional studies of the HE by RNAseq.
A) Schematic of the P1 and P2 specific *runx1* transcripts. Binding sites of the splice-MO against the P2 derived transcript and PCR validation primers are indicated. B) The splice blocking effect of the runx1 MO was demonstrated by PCR. C) In *runx1* morphants, significantly reduced numbers of Citrine^{+} cells can be detected in the CHT around 56 hpf. Citrine expression is not affected in the HE around 35 hpf.

4.2.1.2 Detailed experimental design for the generation of the RNAseq samples

The expression of *runx1* in the HE is detectable by ISH from around 23/24 hpf onwards (Kalev-Zylinska et al., 2002) while the very first EHT events are thought to be initiated at around 32 hpf (Kissa and Herbomel, 2010). In the following period of ~24 hr, additional cells of the HE undergo EHT, whereby the number of events peak at around 40 to 52 hpf (Kissa and Herbomel, 2010). Based on the timing of these events, we reasoned that the optimal time point for the RNAseq analysis of the HE is at around 28-30 hpf, shortly before EHT is initiated (Figure 4-2); changes in gene expression evoked by Runx1 might not be pronounced enough in younger embryos, while the analysis of later time points (e.g. ~36 hpf) might result in the identification of primarily secondary or tertiary Runx1 targets.

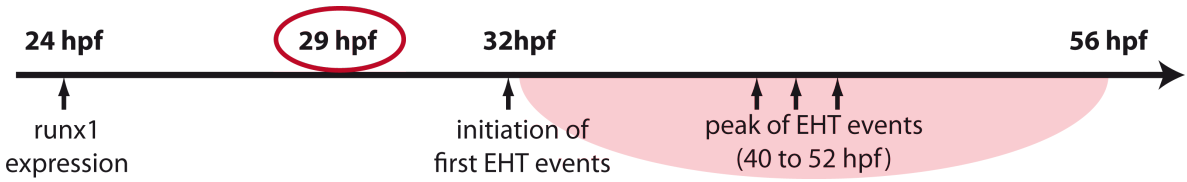


Figure 4-2: Timeline of the most important events during definitive haematopoiesis in the zebrafish HE. *Runx1* expression in the HE becomes detectable around 24 hpf. The first EHT events are thought to be initiated around 32 hpf (Kissa and Herbomel, 2010). EHT events can be observed until 56 hpf with the peak of events around 40-52 hpf). The role of Runx1 in the HE will be studied at around 29 hpf (red circle).

To isolate sufficient amounts of RNA for the next generation sequencing analysis, at least 3000 cells per population were collected. This consistently resulted in ~7 ng of high quality RNA (RIN > 8.0). Our FACS sorting conditions allowed us to isolate ~10 ++HIGH cells per embryo. Hence, at least 300 *TgBAC(runx1P2:Citrine);Tg(flk1:mCherry)* double-transgenic embryos for both the *runx1* knockdown and the control condition were processed (Figure 4-3). To collect sufficient cell numbers, multiple clutches from different parents were combined. To control for potential batch effects, every single clutch was split in half; while one half of the clutch remained unmanipulated, the other half was injected with the *runx1* MO. Afterwards, injected and non-injected embryos from different clutches were pooled accordingly and staged to 28-30 hpf for subsequent FACS sorting. Cells were isolated from the mCherry^{+}, ++LOW, ++MEDIUM, ++HIGH and NEGATIVE gates.

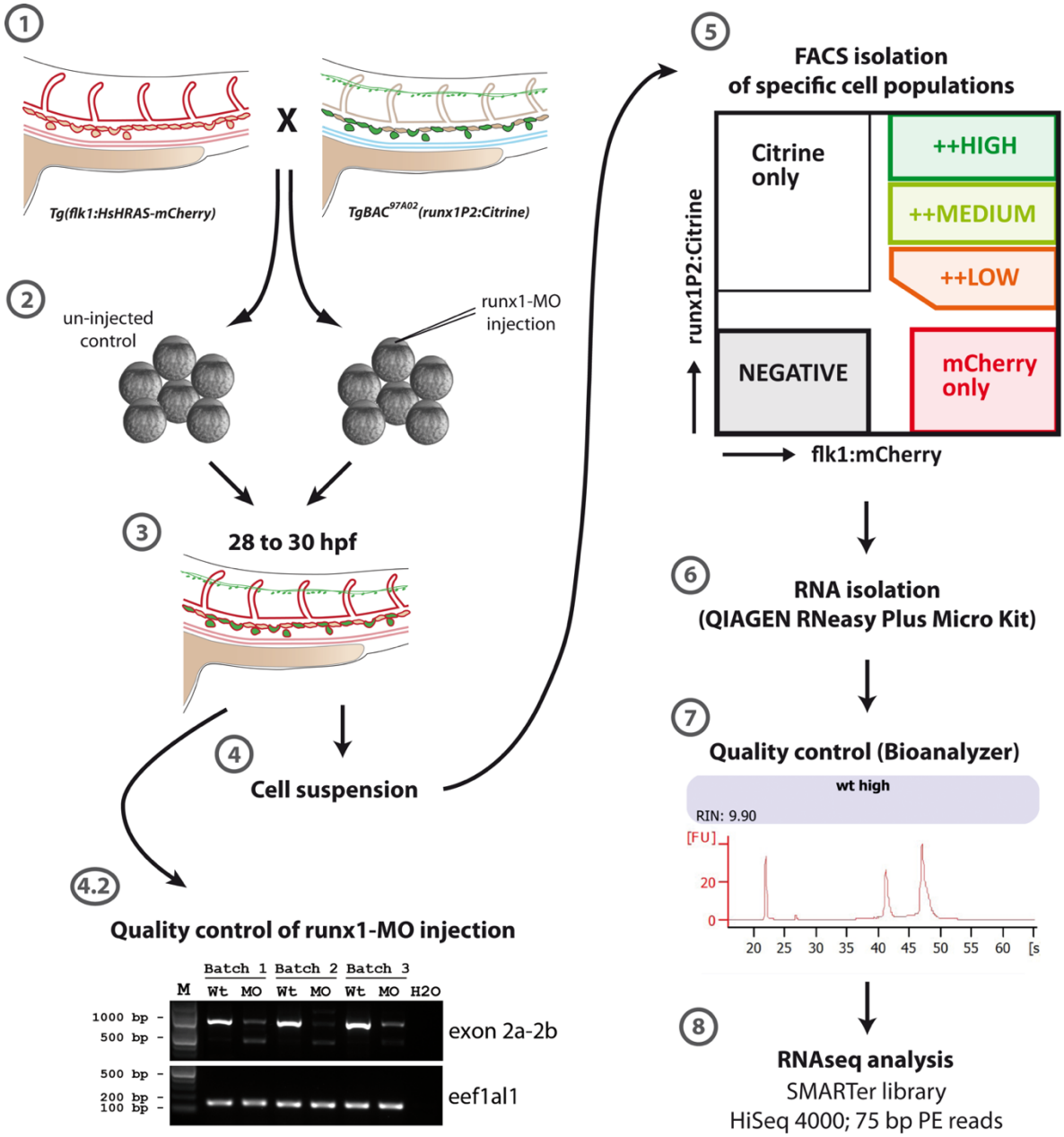


Figure 4-3: Experimental strategy to study the zebrafish HE by RNAseq.
1: The endothelial reporter line *Tg(kdrl:mCherry)* was crossed with the runx1 reporter line *TgBAC(runx1P2:Citrine)*. 2: Clutches of embryos from each pair of parents was split in half. Only one half were injected with a runx1-MO. 3: Double transgenic embryos were separated and raised to 28-30 hpf. 4: Embryos were prepared for subsequent FACS sorting. 4.2: Remaining embryos were used to validate the functionality of the runx1-MO. 5: Cell suspensions of double transgenic embryos were FACS sorted and cells of the NEGATIVE, mCherry^{+}, ++LOW, ++MEDIUM and ++HIGH gates were isolated. 6: RNA was isolated from sorted cells. 7: RNA quality was determined. 8) RNA was send to the Wellcome Trust Centre for Human Genetics (WTCHG) for sequencing.

RNA of FACS sorted cells was isolated with the RNeasy Plus Micro Kit (QIAGEN, Cat No. 74034). With this Kit, we obtained the highest RNA quantities and a consistently high RNA quality compared to the RNeasy Micro Kit (QIAGEN, Cat No. 74004) and the RNAqueous-Micro Total RNA Isolation Kit (Ambion, Cat No. AM1931) (Figure 4-4).

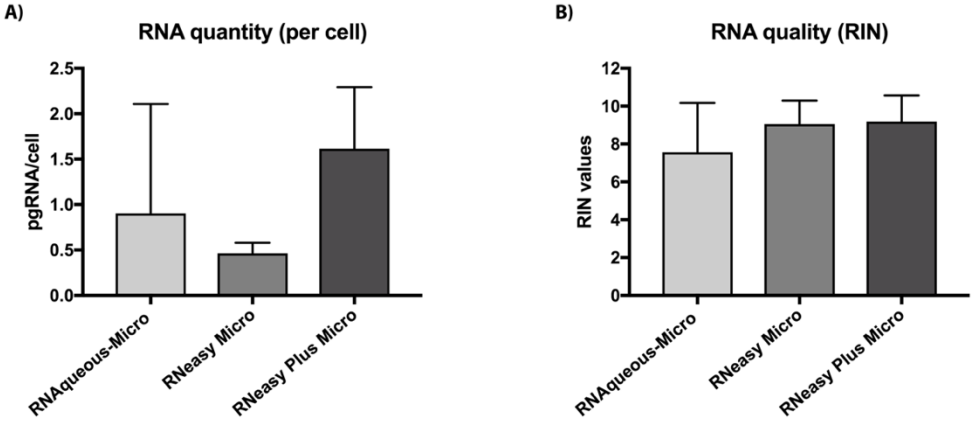


Figure 4-4: Comparison of three kits for the isolation of RNA from cells after FACS.
A) The Ambion RNAqueous-Micro Total RNA isolation Kit showed a great variation in the amount of isolated RNA. The RNeasy Micro kit from QIAGEN resulted in only ~0.5 pg RNA/cell. The RNeasy Plus Micro kit allowed us to isolate the highest amount of RNA (~1.5 pg RNA/cell). B) All tested kits isolated high quality RNA. Both QIAGEN kits generated more consistent results. For the RNAqueous Kit, 5 out of 31 isolations did not result in a RIN value. Error bars represent the standard deviation.

To ensure correct MO injection, additional embryos were used to test for the MO dependent splice-blocking effect (Figure 4-3). N=4 experiments have been performed. Out of those, three were evaluated as being of high enough quality (RNA quality and MO functionality) for subsequent RNAseq analysis. RNA was sent to the Wellcome Trust Centre for Human Genetics (WTCHG). SMARTer libraries for low-input RNA were sequenced on an Illumina HiSeq4000 machine with a 75 bp paired end protocol. Around 24-30 million reads per sample have been generated to ensure for enough sequencing depth. Next generation sequencing analysis was performed using the EdgeR package (Robinson et al., 2010). EdgeR was chosen to stay consistent with our collaborators (Marella de Bruijn laboratory) who analysed global gene expression changes in the murine HE and subsequent blood progenitors also using EdgeR.

4.2.2 Clustering analysis confirms the haemogenic and aortic identity of the ++HIGH and ++LOW populations

First, we used the global gene expression data for a comprehensive comparison of the four isolated endothelial populations (mCherry+, ++LOW, ++MEDIUM and ++HIGH) to describe the respective cell identities in more detail (Figure 4-5 A). The multidimensional scaling plot (MDS plot) highlighted that for each cell population, the three biological replicates clustered together, proving the consistency of the performed protocols (Figure 4-5 B). The MDS plot further suggests inherent differences within the genetic program of the isolated cell fractions, whereby ++LOW and ++MEDIUM cells are more similar to each other and separate from the mCherry^{+} and ++HIGH cells on the Leading logFC dim1. ANOVA analysis across the four endothelial cell populations identified **5574** differentially expressed genes (DEG) (FDR < 0.05). For a subsequent clustering analysis, **361** genes with a log2-FC < +/-0.7 were excluded. Consensus clustering of the remaining 5213 genes identified an optimal setting of 8 clusters from cumulative distribution function (CDF) plots (Figure 4-5). Clusters 1, 2, 4, and 5 consisted of a larger number of genes (Figure 4-6). Cluster 3 only contained 14 genes. Cluster 6, 7, and 8 were relatively small and none of them contained more than 120 genes (Figure 4-6).

Next, using the Database for Annotation, Visualization and Integrated Discovery (DAVID) software (Huang et al., 2008b) gene ontology (GO) analysis was performed to identify overrepresented biological functions in the individual clusters (Table 4-1). Cluster 1, the largest cluster, consisted of 2570 genes with higher expression in the mCherry^{+} and ++LOW fractions (Figure 4-6). It generated the largest number of enriched GO terms (Table 4-1), including “regulation of transcription, DNA templated” and “cell differentiation”. Other prominent GO terms were “cell adhesion”, “sprouting angiogenesis”, “Wnt signalling pathway”, and “Notch signalling pathway”. Many endothelial marker genes including *kdrl*, *efnb2* (*a* and *b*), *ephb4* (*a* and *b*), *tbx20*, *sox7, dab2, nr2f2 (COUP-TFII),* and *flt4* were present in this cluster. Overall, GO term-based analysis of Cluster 1 validated the endothelial character of the ++LOW cell population.

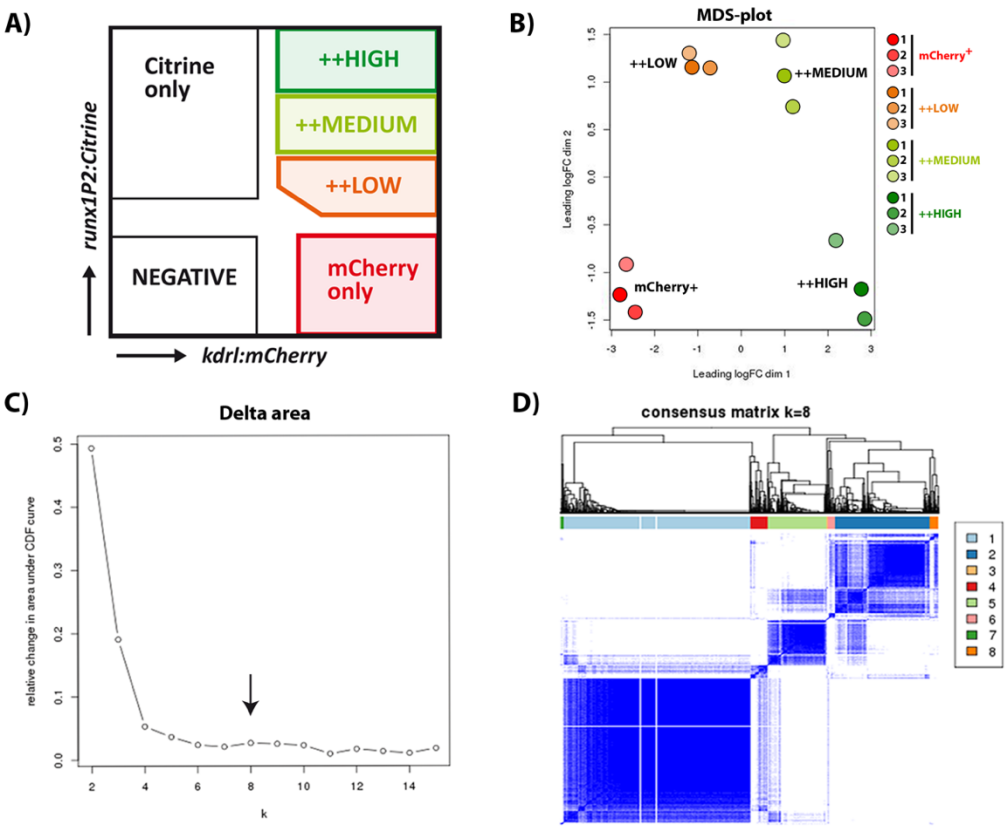


Figure 4-5: Consensus clustering of RNAseq data from Citrine$^{+/-}$mCherry^{+} endothelial populations.
A) Schematic highlighting the populations (mCherry^{+}, ++LOW, ++MEDIUM, and ++HIGH) that were used for consensus clustering analysis. B) Multidimensional scaling plot (MDS-plot). C) Dealta areal plot of consensus clustering identified an optimal setting of k=8 from cumulative distribution function. D) Consensus clustering matrix for k=8. Each row represents one gene. The intensity in blue represents the frequency at which two genes were observed to cluster together.

The second largest cluster with 1305 genes was Cluster 2 (Figure 4-6). It represented genes with higher expression across all three double-positive populations (++LOW/++MEDIUM/++HIGH). Together with the microscopic analysis in Chapter 3, this large common genetic programme further indicates that the combined fraction of ++ cells represent the DA as a whole. Cluster 2 was enriched for GO terms including "erythrocyte differentiation", "embryonic hemopoiesis" and "heme biosynthetic process" (Table 4-1). With respect to the *kdrl*$^{+}$ identity of the analysed cells, this finding highlights a close lineage relationship between the endothelial and the haematopoietic lineages during development (Adamo and García-Cardeña, 2012). Other GO terms like "translation" indicate that these tissues might be undergoing active cellular changes. Furthermore, it was enriched for the GO term "Immune system process", but with a FDR value > 0.05 (-log10 P-value = 2.6). A role of inflammatory signalling for definitive haematopoiesis has been shown before by several laboratories

(Espín-Palazón et al., 2014; He et al., 2015; Li et al., 2014; Sawamiphak et al., 2014). Examples of genes in Cluster 2 are *gfi1aa*, *tal1*, *lmo2*, *cmyb*, *ikzf1* (*ikaros*), *gpr183b*, and *gata2b*. These genes have all been associated either with the DA or the HE (Butko et al., 2015; Gering and Patient, 2005; Patterson et al., 2007; Thambyrajah et al., 2016; Zhang et al., 2015).

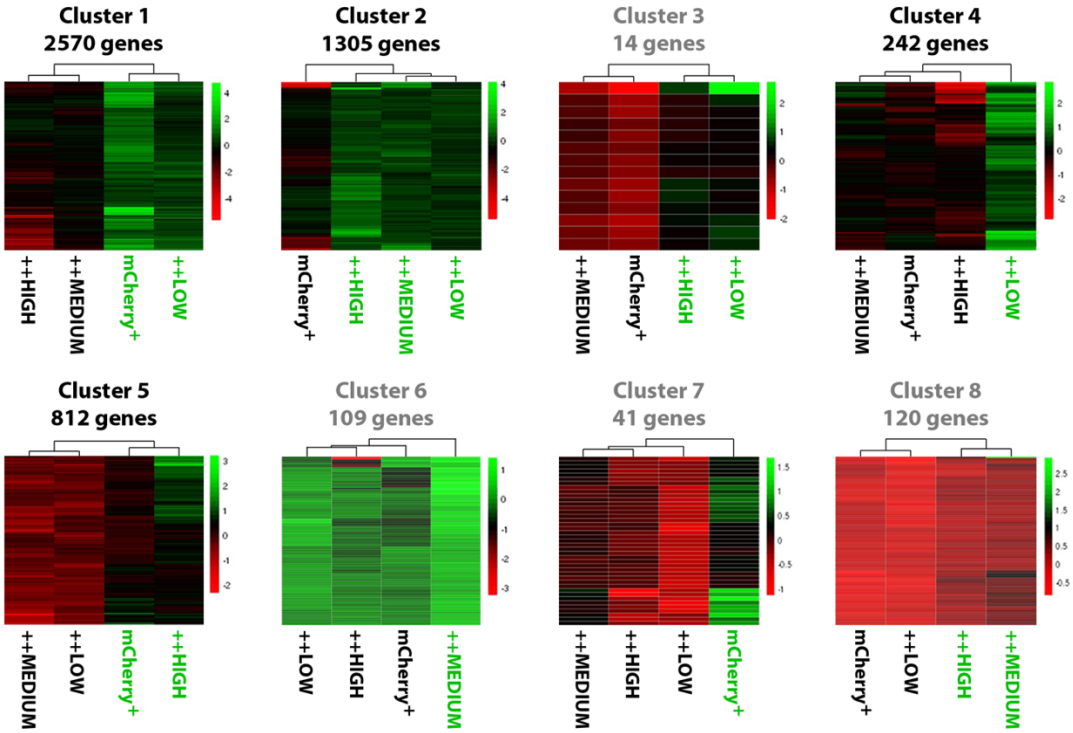


Figure 4-6: Heatmaps of the 8 gene clusters resulting from consensus clustering as shown in Figure 4-5. Shown are Z-score normalised gene expression values. The numbers of genes for each cluster are indicated. The four largest clusters are shown in black. The four smallest clusters are shown in grey. Green name: cell populations for which the genes have a higher expression in the respective clusters.

Cluster 4 contained 242 genes with higher expression in the ++LOW population (Figure 4-6) and generated a low number of enriched GO terms (Table 4-1). The only GO term with an FDR < 0.05 was "regulation of transcription, DNA-templated". In addition, the GO term "Notch signalling pathway" was enriched, but with an FDR > 0.05 (-log10 p-value = 2.5). Importantly, Cluster 4 contained arterial marker genes, including *dll4* and *hey2*, indicating an arterial identity of the ++LOW population.

Cluster 5 contained 812 genes with a higher expression mainly in the ++HIGH population, but also some genes with higher expression in the mCherry^{+} population (Figure 4-6). Enriched GO terms such

as “rRNA processing”, “ribosome biogenesis”, “methylation”, and “DNA replication” (Table 4-1) suggest that these cells are in an active metabolic state, since associated genes are involved in core cellular processes. The *dnmt3bb.1* gene, which was previously shown to be HE-specific and regulated by Runx1, was part of Cluster 5 (Gore et al., 2016).

Cluster 6 and 8 contained genes specific for the ++MEDIUM and ++MEDIUM/++HIGH populations, respectively (Figure 4-6). Genes in Cluster 6 were not informative for the identification of enriched GO terms. Genes in Cluster 8 were enriched for the GO terms “cell cycle” and “cell division” (Table 4-1). It has been shown that HE cells that perform EHT might also undergo cell division shortly after budding (Kissa and Herbomel, 2010). Furthermore, a previous integrative analysis of RUNX1 downstream pathways identified “cell cycle” as a significantly enriched GO term and showed that RUNX1 promotes cell cycle progression from G1 to S phase (Michaud et al., 2008; Strom et al., 2000).

Cluster 7 contained 41 genes with higher expression in the mCherry^{+} endothelium (Figure 4-6) and was enriched for the GO term “cell cycle” (Table 4-1). This was not unexpected since the endothelium is still undergoing angiogenesis.

In summary, this clustering analysis strengthened our hypothesis that cells of the ++LOW fraction represent a purified fraction of the DA roof with a strong endothelial profile but also enriched for arterial marker genes, whereas the ++HIGH gate contains a highly-purified fraction of the HE.

Table 4-1: GO term analysis among the genes of cluster 1 to 8 from Figure 4-6.
Representatives of the most meaningful GO terms are shown. All presented GO terms have a p-value < 0.05. GO terms with a FDR value < 0.05 are highlighted in bold. Genes of Cluster 3 and 6 did not result in the detection of enriched GO terms.

Cluster 1	**-log10 P-value**	**Cluster 2**	**-log10 P-value**
Regulation of transcription, DNA-templated	**35.3**	**Erythrocyte differentiation**	**9.5**
		Embryonic hemopoiesis	**7.1**
Cell differentiation	**14.8**	**Heme biosynthetic process**	**6.0**
Cell adhesion	**10.3**	**Translation**	**5.2**
Sprouting angiogenesis	**9.7**	**Hemopoiesis**	**5.1**
Wnt signalling pathway	**8.5**	**Oxygen transport**	**4.9**
Notch signalling pathway	**8.2**	**Erythocyte development**	**3.8**
Angiogenesis	**8.1**	**Translational initiation**	**3.5**
Blood vessel development	**7.4**	Immune system process	2.6
Cluster 3	**-log10 P-value**	**Cluster 4**	**-log10 P-value**
/	/	**Regulation of transcription, DNA-templated**	**6.0**
		Notch signalling pathway	2.5
Cluster 5	**-log10 P-value**	**Cluster 6**	**-log10 P-value**
rRNA processing	**41.6**	/	/
Ribosome biogenesis	**29.8**		
Methylation	**21.4**		
tRNA processing	**19.9**		
RNA processing	**12.9**		
Cellular response to DNA damage stimulus	**8.4**		
DNA replication	**5.9**		
Translation	**4.8**		
Cluster 7	**-log10 P-value**	**Cluster 8**	**-log10 P-value**
Cell cycle	**5.8**	**Cell cycle**	**8.5**
		Cell division	**6.6**
		Phosphorylation	1.4

4.2.3 Identification of HE-specific genes

The clustering analysis verified the identity of the analysed cell populations. However, known HE marker genes (e.g. *cmyb, gata2b, gfi1aa* and *dnmt3bb.1*) were in different clusters (Cluster 2 and 5). A closer assessment of such marker genes revealed that whereas expression of the genes *cmyb*, *gata2b* and *gfi1aa* was lowest in the mCherry^{+} population (Figure 4-7 A-C), expression of the *dnmt3bb.1* gene was not (Figure 4-7 D). This presumably prompted the algorithm to place such genes in different clusters. Consequently, identification of haemogenic genes was not straight forward using this clustering approach.

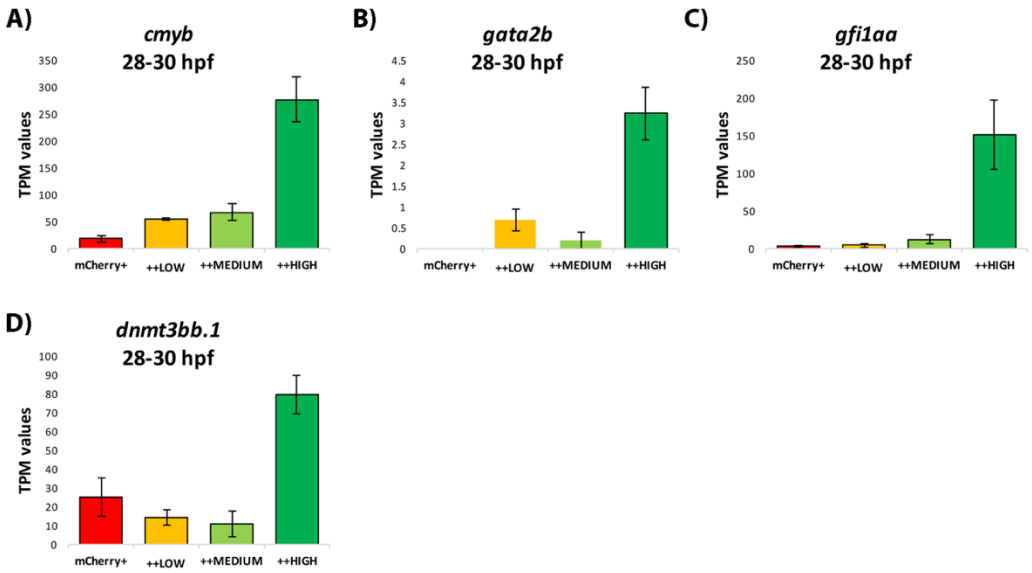


Figure 4-7: Detailed assessment of HE marker gene expression as detected by RNAseq
Graphs show the respective TPM values of the four HE marker genes *cmyb* (A), *gata2b* (B), *gfi1aa* (C) and *dnmt3bb.1* (D) detected within the indicated cell populations sorted from *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* embryos. Error bars represent standard deviations.

4.2.3.1 DEG analysis between ++LOW and ++HIGH populations identifies known and novel HE-specific genes

To identify novel HE-specific genes, we argued that a direct comparison of the HE population with its closest related tissue, the DA roof, would be the most informative (Figure 4-8 A). Using this approach, 2731 genes were found to be differentially expressed between the ++HIGH and the ++LOW populations (Figure 4-8 B). Of those genes, 1099 were up-regulated in the ++HIGH population (HE-

specific) and 1633 were down-regulated (DA-specific). Of the 1099 HE-specific genes, 131 genes had log2-FC > 1.7 (Appendix 5), including most of the known HE markers in zebrafish (*cmyb*, *gata2b*, *gfi1aa* and *dnmt3bb.1*), indicating that we were indeed enriching for haemogenic genes.

The HE-specific set of genes was enriched for GO terms associated with cellular and metabolic processes including "rRNA processing", "ribosome biogenesis", "methylation", and "translations" (Figure 4-9 C). Next, the DAVID software (Huang et al., 2008b) was used to determine enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The most enriched KEGG pathway for HE-specific genes was "ribosome biogenesis in eukaryotes" (Figure 4-8 E). Genes associated with such GO and KEGG terms are involved in core cellular processes, suggesting that the HE is in an active metabolic state. Since mouse RUNX1-deficient HSPCs were shown to have a slow biosynthetic activity (Cai et al., 2011), RUNX1 might influence similar processes also in the HE.

The GO term enrichment analysis for the DA-specific set of genes included terms associated with "angiogenesis", "vascular development", "vasculogenesis" and "patterning of blood vessels" (Figure 4-8 D). Such terms coincides with the arterial identity of the ++LOW cells but also indicates that associated genes are already down-regulated within the HE. The top enriched KEGG pathway was "Notch signalling pathway" (Figure 4-8 F), which is in good agreement with recent publications showing that Notch signalling is repressed in the HE before the onset of EHT (Richard et al., 2013; Zhang et al., 2015).

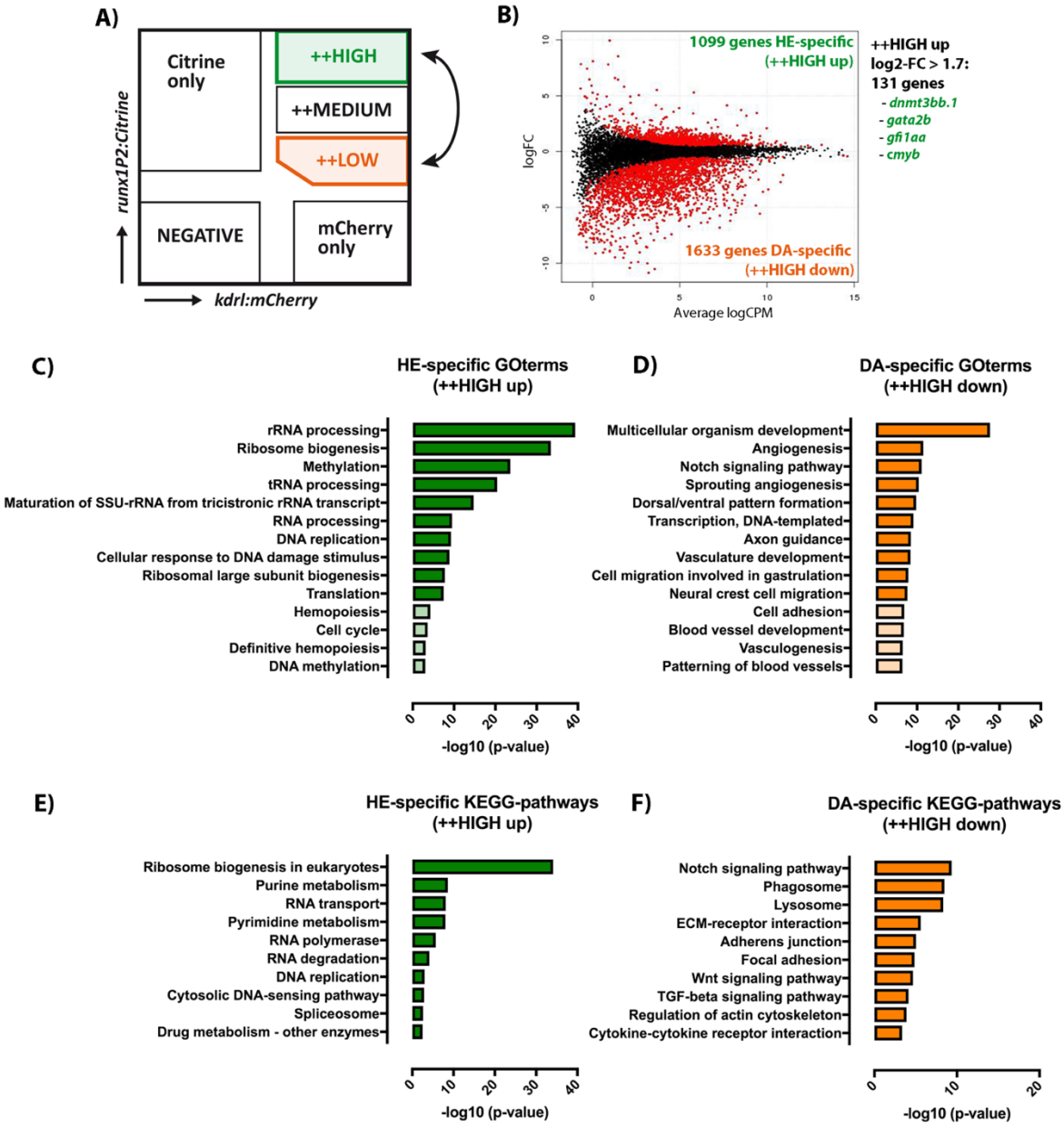


Figure 4-8: Identification of zebrafish HE-specific genes.

A) Schematic highlighting the populations (++LOW and ++HIGH) that were used for the DEG analysis. B) The smear plot shows significantly differentially expressed genes (red) between the ++HIGH and the ++LOW population. C) Top 10 GO terms (dark green) and 4 additional terms (light green) enriched for the HE-specific set of genes. D) Top 10 GO terms (dark orange) and 4 additional terms (light orange) enriched for the DA-specific set of genes. E) Top 10 KEGG-pathways enriched for the HE-specific set of genes. F) Top 10 KEGG-pathways enriched for the DA-specific set of genes

4.2.3.2 Zebrafish HE-specific genes contain a core set of conserved haematopoietic genes

To verify that the 1099 HE-specific genes identified in zebrafish contained genes with an evolutionarily conserved expression during haematopoiesis, we cross-compared this list to two previously published mouse data-sets.

The first study by Kartalaei et al. (2015) made use of *Ly6a-GFP* transgenic embryos and sorted for $CD31^{+}KIT^{+/-}GFP^{+/-}$ cells from the AGM region of E10.5 embryos and distinguished endothelial cells (EC), haemogenic endothelial cells (HECs) and cells they phenotypically identified as HSCs (Kartalaei et al., 2015). Overall 530 DEGs were found, 179 genes of which were associated with HECs and 138 genes associated with HSCs. First, we intersected the zebrafish HE-specific gene list with the combined set of 317 mouse HEC plus HSC genes. This resulted in a small but significant overlap of 33 genes (p-value = 3.4e-5) (Figure 4-9 A). Strikingly, 27 of these genes were found within the HSC associated list (Figure 4-9 B), containing TFs including *gfi1aa*, *gfi1ab*, *gfi1b*, *cmyb,* and *myc,* for which mouse homologous genes have known roles in HSC development (Cooney et al., 2013; Thambyrajah et al., 2016; Wilson et al., 2004; Zhang et al., 2011). We also identified the *angpt1* and *syk* genes which have known roles in HSCs and T-cell differentiation (Sebzda et al., 2006; Zhou et al., 2015). In addition, all zebrafish duplicated versions of the mouse *Dnmt3a* and *Dnmt3b* genes were found within this list (*dnmt3aa*, *dnmt3ab*, *dnmt3ba*, *dnmt3bb.1*, *dnmt3bb.2* and *dnmt3bb.3*).

The second study by Li et al. (2014) sorted for $CD31^{+}VEC^{+}ESAM^{+}KIT^{+/-}$ cells from the AGM regions and the umbilical and vitelline arteries from E10.5 embryos (Li et al., 2014). They distinguished ECs from intra-aortic hematopoietic cluster cells (HCCs) and separated them further based on the expression of a *Ly6a-GFP* reporter. Eight clusters of genes were identified, four of which were associated with cells of the HE or developing HSPCs: Ly6a-GFP^{+} HCCs (C1), Ly6a-GFP^{+} and Ly6a-GFP^{-} HCCs (C2), $Ly6a^{+}$ ECs (C3) and $Ly6a^{+}$ HCCs and ECs (C4). Overall, these four clusters contained together 4498 genes.

To proceed with our comparison, we first intersected our HE-specific gene list with the combined genes of C1-4 which resulted in an overlap of 241 genes (P-value = 0.047) (Figure 4-9 C). We then performed the analysis with different combinations of clusters. Intriguingly, no significant overlap

could be identified for either C3 or C4. On the other hand, combining C1 and C2 resulted in an overlap of 162 genes with a much higher significance (P-value = 1.5e-9) (Figure 4-9 D). Of these, 78 genes were found in C1 and 84 genes were found in C2. Both, C1 and C2 contained only HCCs, hence, the most differentiated haematopoietic sub-types within this analysis. This intersected gene list of 162 genes contained several TFs including *gata2b*, *gfi1aa*, *gfi1ab*, *gfi1b*, *cmyb, stat5a* and *tal1,* for which mouse homologous genes have known roles in haematopoiesis (Butko et al., 2015; Cooney et al., 2013; Patterson et al., 2007; Thambyrajah et al., 2016; Wang et al., 2009b; Zhang et al., 2011). Furthermore, additional genes with known roles in HSCs and in blood differentiation were identified, including *irf1b*, *mpl*, *syk*, *pik3cd* and its regulatory partner *pik3r1* (Abdollahi et al., 1991; Fragale et al., 2008; Kim et al., 2007; Sebzda et al., 2006; Solar et al., 1998). In contrast to the study performed by Kartalaei et al. (2015), only the zebrafish versions of the mouse *Dnmt3b* gene but not of the *Dnmt3a* gene were identified here.

Overall, the here identified list of 1099 HE-specific genes contained several haematopoietic factors with an evolutionarily conserved expression during murine definitive haematopoiesis. Interestingly, we identified the up-regulation of such genes already during the HE stage in zebrafish.

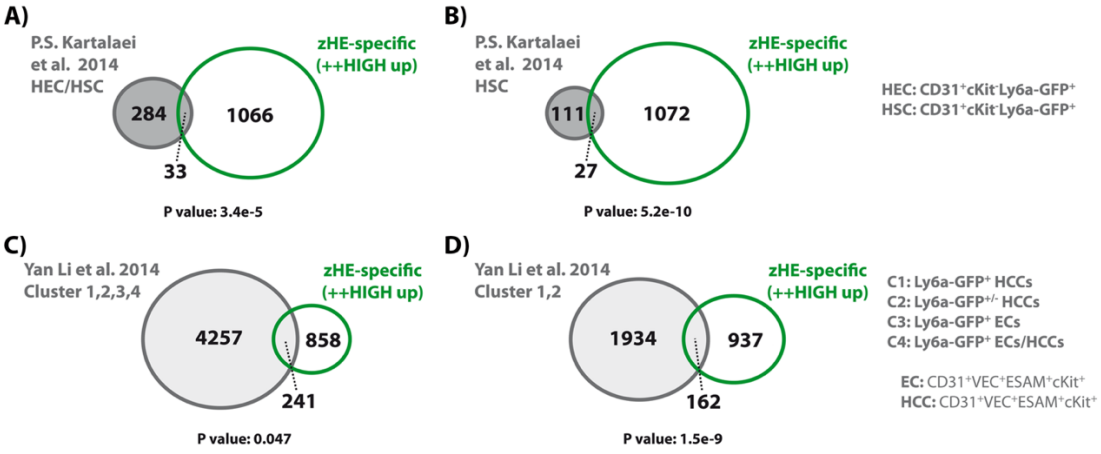


Figure 4-9: Identification of HE-specific genes with an evolutionarily conserved expression during murine haematopoiesis.

Venn diagrams show the overlap of zebrafish HE-specific genes (zHE; green) with DEGs identified by RNAseq of mouse embryo-derived AGM cells. A-B) Overlap of genes with highest expression in HEC/HSCs (A) or HSCs (B) as identified by Kartalaei et al (2015). A detailed description of cell types can be found on the right. C-D) Overlap of genes with highest expression in cluster 1,2,3,4 (C) or cluster 1,2 (D) as identified by Li et al. (2014). A detailed description of cluster cell types can be found on the right. EC: endothelial cell; HCC; haematopoietic cluster cell; HEC: haematopoietic endothelial cell; HSC haematopoietic stem cell.

4.2.4 Identification of Runx1-regulated genes

4.2.4.1 DEG analysis between *runx1* morphants and control embryos identifies novel Runx1 targets

As a next step, we aimed to identify Runx1-regulated genes within the HE. Therefore, we focused on the ++HIGH populations from control and *runx1* MO injected embryos. The NEGATIVE populations were included in this analysis as a reference tissue (Figure 4-10 A).

The MDS plot showed that the samples of the ++HIGH control population clustered away from the samples of the ++HIGH *runx1* MO population (Figure 4-10 B). In contrast, all samples of the NEGATIVE populations clustered tightly together, irrespective of whether they were derived from the control or the *runx1* MO conditions (Figure 4-10 B).

The DEG analysis between the ++HIGH populations of the control and *runx1* MO samples identified 676 genes (Figure 4-10 C). Of these, 255 genes were down-regulated in the *runx1* MO condition (Runx1-activated) and 421 genes were up-regulated (Runx1-repressed).

To correct for potential morpholino off-target effects, we also performed a DEG analysis between the NEGATIVE populations of the control and *runx1* MO conditions (Figure 4-10 D). Cells of the NEGATIVE population do not express *Citrine* and hence should also not express the MO target *runx1*. Overall, 176 were identified as being differentially expressed in the NEGATIVE population; 57 genes were down-regulated in the *runx1-MO* condition and 119 genes up-regulated. We considered only those genes as HE-specific off-targets, whose up- and down-regulation correlated positively within the DEGs identified in the ++HIGH population (Figure 4-10 E). That resulted in a total of 42 potential off-targets (Table 4-2). After subtraction of the identified off-target genes, we ended up with 245 Runx1-activated (see top 100 in Appendix 6) and 389 Runx1-repressed (see top 100 in Appendix 7) genes.

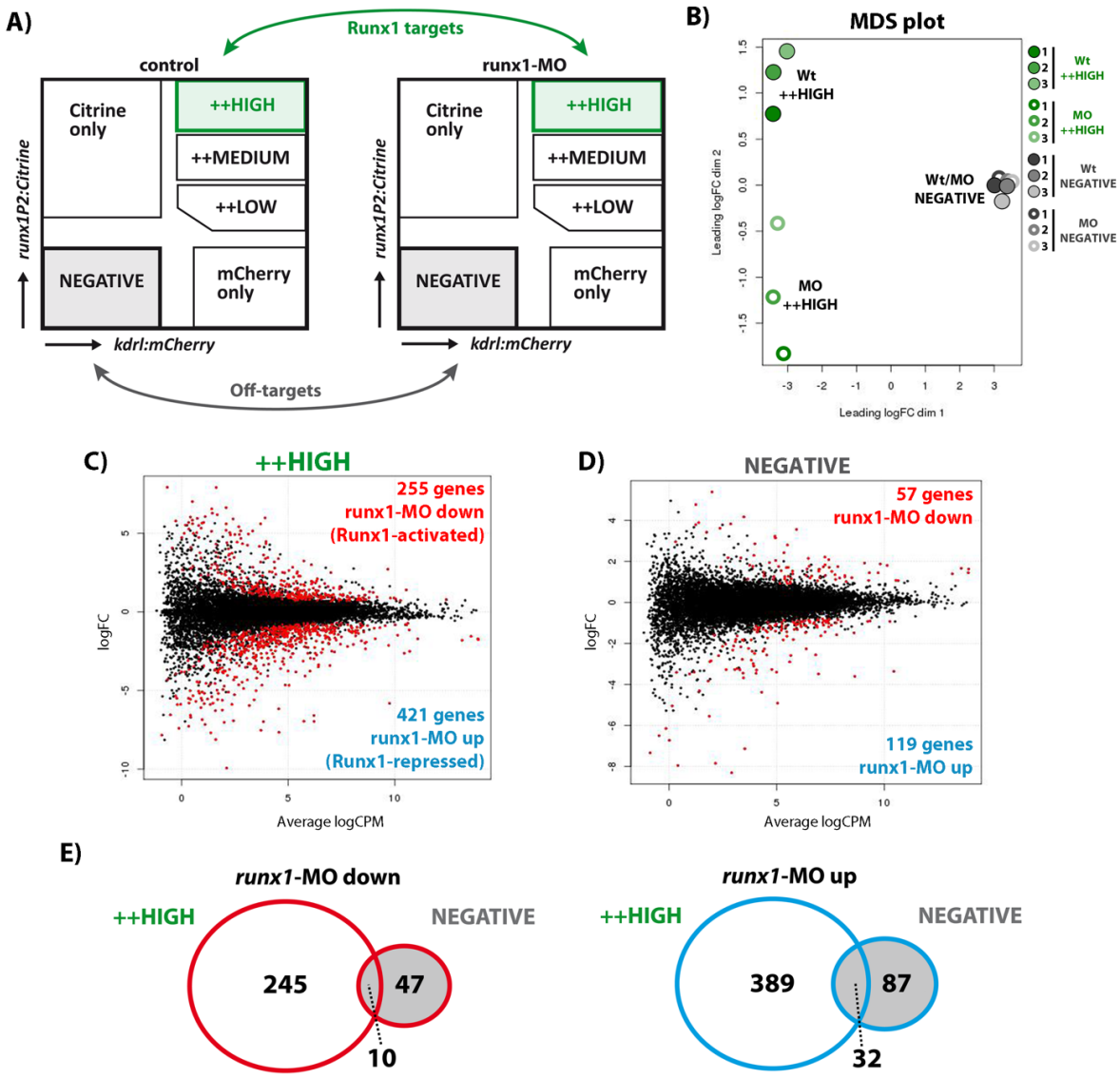


Figure 4-10: Identification of Runx1-regulated genes.

A) Schematic highlighting the populations (++HIGH and NEGATIVE) that were used for the DEG analysis. B) Multidimensional scaling plot (MDS plot) generated by EdgeR for the populations highlighted in (A). C-D) The smear plot shows significantly differentially expressed genes (red) between control and *runx1* MO conditions for cells of the ++HIGH population (C) or the NEGATIVE population (D). E) Differentially expressed genes with a positive correlation between the analysis ++HIGH cells (C) and NEGATIVE cells (D) will be defined as MO off-targets.

Table 4-2: Lists of *runx1* MO off-targets in the zebrafish HE.

DEG analyses for Wt and *runx1* MO conditions were performed with the ++HIGH and NEGATIVE populations. DEGs whose up- and down-regulation correlated positively across both tissues were classified as off-targets.

Off-targets	
Up-regulated in *runx1-MO* condition	**Down-regulated in *runx1-MO* condition**
32 genes	**10 genes**
acss1, atf5a, bbc3, bmb, ccng1, cdkn1a, enpp1, fam160b1, fosl1a, foxi1, foxo3b, gpd1a, hsp90aa1.1, lnx1, mat2al, mdm2, mmp2, mycb, phlda2, phlda3, pik3r3a, rhoub, sesn3, stx18, tp53, trim35-7, tspan18a, ubtd1b, upb1, zgc:123335, zgc:158257, zgc:92287	*g3bp2, itpa, ccni, cnpy1, pbx2, mpx, rpl22l1, gstr, selenbp1, plp1a*

GO term analysis on the Runx1-activated and -repressed genes resulted in a generally low number of enriched GO terms, most of which only contained a few genes (Table 4-3). The most enriched GO term for Runx1-activated genes was “DNA methylation”. Here, we identified all 6 members of the zebrafish dnmt3 family (*dnmt3aa*, *dnmt3ab*, *dnmt3ba*, *dnmt3bb.1*, *dnmt3bb.2* and *dnmt3bb.3*).

The Runx1-repressed genes were enriched for GO terms including “heme biosynthetic process” and “erythroid differentiation”. This is of interest since we detected the same GO terms in Cluster 2 associated genes expressed in the three ++LOW/++MEDIUM/++HIGH populations (Table 4-1). However, it was also shown that RUNX1 can repress the erythroid gene expression programme during megakaryocytic differentiation (Kuvardina et al., 2015). Potentially, zebrafish Runx1 fulfils a similar role in the HE to prevent these cells from activating an erythroid programme during HSPC *de novo* generation.

Table 4-3: GO term analysis on Runx1-activated and Runx1-repressed genes.
Top 10 enriched GO terms are highlighted by a grey background. Additional interesting GO terms are shown below. GO terms with a FDR < 0.05 are highlighted in bold. The biologically most interesting GO terms are highlighted in red (activated) and blue (repressed), respectively.

Runx1-activated	-log10 P-value	Runx1-repressed	-log10 P-value
DNA methylation	**5.5**	**Heme biosynthetic process**	**10.6**
Retina layer formation	3.1	**Cellular iron ion homeostasis**	**6.9**
Phosphorylation	2.8	**Iron ion transport**	**6.9**
Retina morphogenesis in camera-type eye	2.1	**Intracellular sequestering of iron ion**	**5.9**
Cellular amino acid biosynthetic process	1.8	**ATP hydrolysis coupled proton transport**	**4.8**
Brain development	1.7	**Oxygen transport**	**4.6**
Methylation	1.7	**Proton transport**	**4.6**
Organ regeneration	1.6	**Cellular response to xenobiotic stimulus**	**4.6**
Lens development in camera-type eye	1.4	Tetrapyrrole biosynthetic process	3.9
DNA methylation on cytosine	1.4	Erythrocyte differentiation	3.8
Myeloid cell differentiation	1.2	Blood vessel morphogenesis	2.0
Mitotic cell cycle	1.1	Angiogenesis	2.0

4.2.4.2 Identified Runx1-regulated genes contain direct RUNX1 targets in murine haematopoietic progenitor cells

To investigate if the here identified Runx1-regulated genes might contain direct targets, we intersected these lists with a published data-set of murine RUNX1 bound genes generated from the multipotent haematopoietic progenitor cell line 7 (HPC-7) (Wilson et al., 2010a). HPC-7 cells were shown to be a valid model system for HSPCs *in vitro* (Pinto Do Ó et al., 1998; Wilson et al., 2009).

Here, 70 out of the 389 identified Runx1-repressed genes were bound by RUNX1 in HPC-7 cells (Figure 4-11 A). The overlap was small but significant (P-value = 0.049). On the other hand, 68 out of the 245 Runx1-activated genes were bound by RUNX1 in HPC-7 cells (Figure 4-11). Here, the overlap was highly significant (P-value = 1.4e-7). The intersected list for Runx1-activated targets contained genes that we have already identified as evolutionarily conserved in their expression during the transition from the HE to HSPC including *angpt1*, *dnmt3aa*, *dnmt3ab*, *irf1b*, *mpl*, *pik3r1* and *syk*.

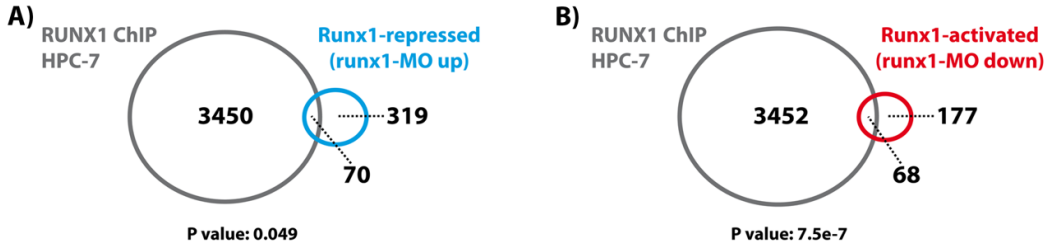


Figure 4-11: Comparison of Runx1-regulated genes with RUNX1 ChIPseq targets in HPC-7 cells.

A) Venn diagram shows the overlap of RUNX1 targets of Wilson et al (2010) with Runx1-repressed genes.

B) Venn diagram shows the overlap of RUNX1 targets of Wilson et al (2010) with Runx1-activated genes.

4.2.4.3 Newly identified Runx1-regulated genes contain RUNX1 targets in t(8;21) AML cells

To investigate if the here identified Runx1-regulated genes might also have implications for Runx1-dependent AML, we cross-compared our gene list to ChIPseq data-sets generated from Kasumi-1 cells. Kasumi-1 cells are a well-studied model for AML that contain a t(8;21) translocation generating a RUNX1/ETO fusion protein (Asou et al., 1991; Dunne et al., 2006). Several studies have investigated the genome-wide occupancy of RUNX1 and RUNX1/ETO in these cells (Li et al., 2016; Ptasinska et al., 2012; Trombly et al., 2015).

In one study performed by Li et al. (2016), the genome occupancy of both RUNX1 and RUNX1-ETO was analysed (Li et al., 2016). To identify high confidence binding peaks, two independent antibodies recognizing different epitopes of the RUNX1 and RUNX1-ETO proteins were used. In total, 6612 high-confidence target genes were found. An intersection with the here identified 634 Runx1-regulated genes revealed that 240 genes were bound by both factors in Kasumi-1 cells (P-value = 0.002). Focusing on the Runx1-activated genes, this list included transcription factors such as *bcl6a*, *erg*, *gfi1ab, irf1b* and *smad3b*, but also members of various signalling pathways including *angpt1*, *notch1b*, *pik3cd*, *pik3r1*, *syk* and *vegfa*. Finally, the *de novo* DNA methyltransferases *dnmt3ab*, *dnmt3ba*, *dnmt3bb.1* and *dnmt3bb.2* were also identified.

In an independent study, Ptasinska et al. (2012) analysed the genome occupancy of RUNX1/ETO in both Kasumi-1 cells and within two t(8;21) patient samples (Ptasinska et al., 2012). High confidence binding events were obtained by overlaying the ChIPseq binding peaks with DNase hypersensitivity

sites (DHS). Overall 1336 common genes were identified to be bound by RUNX1/ETO. An intersection with the here identified 634 Runx1-regulated genes revealed that 58 genes were also bound by RUNX1/ETO in Kasumi-1 cells and t(8;21) patient samples (p-value = 0.007). This list included some of the genes that were identified in the previous comparison including *gfi1ab*, *dnmt3ab*, *irf1b*, *pik3cd* and *smad3b*.

Overall, this cross-comparisons highlights a number of genes involved in cell signalling and gene regulation that are potentially regulated by RUNX1 during haematopoiesis and that might be misregulated by RUNX1/ETO in t(8;21) dependent AML. It also suggests that the control of DNA methylation by RUNX1 is a key process in both, normal and malignant haematopoiesis.

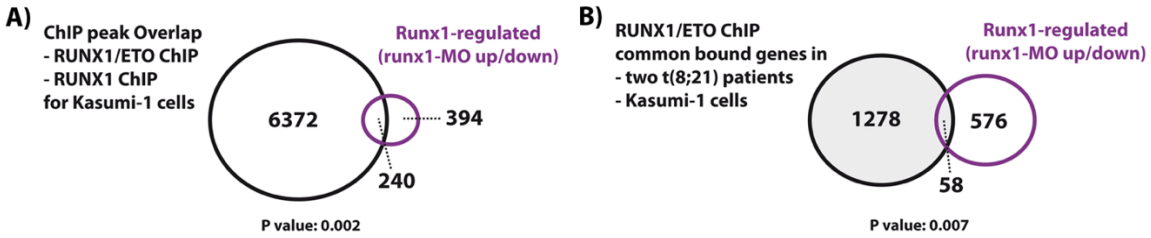


Figure 4-12: Comparison of Runx1-regulated genes with RUNX1 and/or RUNX1/ETO ChIPseq targets in AML cells.

A) Venn diagram shows the overlap of the common RUNX1/ETO- and RUNX1-targets in Kasumi-1 cells of Li et al. (2016) with Runx1-regulated genes. B) Venn diagram shows the overlap of the common RUNX1/ETO-targets in Kasumi-1 cells and t(8;21) patient samples of Ptasinska et al. (2012) with Runx1-regulated genes.

4.2.5 Identification of high-confidence HE-specific Runx1-regulated genes

To restrict our focus to the most interesting genes, we first intersected the HE/DA-specific gene lists with the Runx1-activated/repressed gene lists. Of the 389 Runx1-repressed genes 68 % were DA-specific (Figure 4-13 A). Part of this list were the known Notch targets *her6*, *her9* and *hey2*. These findings imply that Runx1 might be involved in the down-regulation of the endothelial program but also of the Notch signalling pathway in the HE (Zhang et al., 2015).

In contrast, intersecting the HE-specific and the Runx1-activated gene lists resulted in an overlay of 85 genes (Figure 4-13 B) (Appendix 8). This restricted gene list still contained most of the genes that we identified as being evolutionary conserved in their expression during the transition from HE to HSPCs. KEGG pathway analysis showed that some of these genes, including *akt1*, *pik3cd*, *pik3r1*, *prkcbb* and *prkab1a,* show broader functions across several cellular signalling pathways like the FoxO-, Insulin-, mTOR-, VEGF-, Jak-STAT- and ErbB-signaling pathways (Table 4-4). GO term analysis highlighted that "DNA methylation" might be an important mechanism during HSPC *de novo* generation (Table 4-5). Furthermore, 30 of these 85 genes were bound by RUNX1 in HPC-7 cells.

To further centre our focus, we asked which of the 85 HE-specific Runx1-activated genes were differentially expressed during murine haematopoiesis in a data-set curated by our collaborators. Lucas Greder from Marella de Bruijn's laboratory analysed gene expression data from six cell types isolated from the AGM region from E8.5 to E10.5 old mouse embryos. These cell types included endothelium (Endo), different stages of HE (competent (cHE), and specified (sHE)), and pro- and pre-HSCs (Figure 4-13 C) (Rybtsov et al., 2011; Rybtsov et al., 2014; Taoudi et al., 2008). Here, 5834 genes were identified as being differentially expressed across these cell types.

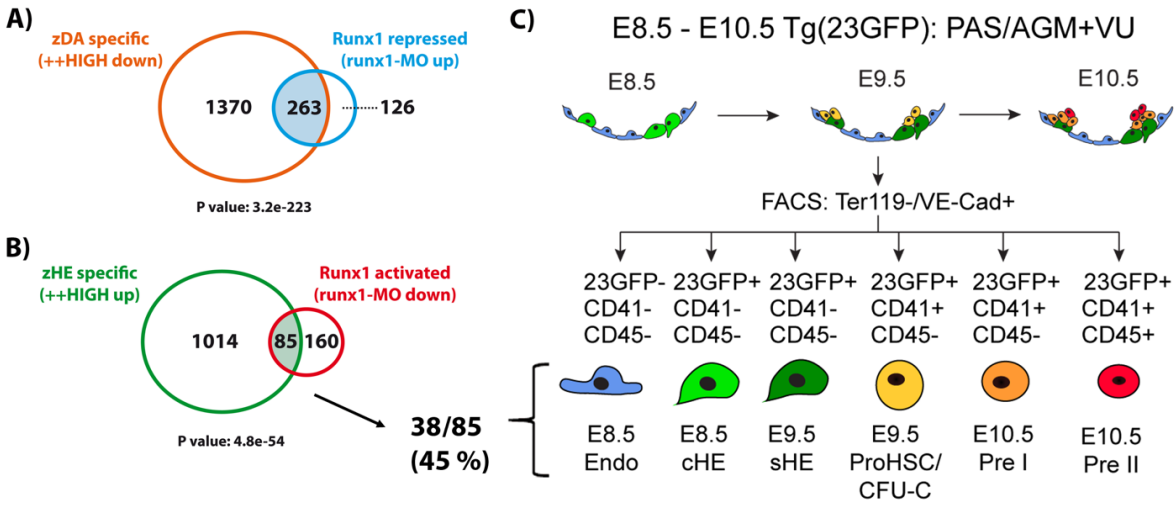


Figure 4-13: Identification of high-confidence zHE-specific Runx1-activated genes.

A) Gene list overlay between the zDA-specific and Runx1-repressed genes. B) Gene list overlay between the zHE-specific and Runx1-activated genes. C) Analysis of mouse AGM cells from E8.5 to E10.5 embryos identified 5834 DEG between endothelium (Endo), competent HE (cHE), specified HE (sHE), Pro-HSCs, and Pre-HSCs type-I and -II. Figure C is adapted from Lucas Greder (unpublished).

Of our 85 zHE-specific Runx1-activated genes 38 homologs were enriched in haemogenic or haematopoietic cell types in this mouse data-set (Figure 4-13 C). Table 4-6 summarizes these 38 genes and provides further information about their expression in the Li et al. (2014) and Kartalaei et al. (2015) data-sets. Additionally, it summarises if these genes were bound by RUNX1 (or RUNX1/ETO) in any of the three mentioned ChIPseq data-sets (Li et al., 2016; Ptasinska et al., 2012; Wilson et al., 2010a). These genes will be taking forward for further validation experiments presented in Chapter 5.

Table 4-4: KEGG Pathway analysis on the 85 HE-specific Runx1-activated genes.
Shown are all identified pathways with their respective P-value and the associated genes.

Term	-log10 (P-value)	Genes
Insulin resistance	2.57	*prkab1a, akt1, pygma, pik3cd, pik3r1, slc27a4*
Cysteine and methionine metabolism	2.40	*dnmt3bb.2, dnmt3bb.1, dnmt3ba, dnmt3ab*
FoxO signaling pathway	2.12	*prkab1a, akt1, pik3cd, bcl6a, agap2, pik3r1*
Insulin signaling pathway	2.09	*prkab1a, akt1, pygma, pik3cd, mknk2b, pik3r1*
Metabolic pathways	1.89	*bcat1, pdxka, dnmt3bb.2, pygma, dnmt3bb.1, dnmt3ba, dnmt3ab, adh5, impdh1b, mthfr, tkfc, ptdss1a, rrm1, nme2b.2, b3gnt2b, acsl1a, nat8l*
mTOR signaling pathway	1.81	*akt1, pik3cd, pik3r1, prkcbb*
VEGF signaling pathway	1.65	*akt1, pik3cd, pik3r1, prkcbb*
Jak-STAT signaling pathway	1.33	*akt1, pik3cd, mpl, pik3r1*
ErbB signaling pathway	1.33	*akt1, pik3cd, pik3r1, prkcbb*

Table 4-5: GO term analysis on the 85 HE-specific Runx1-activated genes.
Shown are all identified GO terms with their respective P-value and the associated genes.

Terms	-log10 (P-value)	Genes
DNA methylation	5.15	*dnmt3bb.2, dnmt3bb.1, dnmt3ba, dnmt3ab*
methylation	2.62	*dnmt3bb.2, ezh2, dnmt3bb.1, dnmt3ba, dnmt3ab*
myeloid cell differentiation	2.07	*brf1a, gfi1ab, bcl6a*
phosphorylation	1.57	*prkab1a, akt1, pik3cd, mknk2b, nme2b.2, pdk2b, prkcbb, syk*
rhythmic process	1.28	*ezh2, mybbp1a*
covalent chromatin modification	1.26	*uhrf1, ezh2, prkcbb*
GTP biosynthetic process	1.18	*impdh1b, nme2b.2*
adaptive immune response	1.18	*prkcbb, syk*
regulation of neurogenesis	1.04	*gfi1ab, bcl6a*

Table 4-6: List of HE-specific Runx1-activated genes with an evolutionarily conserved expression during mouse haematopoiesis.

HE-specific and Runx1-activated genes identified in zebrafish were cross-compared to mouse/human RNAseq/ChIPseq data-sets. Detailed information about the cell-types analysed by L. Greder (unpublished) can be found in Figure 4-13 C. Detailed information about the cell types analysed by Y. Li et al. (2014) and P.S Kartalaei et al. (2012) can be found in Figure 4-9. Green: zebrafish genes that overlaps with the gene-list of the respective study. Grey: zebrafish genes that does not overlap with the respective gene list.

	Zebrafish gene names	**HE to pre-HSCs** (L. Greder unpublished)	**Cluster (C) 1,2,3,4** (Y. Li et al. 2014)	**HECs and HSCs** (P.S. Kartalaei et al. 2015)	**Runx1 binding** Within 3 ChiPseq studies	**Legend**
1	acsl1a	Pre II	/	/	A, B, C	**A: Runx1 ChiP in HCP7 cells** (Wilson et al. 2010)
2	angpt1	Pre II	C1	HSCs	A, B	
3	ankrd10a	Pre I	C4	/	/	
4	b3gnt2b	sHE	/	/	A, B	
5	coro1b	cHE	/	/	/	
6	cyfip2	Pre II	C1	/	/	**B: RUNX1 and RUNX1/ETO ChiP in Kasumi-1 cells** (Y. Li et al. 2016)
7	dnmt3ab	Pre II	C4	HSCs	A, B, C	
8	dnmt3ba	Pre II	C1	HSCs	B	
9	dnmt3bb.1	Pre II	C1	HSCs	B	
10	dnmt3bb.2	Pre II	C1	HSCs	B	
11	ezrb	Pre II	C4	/	/	
12	gfi1ab	Pre I	C1	HSCs	B, C	**C: RUNX1 ChiP in Kasumi-1 cells and two t(8;21) patients** (A. Ptasinska et al 2012))
13	iqgap1	Pre I	/	/	A, B	
14	irf1b	Pre II	C1	/	A, B, C	
15	itm2cb	sHE	/	/	A, B. C	
16	mcm5	Endo	/	/	A	
17	mpl	Pre II	C1	/	A	
18	mrpl28	Endo	/	/	/	
19	mybbp1a	Pre II	C1	/	/	
20	parp1	Pre II	/	/	B	
21	pbrm1l	sHE	C2	/	B	
22	pdxka	Pre II	C1	/	B	
23	pik3cd	Pre II	C1	/	B, C	
24	pik3r1	Pre II	C1	/	A, B	
25	pmp22b	cHE	/	/	/	
26	prkcbb	sHE	/	/	B	
27	selp	Pre II	C1	/	A	
28	sept9b	sHE	/	/	/	
29	skap1	cHE	/	/	A, B	
30	slc27a4	sHE	/	/	/	
31	stmn1a	Pre II	C1	/	A, B	
32	syk	Pre II	C1	HSCs	A, B	
33	tfr1b	Endo	/	/	A	
34	tp53i11a	cHE	/	/	/	
35	wdhd1	Pre II	C2	HSCs	B	
36	xbp1	sHE	/	/	B, C	
37	zgc:92791	Pre II	C1	/	/	
38	znf296	Pre II	C1	HSCs	/	

4.3 Discussion

4.3.1 The executed RNAseq experiment allowed for a robust analysis of DEGs and controlled for MO off-targets

Here I described the global gene expression analysis of the HE and related endothelial tissues. We carefully designed the RNAseq experiment to avoid possible technical and biological adverse effects. First, as many cells as possible were isolated for each population to reduce variation potentially introduced by smaller batches. The MDS plot analysis proved the robustness of the performed protocol (Figure 4-5 B).

Second, we concluded that analysing the HE at a time point around 28-30 hpf is most informative for multiple reasons: I) Even though *runx1* expression becomes detectable in the first HE cells from 23/24 hpf onwards, prominent expression in most HE cells becomes evident only within the next couple of hours. II) Transcriptional changes introduced by Runx1 will not be immediately detectable due to a transcriptional delay of down-stream targets (Darzacq et al., 2007; Swinburne and Silver, 2008). For example, the recently identified Runx1 target *dnmt3bb.1* is 89,746 bp long, which suggests a transcriptional delay of several hours (Gore et al., 2016). III) It is not known if the loss of *runx1* in the HE will cause a general block in cell differentiation. This would introduce the risk of comparing a more mature HE (control) and a less mature HE (*runx1* MO) rather than the same cell types. To limit the risk of these potential adverse effects, no embryo older then 30 hpf was used. Indeed, the large number of identified genes with an evolutionarily conserved expression during definitive haematopoiesis (Figure 4-9) and the identification of over 634 Runx1-regulated genes, many of which are bound by RUNX1 in other haematopoietic cell types (Figure 4-11 and Figure 4-12), suggests that the time point of analysis was well chosen.

For the functional experiment, a major focus was on finding the best controls for the *runx1* MO. Embryonic clutches from different parents can show slight differences in their speed of development. To distinguish MO-induced effects from such artefacts, every embryonic batch was split in half (control vs *runx1* MO) and then processed accordingly. Next, even though we and others have shown that the here used *runx1* MO phenocopies the *runx1*$^{W84X/W84X}$ mutant (Sood et al., 2010) (Figure 4-1 C), additional off-target effects might be introduced through the usage of MOs in general. On the

other hand, control MOs can potentially introduce their own off-target effects. Instead, we argued that the comparison of the NEGATIVE populations, which do not express the *Citrine* transgene and hence should also not express any endogenous *runx1*, is more informative. In our setup, the NEGATIVE population represents a multitude of different cell types and serves as the best reflection of overall embryonic gene expression. The MDS plot analysis showed that the *runx1* MO separated the ++HIGH samples but not the NEGATIVE samples (Figure 4-10 B), indicating good specificity of the runx1 MO. One of the most up-regulated genes of the NEGATIVE fraction was *tp53*. The *tp53* gene is a well-known response gene for cellular stress, which in this case might be up-regulated by introducing *runx1* MO as a cellular stressor (Hu et al., 2012). In fact, we tested a selection of 5 *runx1* MO off-target genes including *tp53* by qPCR analysis using the *runx1*$^{W84X/W84X}$ mutant (Chapter 5 Figure 5-7). None of these genes showed clear expression changes due to the loss of Runx1. This supports our strategy of MO off-target identification.

4.3.2 The zebrafish HE shows an early loss of the endothelial programme including *dll4*

As discussed above, our DEG analysis for the identification of HE-specific genes was confirmed by the detection of known HE marker genes (*tal1*, *cmyb*, *gfi1aa*, *gata2b* and *dnmt3bb.1*). On the global level DA-specific genes were enriched for GO terms including "angiogenesis", "vascular development", "vasculogenesis", and "patterning of blood vessels", which in turn indicates, that the HE was depleted for such endothelial genes, even though the HE was still embedded in the DA framework around 28 to 30 hpf. Similar observations have been made in mouse where changes towards a haematopoietic fate in the HE were discovered two days before the emergence of intra-aortic clusters (IAC) (Swiers et al., 2013a). Despite being still part of the endothelial wall, these mouse HE cells already showed an early loss of their previous endothelial potential (Swiers et al., 2013a).

Furthermore, DA-specific genes contained 24 genes associated with the KEGG pathway "Notch signalling pathway". These genes included the receptors *notch1a, notch2* and *notch3*, the ligands *dlb*, *dlc*, *dld*, *dll4*, *jag1b* and *jag2b,* and the Notch target genes *her6*, *her9* and *her15.1*. We also identified the Notch target gene *hey2* as being HE down-regulated. These findings agree with Richard et al.,

(2013) who have shown that *runx1* and *jag2* have mutually exclusive expression patterns in the DA endothelium in chick embryos at the stage of cluster formation and that *dll4* expression was down-regulated in IAC cells (Richard et al., 2013). These results further indicate that the Notch pathway, which is initially required for the establishment of the HE and for *runx1* expression (Burns et al., 2005; Butko et al., 2016; Hadland et al., 2004; Kumano et al., 2003), needs to be down-regulated to enable further haematopoietic differentiation (Butko et al., 2016; Tang et al., 2013). A similar mechanism has been shown whereby Notch1 is degraded in a Gpr183-dependent manner (Zhang et al., 2015). In agreement with this, we found the highest expression of *gpr183a* and *gpr183b* within the DA (Cluster 2).

HSC specification requires multiple Notch signalling inputs and the complexity is further increased by the different ligands involved in the Notch signalling pathway (Butko et al., 2016). As it was shown in mouse, both DLL4 and JAG1 can signal through NOTCH1 but with different signalling strength (Gama-Norton et al., 2015). Thus, a stronger NOTCH1 signalling, which is induced by the DLL4 ligand, drives the arterial programme, whereas JAG1 stimulates a low NOTCH1 signalling strength that is required for HSC specification (Gama-Norton et al., 2015; Robert-Moreno et al., 2008). A similar requirement was shown in zebrafish where the loss of Jag1a leads to reduced expression of *runx1* in the HE at 28 hpf (Espín-Palazón et al., 2014; Monteiro et al., 2016). In the Espin-Palazon et al. (2014) study, the endothelial population was identified as a potential Jag1a-presenting cell in contrast to developing HSPCs. Interestingly, *jag1a* was not found to be differentially expressed within the DA in our dataset. This discrepancy is likely caused by the different experimental designs. In the study by Espin-Palazon et al. (2014), *kdrl*$^{+}$ endothelial cells and *cmyb*$^{+}$*kdrl*$^{+}$ HSCs were analysed at 48 hpf. In contrast, we analysed purified cell populations of the HE and the DA roof at 28-30 hpf without the contamination of "general" endothelium. In fact, *jag1a* showed the highest expression in the mCherry^{+} population (see also Chapter 5 Figure 5-5 I). In contrast to the uniform expression of *jag1a* identified in all DA cells, *dll4* expression was depleted from the HE, suggesting that the ratio between Dll4 and Jag1a signalling, that define the arterial and haemogenic potential, are regulated by a specific down-regulation of *dll4* in the HE population. These mechanisms will be further investigated in Chapter 5.

Another interesting KEGG-pathway enriched only for the DA roof population and hence depleted from the HE was "TGF-beta signalling pathway". Identified genes contained 23 genes of the BMP and TGFβ signalling branches including the *tgfbr2* receptor. We have recently shown that TGFβ signalling through the Tgfbr2 receptor is necessary for the expression of *runx1*, *gata2b*, *gfi1aa,* and the Notch ligand *jag1a* in the early HE (Monteiro et al., 2016). Our new findings suggest that the TGFβ signalling pathway is dynamically regulated which involves a repression during differentiation into HSPCs. Supporting this hypothesis, recent studies reported an inhibitory effect of the TGFβ pathway during AGM stage haematopoiesis (Sato et al., 2008; Vargel et al., 2016).

In summary, these data imply that an early loss of the arterial programme is occurring in the zebrafish HE. It also suggests that signalling pathways that are required for the differentiation of the HE (including Notch and TGFβ) need to be down-regulated for the subsequent differentiation from the HE state into haematopoietic progenitor cells.

4.3.3 *Runx1* positively regulates expression of *dnmt3* genes which might have implications for AML

Genes activated by Runx1 were relatively diverse with an in general low enrichment for identified GO terms (Table 4-3). The most prominent GO terms were "DNA methylation", "phosphorylation" and "methylation", suggesting that Runx1 acts on genes that either modify protein-protein networks during differentiation (Van Hoof et al., 2009) or regulate the epigenetic state by DNA methylation. Indeed, 4 of the 6 *de novo* DNA methylases present in zebrafish were Runx1-activated and all 6 isoforms were HE-specific. Furthermore, only 18 % of all Runx1-repressed genes were bound by RUNX1 in HPC-7 cells (Figure 4-11 A). This suggests that the repressive function of Runx1 is mostly accomplished by the regulation of transcriptional repressors, as it has been shown for the murine genes *Gfi1* and *Gfi1b* (Lancrin et al., 2012). Additionally, DNA methylation is often associated with gene repression, especially when promoter regions are methylated (Jones, 2012). Since we identified most *dnmt3* genes present in zebrafish to be regulated by Runx1, we hypothesize that they might be involved in a broader role during haematopoiesis. A specific role for *dnmt3bb.1* during definitive haematopoiesis has already been reported (Gore et al., 2016).

Our findings further suggest that a Runx1-dependent regulation of DNA methylation might also have implications for AML. Of note, both the human *DNMT3A* and *DNMT3B* genes were bound by RUNX1 and RUNX1/ETO in Kasumi-1 cells (Li et al., 2016) and *DNMT3A* was also bound by RUNX1/ETO in t(8;210) AML patient samples (Ptasinska et al., 2012). DNMT3A mutations are frequently detected in patients with *de novo* AML and are associated with a poor outcome (Ley et al., 2010; Thol et al., 2011; Yan et al., 2011). *Dnmt3a* mutations mostly result in a defective methylase activity (Holz-Schietinger et al., 2012; Russler-Germain et al., 2014). Consequently, DNMT3A-mutated AML samples are DNA hypo-methylated (Voigt and Reinberg, 2013). Furthermore, it was shown that a deletion of *Dnmt3a* in murine HSCs blocks differentiation and leads to the up-regulation of self-renewal genes resulting in HSC expansion (Challen et al., 2011). These and our findings suggest that *DNMT3A* expression levels might be affected in RUNX1 loss of function-dependent leukaemia. Interestingly, *DNMT3A* mutations are not found in patients with a t(8;21) mutations (Ley et al., 2010). Instead, a mechanism was identified in which RUNX1/ETO cooperates with HIF1α, which then promotes the transcriptional activation of *DNMT3A* leading to DNA hyper-methylation (Gao et al., 2015). This study showed that the DNMT3A promoter contains at least four RUNX1/ETO binding sites and direct binding was validated.

Overall, these data indicate that DNA methylation in the haematopoietic system needs to be tightly regulated. The role of the *dnmt3* isoforms during haematopoiesis will be further analysed and discussed in Chapter 5.

Significantly, other genes that we identified as being Runx1-activated and HE-specific (*gfi1ab*, *irf1b*, *pik3cd* and *smad3b)* were also bound by RUNX1 in the above mentioned AML studies. Direct binding of RUNX1 to the *PIK3CD* gene in acute megakaryocytic leukaemia cells has been shown previously (Edwards et al., 2009).

4.3.4 Runx1 keeps the erythroid programme in check and presumably represses endothelial genes via *gfi1ab*

Genes repressed by Runx1 were enriched for GO terms including "heme biosynthetic process" and "erythroid differentiation" (Table 4-3). Interestingly, genes in Cluster 2 (higher expressed in all ++ populations) were also strongly enriched for the very same GO terms (Table 4-1). This indicates a general close relationship between the haematopoietic system and the arterial cells of the embryonic DA (Eilken et al., 2009; Zhang et al., 2013). Here, Runx1 seems to hold the erythroid differentiation in check while the HE is down-regulating its endothelial program and undergoing haematopoietic differentiation into HSPCs. Consistent with this notion, it has been shown that RUNX1 can repress the erythroid programme during megakaryocytic differentiation (Kuvardina et al., 2015).

Furthermore, GO terms enriched for the Runx1-repressed genes included "blood vessel morphogenesis" and "angiogenesis". It has been shown in mouse that RUNX1 directly binds to and activates GFI1 and GFI1b expression, which in turn repress the endothelial programme (Lancrin et al., 2012; Thambyrajah et al., 2015). In agreement with this, we found the two isoforms *gfi1aa* and *gfi1ab* to be specifically expressed in the HE. As shown recently, we also found that the expression of *gfi1aa* was Runx1 independent (Thambyrajah et al., 2016). Instead, *gfi1ab* was the gene with the highest fold enrichment for the ++HIGH population and with the 5th strongest fold-change of all the Runx1-activated genes. The *gfi1aa* and *gfi1ab* genes are the result of a genome duplication event (Cooney et al., 2013). So far it is not resolved how the important role for *Gfi1* genes during mouse haematopoiesis is precisely recapitulated in zebrafish. Loss of *gfi1aa* does not affect haematopoiesis but leads to an enhanced expression of *gfi1ab* (Thambyrajah et al., 2016). The role of *gfi1ab* for zebrafish haematopoiesis remains to be defined.

Interestingly, RUNX1 showed no binding to the *Gfi1* locus in HPC-7 cells but to the *Gfi1b* locus (Wilson et al., 2010a). This might reflect the greater maturation of HPC-7 cells compared to the here analysed HE cells (Pinto Do Ó et al., 1998). It was shown that *Gfi1* is only transiently expressed in the HE and in cells undergoing EHT but becomes subsequently replaced by *Gfi1b* in IAC cells (Thambyrajah et al., 2015). In agreement with this, another study that focused on early blood progenitors (EBPs) development *in vitro* found the *Gfi1* locus to be indeed bound by RUNX1 (Tanaka

et al., 2012). Overall, this EBP ChiPseq data-set showed only little overlap with the here identified Runx1-regulated genes (data not shown). This discrepancy might have resulted from the generally lower number of identified RUNX1 targets (718 genes) in that study. Otherwise, it could reflect the likelihood that *in vitro* differentiation mimics the YS blood production rather than the intra-embryonic haematopoiesis (Goode et al., 2016). In the context of different cell types with specific gene regulatory networks, RUNX1 might bind to a distinct set of targets (Ito, 1999; Levine and Davidson, 2005; Lichtinger et al., 2012). Nevertheless, the identification of *Gfi1* as a RUNX1 target in EBPs proves that *Gfi1* has an important and conserved role during the EHT-like cell differentiation process in general, likely representing a core mechanism. The role of *gfi1ab* during zebrafish haematopoiesis and its relationship to *gfi1aa* will be further addressed in Chapter 5.

4.3.5 High-confidence Runx1 targets include genes with up-regulated expression during definitive haematopoiesis

Our sequencing strategy allowed us to identify a large number of HE-specific genes and genes that are potentially regulated by Runx1. To identify target genes that are involved in haematopoietic differentiation, we focussed primarily on the Runx1-activated genes. The Runx1-repressed genes were mostly DA-specific and therefore endothelial (Figure 4-13 A).

To restrict further analysis to the most promising Runx1-activated target genes, we decided to carry forward only those genes that are also HE-specific. That was true for 85 of the 245 Runx1-activated genes. Next, to identify genes with an evolutionarily conserved expression during haematopoiesis, we overlaid this list with several RNAseq and RUNX1-ChIP data-sets from mouse and human. Thereby, we centred our analyses on 38 genes that showed expression during haematopoietic development in our collaborators' data-set for which they were analysing cells from the endothelial stage to pre-HSC type-II cells (Marella de Brujin laboratory; unpublished) (Figure 4-13). Intriguingly, 15 of these 38 genes were also bound by RUNX1 in HPC-7 cells. An additional 11 genes were also bound by RUNX1 or RUNX1/ETO in Kasumi-1 cells. Strikingly, most of these 38 genes showed the highest expression levels in the most mature haematopoietic cell types in the respective gene expression

studies (Table 4-6). This finding implies that the here identified genes are involved in the establishment of the definitive haematopoietic fate, rather than in the temporarily occurring transition of endothelial to haematopoietic cells. This agrees with the finding that expression of *Gfi* or *Gfi1b* in the HE is sufficient to drive EHT (Lancrin et al., 2012).

The restricted list of 38 genes with an evolutionarily conserved expression during haematopoiesis contained several other genes with a reported expression and function within the haematopoietic system including haematopoietic progenitors in the FL or in HSCs. In addition to the already mentioned *dnmt3* and *gfi1ab* genes, this list included *angpt1*, *irf1b*, *mpl*, *pik3cd*, *pik3r1*, *prkcbb* and *syk*. Besides a reported expression in the DA for the mouse *Dnmt3b* and the zebrafish *dnmt3bb.1* genes (Gore et al., 2016; Watanabe et al., 2004), none of these genes have been linked to the HE so far. For most of these genes a regulative connection to RUNX1 has not been established yet. All these genes will be further analysed and discussed in Chapter 5.

4.4 Conclusion and directions

Here, I have presented a detailed analysis of the gene expression in the zebrafish HE and of the Runx1-regulated genes within that tissue. The analysis resulted in 85 genes that are both HE-specific and Runx1-activated. The validation of our analysis strategy is confirmed by the identification of known HE marker genes (*tal1*, *cmyb*, *gfi1aa*, *gata2b* and *dnmt3bb.1*), as well as genes dependent on Runx1 (*dnmt3bb.1*). Interestingly, several *dnmt3* isoforms seem to play an important role during haematopoiesis, a finding that will be further investigated.

The here generated data also identify the *gfi1aa* gene as being HE-specific but independent of Runx1. In contrast, the *gfi1ab* gene is not only specific in its expression in the HE but appears to be dependent on Runx1 and thus will be further investigated as well.

Further analysis focused on identifying additional genes with evolutionarily conserved expression during haematopoiesis. This resulted in a restricted list of 38 targets that contains genes with known roles in the haematopoietic system (*angpt1*, *dnmt3* family members, *gfi1ab*, *irf1b*, *mpl*, *pik3cd*, *pik3r1*,

prkcbb, and *syk)*. Moreover, 26 of these 38 can be bound by RUNX1 or RUNX1-ETO in HPC-7 or Kasumi-1 cells, including all the above mentioned genes.

In the next chapter, we will perform validation experiments for the here identified genes. First, ISHs will verify the correct spatial expression in the HE. Next, qPCR experiments using the *runx1*$^{W84X/W84X}$ mutant background will validate a true regulation by Runx1. This qPCR analysis will be combined with a time-course analysis to gain insights into the gene expression dynamics during HE maturation.

- Chapter 5 –

Detailed analysis of the consequences of the loss of Runx1 on the HE *in vivo*

5.1 Introduction

RUNX1 is often regarded as a key TF and haemogenic expression of *Runx1* was found to be controlled by a multitude of upstream signalling pathways and regulators, which include among others Notch signalling (Burns et al., 2005; Kobayashi et al., 2014), TGFβ and BMP signalling (Monteiro et al., 2016; Wilkinson et al., 2009), FGF signalling (Pouget et al., 2014), inflammatory signalling (Espín-Palazón et al., 2014; He et al., 2015; Li et al., 2014; Lu et al., 2013; Sawamiphak et al., 2014), PI3K signalling (Li et al., 2015), prostaglandin synthesis (North et al., 2007), GATA2 (Butko et al., 2015), TAL1 (Zhen et al., 2013), and blood flow (North et al., 2009). Due to its pivotal role in the HE, *Runx1* is also widely used as a functional readout in studies that focus on early events during definitive haematopoiesis (Huang et al., 2013).

Despite this large body of data about pathways acting upstream of *Runx1* expression in the HE, detailed knowledge about processes downstream of RUNX1 during definitive haematopoiesis remain limited. Hemizygous *Runx1* mutants reveal changes in the temporal and spatial appearance of HSCs indicating a dosage dependency (Cai et al., 2000), whereas homozygous mutants fail to undergo EHT (North et al., 1999; North et al., 2002). In such *Runx1* knockout mice, the HE can still mature to a stage at which it expresses *Itga2* (*CD41*) (Liakhovitskaia et al., 2014), a marker of haematopoietic commitment (Ferkowicz, 2003; Mikkola et al., 2003; Mitjavila-Garcia et al., 2002).

Thus far, effects of loss of RUNX1 on the global gene expression programme during haematopoietic development have only been studied in YS-derived cells or in HE-like cells derived from *in vitro* differentiation cultures (Lancrin et al., 2012; Lichtinger et al., 2012; Lie-A-Ling et al., 2014; Sakai et al., 2009; Tanaka et al., 2012). These studies reported that RUNX1 is involved in the repression of the endothelial programme to facilitate EHT, that it might be involved in orchestrating cell positioning and assembly via regulation of genes involved in cell adhesion and migration, and that RUNX1 induction causes a global reorganization of TAL1 and FLI1 binding sites (Lancrin et al., 2012; Lichtinger et al., 2012; Lie-A-Ling et al., 2014). Nevertheless, how RUNX1 facilitates the full commitment to a haematopoietic fate in the developing HSCs is still not fully understood. Strikingly, in mouse embryos RUNX1 was shown to be still required in at least a subset of forming HSCs at a time when most HE cells have already differentiated into intra-aortic cluster (IAC) cells (Tober et al.,

2013). Therefore, it is likely that RUNX1 is involved in the regulation of additional yet unidentified genes. To our knowledge, a global gene expression analysis of the intra-embryonic HE in a loss of RUNX1 situation has not been reported yet.

Even though cells in most haematopoietic *in vitro* culture systems transit through an HE-like state, it is not clear to what extent these cells resemble their intra-embryonic counterparts. It can be speculated that certain inherent differences exist between such *in vitro* generated HE-like cells and the aortic HE, capable of producing HSC, that are responsible for the gap in understanding the full haematopoietic activity of RUNX1. Furthermore, the mouse embryo harbours at least two different types of HE in the YS and in the embryo proper (Chen et al., 2011; Frame et al., 2016; Hirschi, 2012). The HE that gives rise to HSCs is always found in the lining of the major arteries of the early embryo, mainly in the DA (De Bruijn et al., 2002; Medvinsky and Dzierzak, 1996). Lymphoid potential, which is usually used as a readout for definitive haematopoiesis *in vitro,* can also be detected during YS haematopoiesis (Lin et al., 2014; Rhodes et al., 2008; Yoshimoto et al., 2011). In this regard, the *in vitro* generation of transplantable HSCs without the need of genetic manipulations has yet to be reported (Lis et al., 2017; Sugimura et al., 2017). To achieve this goal, it is important to reconstruct the exact lineage path that results in the generation of intra-embryonic HE found in the DA.

The relationship between the haemogenic and the arterial lineages remains a controversial question. In all vertebrate model organisms, the aortic HE is always found in close association with the arterial vasculature (Ciau-Uitz et al., 2013; De Bruijn et al., 2002; Gering and Patient, 2005; North et al., 1999; Richard et al., 2013). This led to the hypothesis that the aortic HE lineage might be derived from the arterial endothelium. However, a recent study that utilized an optimized *in vitro* protocol generating definitive progenitor cells with lymphoid potential suggested that the arterial and the haemogenic lineages might be separate (Ditadi et al., 2015). Therefore, more comprehensive data about the intra-embryonic HE and its relation to the arterial lineage are needed to resolve this issue.

In the previous Chapter, I described the identification of HE-specific and Runx1-regulated genes by performing next generation sequencing analysis on cells isolated from zebrafish embryos. This chapter aims to validate the RNAseq results and to examine the consequences of the loss of Runx1 for the aortic HE in greater detail.

5.1.1 Chapter aims

There are three main aims in this chapter:

1. The validation of high-confidence Runx1 target genes identified by RNAseq.
2. Functional analysis of novel Runx1 target genes.
3. The characterization of the HE in *runx1* mutants.

First, I present an extensive multilevel validation screen that is based on ISH and qPCR based methods. ISH was performed for 29 candidate genes to prove the correct spatial expression in the HE. qPCR analysis on sorted cells from Wt and *runx1*$^{W84X/W84X}$ mutant embryos validated the functional data generated by the *runx1* MO. Second, loss of function experiments were performed for the *gfi1ab* and *gfi1aa* genes to clarify their role during EHT in zebrafish. Third, the analysis of *dll4* expression in the HE of *runx1*$^{W84X/W84X}$ mutants showed that HE cells become only *dll4*$^{-}$ after *runx1* induction, suggesting that the aortic HE is very likely derived from *dll4*$^{+}$ aortic endothelium/precursors.

5.2 Results

5.2.1 Identification of high-confidence potential Runx1 target genes

In the previous chapter, we analysed the global gene expression of the HE and additionally identified 634 differentially expressed genes (DEGs) in *runx1* morphants. To identify the most relevant target genes, a selection logic was applied that incorporated several published and unpublished genome-wide gene expression and RUNX1 ChIPseq data-sets from mouse and human (Kartalaei et al., 2015; Li et al., 2014; Li et al., 2016; Ptasinska et al., 2012; Wilson et al., 2010a). In that way, we could isolate genes with an evolutionarily conserved expression during definitive haematopoiesis, which we assumed will be the most relevant genes. Genes that have the potential to be bound by RUNX1 in other haematopoietic cell types will be considered as high-confidence targets.

Focusing on the 1099 HE-specific genes, 85 were identified as being positively regulated by Runx1 (Figure 5-1). Of these 85 genes, 38 (45 %) were expressed in haematopoietic cells of our collaborator's data-set (Marella de Brujin, personal communication) (Figure 5-1). Further intersection with genes identified by the Li et al. (2014) and Kartalaei et al. (2015) studies condensed the list to 22 genes (Figure 5-1). Lastly, we retained only those genes that showed a RUNX1 or RUNX1/ETO binding peak in at least one of the three incorporated ChIPseq studies using HPC-7 or Kasumi-1 cells (Figure 5-1) (Li et al., 2016; Ptasinska et al., 2012; Wilkinson et al., 2009). In total, we identified 16 high-confidence RUNX1 target genes that we aimed to validate and study in more detail (Table 5-1).

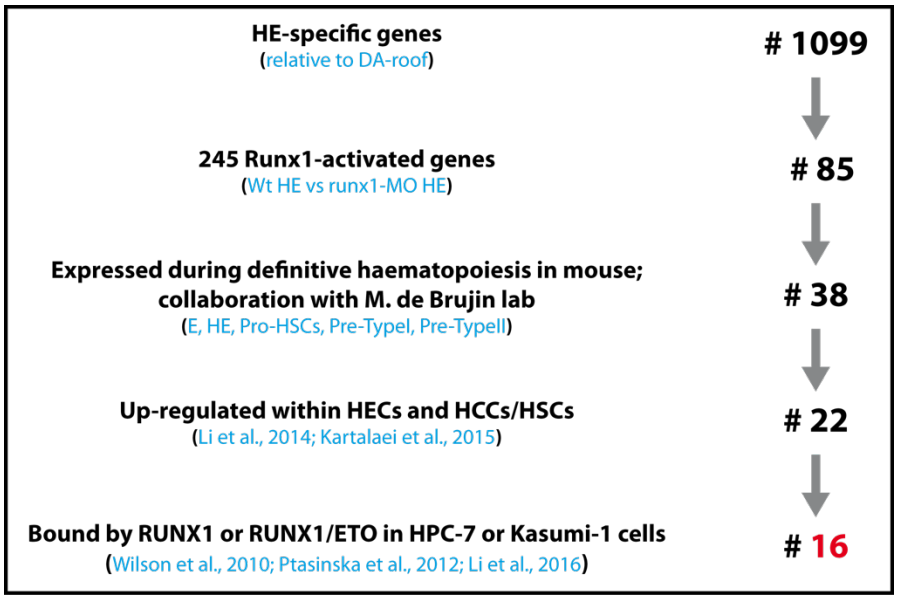


Figure 5-1: Selection logic for the identification of high-confdence potential Runx1 target genes.
The left side of the diagram describes which expression data-sets were used for the gene list intersection. The right side shows the number of retained genes.

Table 5-1: List of the 16 identified high-confidence potential Runx1 target genes.
The genes of this list were retained by following the selection logic outlined in Figure 5-1. Green: Gene is present in the respective data-set. Grey: Gene is absent in the respective data-set. C1/C2/C4: Cluster 1/2/4 as defined by the Y.Li et al (2014) study. Pre I/II: pre-HSCs type-I/II. sHE: specified HE.

	Zebrafish gene names	HE to preHSCs (L. Greder unpublished)	Cluster 1,2,3,4 (Y. Li et al. 2014)	HECs and HSCs (P.S. Kartalaei et al. 2015)	Runx1 binding Within 3 ChiPseq studies	Legend
1	angpt1	Pre II	C1	HSCs	A, B	**A:** Runx1 ChiP in HCP7 cells (Wilson et al. 2010) **B:** RUNX1 and RUNX1/ETO ChiP in Kasumi-1 cells (Li et al. 2016) **C:** RUNX1 ChiP in Kasumi-1 cells and two t(8;21) patients (Ptasinska et al. 2012)
2	dnmt3ab	Pre II	C4	HSCs	A, B, C	
3	dnmt3ba	Pre II	C1	HSCs	B	
4	dnmt3bb.1	Pre II	C1	HSCs	B	
5	dnmt3bb.2	Pre II	C1	HSCs	B	
6	gfi1ab	Pre I	C1	HSCs	B, C	
7	irf1b	Pre II	C1	/	A, B, C	
8	mpl	Pre II	C1	/	A	
9	pbrm1l	sHE	C2	/	B	
10	pdxka	Pre II	C1	/	B	
11	pik3cd	Pre II	C1	/	B, C	
12	pik3r1	Pre II	C1	/	A, B	
13	selp	Pre II	C1	/	A	
14	stmn1a	Pre II	C1	/	A, B	
15	syk	Pre II	C1	HSCs	A, B	
16	wdhd1	Pre II	C2	HSCs	B	

5.2.2 Expression analysis of high-confidence potential Runx1 target genes by ISH

The first validation experiments were done by whole mount ISH. This initial screen aimed to investigate the correct spatial expression of the 16 high-confidence Runx1 target genes we identified (Table 5-1). ISH probes were successfully generated for 13 of these 16 genes (Table 5-2). Additionally, ISH probes for a second set of 16 potential target genes were generated (Figure 5-2). These additional genes were identified within the list of the 85 HE-specific Runx1 target genes and considered for various reasons (members of the *dnmt* family, kinases, TFs, etc.). The *itga4* gene was also considered because it has been identified as a marker that distinguishes haemogenic from non-haemogenic $Cdh5^{+}$ (VEcad) cells in culture (Ogawa et al., 1999). *Itga4* wasn't identified by the DEG analysis due to a missing annotation, but the gene tracks for the *itga4* genomic region suggest a HE-specific Runx1-dependent expression (Appendix 9). Thus, we decided to include *itga4* into the screen.

Table 5-2: List of genes that were probed for expression by ISH.
ISH probes were generated for 13 of the 16 identified high-confidence potential target genes (Table 5-1). Additional 16 ISH probes were generated for additional potential target genes identified in zebrafish

ISH probe generated (13); high-confidence potential target genes
angpt1, dnmt3ab, dnmt3ba, dnmt3bb.1, dnmt3bb.2, gfi1ab, irf1b, mpl, pbrm1l, pik3cd, pik3r1, syk, wdhd1
ISH probe generated (16); additional potential target genes
dnmt1, dnmt3aa, dnmt3bb.3, agap2, bcl6a, itga4, itm2cb, mknk2b, prkab1a, prkcbb, skap1, tfr1b, uhrf1, xbp1, ppm1g, map2k6

The ISH screen showed that 8 of the 13 high-confidence potential Runx1 target genes were expressed in the HE (Figure 5-2 A-H). This was true for the 4 *dnmt3* isoforms *dnmt3ab*, *dnmt3ba*, *dnmt3bb.1* and *dnmt3bb.2* (Figure 5-2 A-D; red arrows) and for the genes *angpt1*, *gfi1ab*, *irf1b*, *pik3cd* (Figure 5-2 E-H; red arrows). A specifically strong expression in the HE was detected for *dnmt3ba* and *dnmt3bb.1 (*Figure 5-2 *B, C)*, whereas *dnmt3ab* and *dnmt3bb.2* were weaker and had a generally broader expression (Figure 5-2 A, D). The *angpt1* gene showed, in addition to expression in the HE, a strong expression in the hypochord right above the DA (Figure 5-2 E). Expression of *gfi1ab* was not

always detectable in all analysed embryos and was generally relatively weak. However, for some embryos a clear expression in the HE was detectable (Figure 5-2 F). Similarly, a previous study detected *gfi1ab* expression in the HE only in *gfi1aa*$^{-/-}$ mutants but not in Wt embryos (Thambyrajah et al., 2016). The *irf1b* and *pik3cd* genes showed restricted expression in the HE (Figure 5-2 G-H).

For the other 5 high-confidence Runx1 target genes expression in the HE was not detectable by ISH (Figure 5-2 I-M). For the *mpl* gene, no expression was detectable within the whole embryo (Figure 5-2 I). The *pbrm1l*, *pik3cd* and *wdhd1* genes showed strong expression in the head region in addition to a relatively broad expression in the trunk (Figure 5-2 J-K, M). The *syk* gene showed some expression in the region of the PBI (Figure 5-2 L).

ISH analysis for the additional 16 potential Runx1 target genes did not show specific HE expression for any of the analysed genes (Figure 5-3). This was also true for the remaining members of the *dnmt3* family (*dnmt3aa* and *dnmt3bb.3*) (Figure 5-3 B-C) as well as for the *dnmt1* gene and *uhrf* (Figure 5-3 A, D), which encodes an important interaction partner of Dnmt1 (Liu et al., 2013). The *mknk2b*, *prkab1a* and *xbp1* genes showed a more ubiquitous expression already in embryos at 25/27 hpf and have not been further analysed at later time points (Figure 5-3 I, J, N). The *itga4* gene showed expression in the trunk and tail region where the major vessels are located (Figure 5-3 G), which made it impractical to assess the presence or absence of its expression in the HE by whole mount ISH.

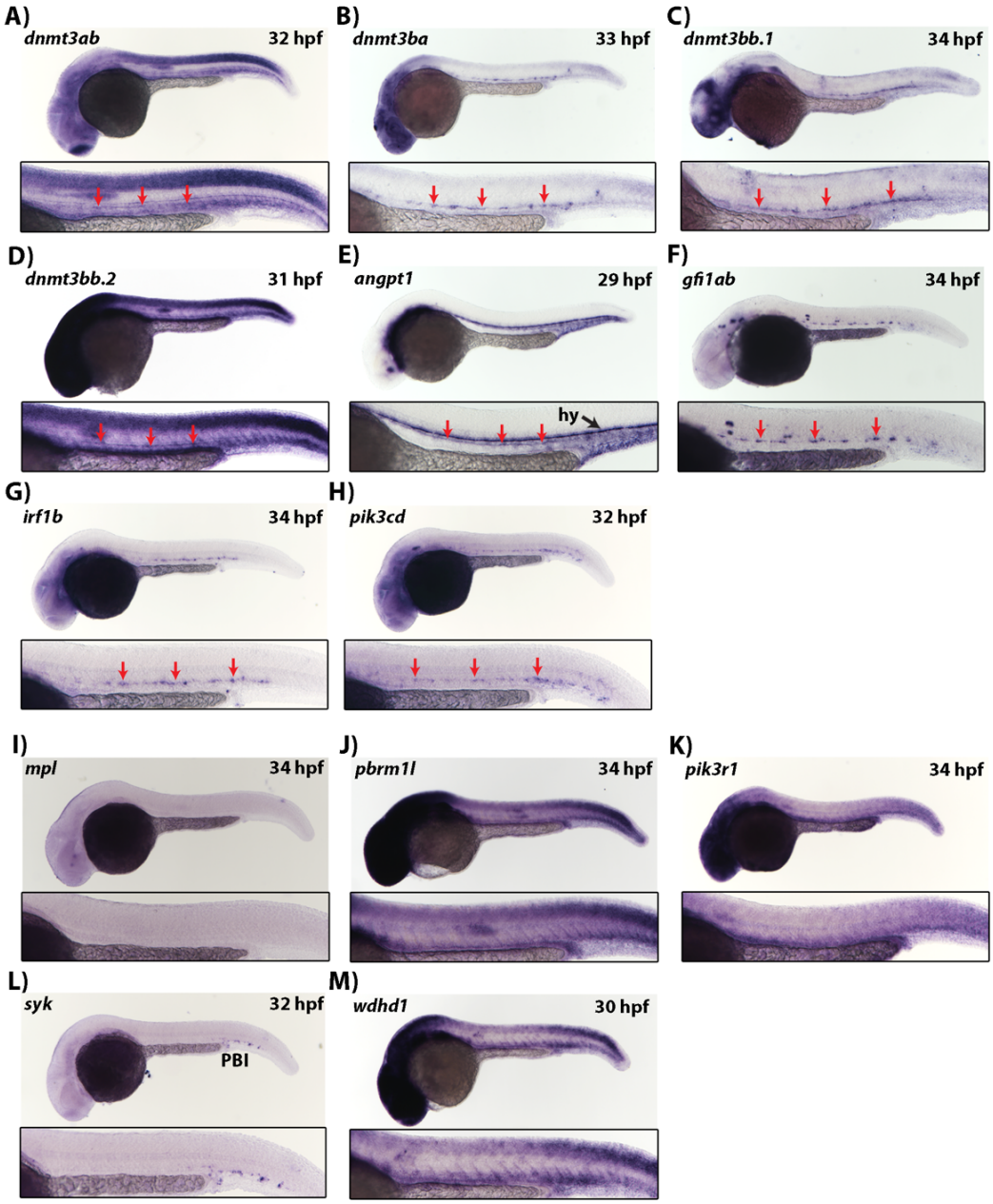


Figure 5-2: ISH based expression screen for high-confidence potential Runx1 target genes.

ISH probes were generated for 13 genes that are high-confidence Runx1 targets in the zebrafish HE. A-H) Genes with expression in the HE (red arrow). I-M) Genes for which expression in the HE was not detactable by ISH. Hy: hypochord; PBI: posterior blood island region. Experiments were performed at multiple time points and the best representative examples are shown.

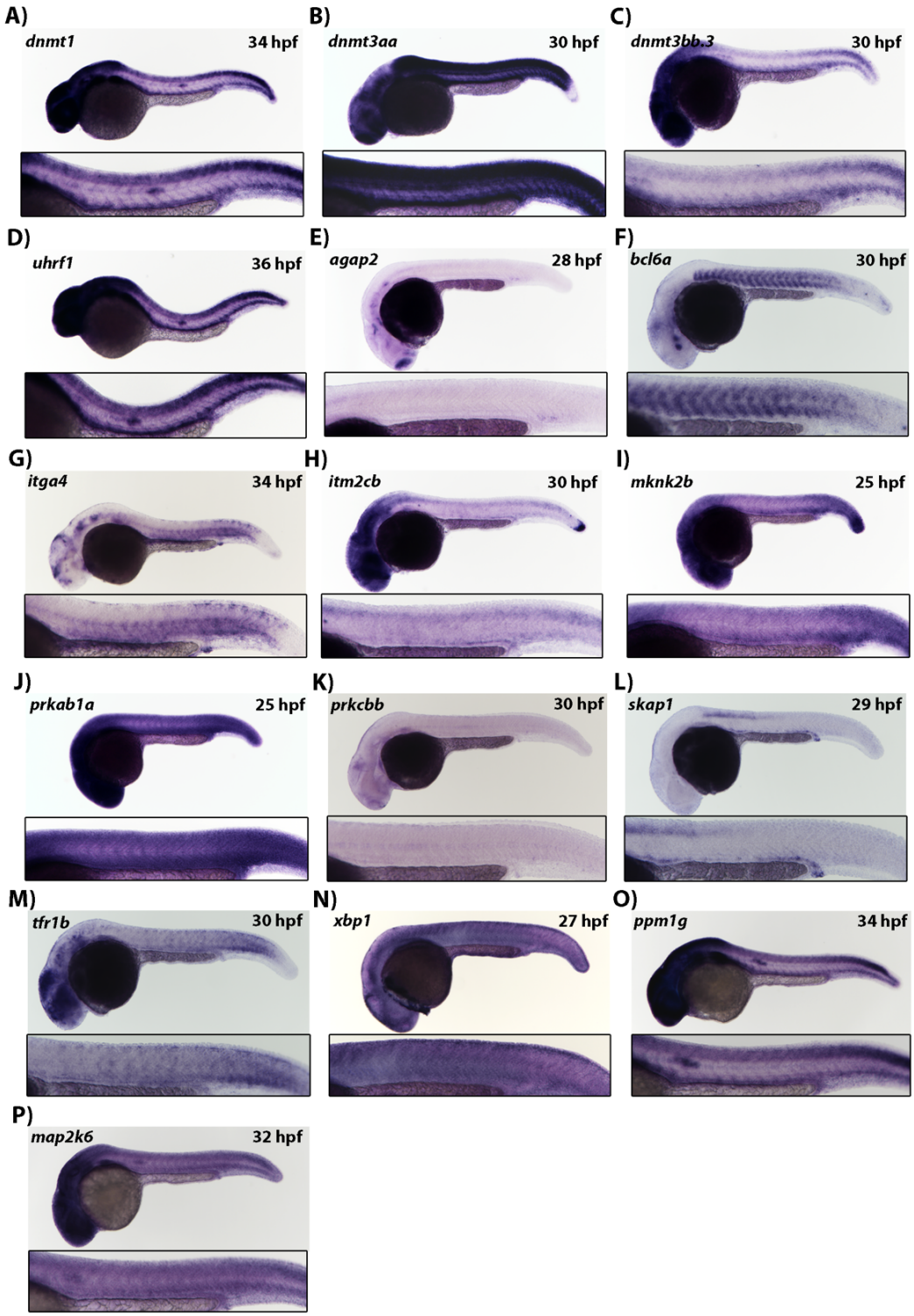


Figure 5-3: ISH based expression screen for additional potential Runx1 target genes.

ISH probes were generated for additional 16 genes that might be regulated by Runx1 in the zebrafish HE. Expression in the HE could not be detected for any of the additonal 16 genes analysed by ISH. Experiments were performed at multiple time points and the best representative examples are shown.

5.2.3 Set-up of a qPCR based validation screen

5.2.3.1 Design of the BioMark-based expression screen

A second validation experiment based on qPCR analysis was designed to complement the RNAseq and ISH experiments. This screen included three different time points (24/25 hpf, 28/29 hpf and 33/34 hpf) to address expression dynamics in the HE (Figure 5-4 A).

For that purpose, the *kdrl:mCherry* and the *runx1P2:Citrine* reporter were both crossed onto a homozygous *runx1*$^{W84X/W84X}$ background (Figure 5-4 B). This was possible because in zebrafish some of the homozygous *runx1*$^{W84X/W84X}$ embryos recover from a larval "bloodless" phase and develop to fertile adults with multi-lineage haematopoiesis (Sood et al., 2010). Cells of double transgenic embryos that were either derived from Wt or homozygous *runx1*$^{W84X/W84X}$ mutant embryos were isolated by FACS. Cells of the mCherry^{+}, ++LOW and ++HIGH gates were sorted from Wt embryos and cells of the ++HIGH gate were sorted from *runx1*$^{W84X/W84X}$ mutants. For each sample, 100 cells were directly sorted into cDNA library buffer and analysed by a TaqMan probe based qPCR assay on the Fluidigm BioMark platform.

We included a high number of control genes (Table 5-3); 10 genes with known roles in embryonic haematopoiesis plus 4 endothelial marker genes. Furthermore, 5 genes that were identified as potential *runx1*-MO off-targets were also incorporated. Finally, we screened for 11 high confidence Runx1 target genes (TaqMan probes were not available for all genes) and 9 additional potential target genes (Table 5-3). These included 3 genes that were identified as being Runx1-repressed. Of note, *gata2a*, which we included as one of the haematopoietic control genes was also identified as Runx1-repressed.

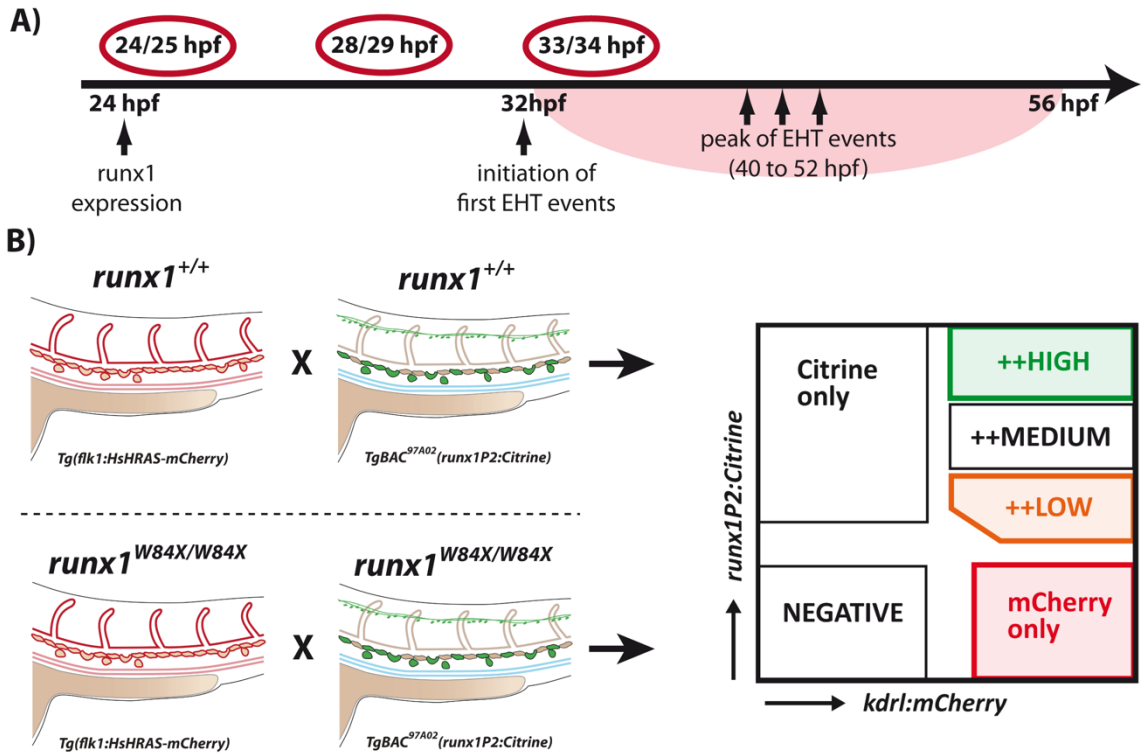


Figure 5-4: Setup of the Biomarkbased expression screen.

A) Timeline of the most important events during definitive haematopoiesis in the zebrafish HE. Three different time points have been analysed (red circles). B) Double-transgenic offspring of reporter line incrosses on a *runx1*$^{+/+}$ or *runx1*$^{W84X/W84X}$ mutant background were used for FAC-sorting of the indicated cell populations (colored). An N=5 experiments were performed for Wt embryos and N=6 for mutants.

Gene expression levels were analysed in relation to the geometric mean of the two housekeeping genes *eef1a1l* and *rpl13a*. Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and the ++HIGH (Wt and W84X) expression data. Detected gene expression levels in the mCherry^{+} fraction provide additional insights into the "general" endothelial population.

Table 5-3: List of genes that were analysed using the BioMark-based expression screen.

Haematopoietic genes (10)	*cbfb*, *cmyb*, *gata2a*, *gata2b*, *gfi1aa*, *jag1a*, *runx1*, *tal1*	Figure 5-5
	itga2b (*CD41)*, *ifng1-2*	Not detected
Endothelial genes (4)	*dll4*, *efnb2a*, *ephb4b*	Figure 5-19
	tbx20	Figure 5-5
Runx1-MO off-target genes (5)	**Up-regulated (Runx1-repressed):** *tp53*, *mdm2*, *foxc3b* **Down-regulated (Runx1-activated):** *rpl22l1*, *pbx2*	Figure 5-7
High-confidence potential target genes (11)	**Down-regulated (Runx1-activated):** *angpt1*, *gfi1ab*, *irf1b*, *pik3cd*,	Figure 5-11
	dnmt3ab, *dnmt3ba*, *dnmt3bb.1*, *dnmt3bb.2*,	Figure 5-8
	mpl, *syk*, *wdhd1*	Figure 5-12
Additional potential target genes (9)	**Up-regulated (Runx1-repressed):** *abcg2a,*	Appendix
	sox17, *sox32*	Figure 5-20
	Down-regulated (Runx1-activated): *dnmt1*, *dnmt3aa*, *uhrf1*,	Figure 5-8
	map2k6, ppm1g, *prkcbb*	Appendix
House-keeping genes (2)	*eef1a1l*, *rpl13a*	/

5.2.3.2 qPCR expression analysis of haematopoietic and endothelial genes

To validate the general experimental set-up, a panel of 9 widely used marker genes were analysed (Figure 5-5). This panel included the HE markers *tal1*, *cmyb* and *gfi1aa* and the DA roof marker *tbx20* (Figure 5-5 A-D), all of which were also used in the initial qPCR analysis (Chapter 3 Figure 3-14.). This panel also included *runx1* and its binding-partner *cbfb*, as well as *gata2b*, *gata2a* and *jag1a* (Figure 5-5 E-I). Expression of another 3 endothelial marker genes (*dll4*, *efnb2a,* and *ephb4b*) will be presented and discussed in a different context further below (Figure 5-19). Of note, expression of *itga2b* and *ifng1-2* could not be detected by qPCR for any time point.

The analysis of *tal1*, *cmyb*, *gfi1aa and tbx20* expression (Figure 5-5 A-D) recapitulated the expression patterns detected by the initial qPCR analysis performed in Chapter 3 (Figure 3-14), validating the reliability of the here performed protocol. Furthermore, *gfi1aa* expression was not affected by the loss of Runx1, confirming a previous report (Thambyrajah et al., 2016).

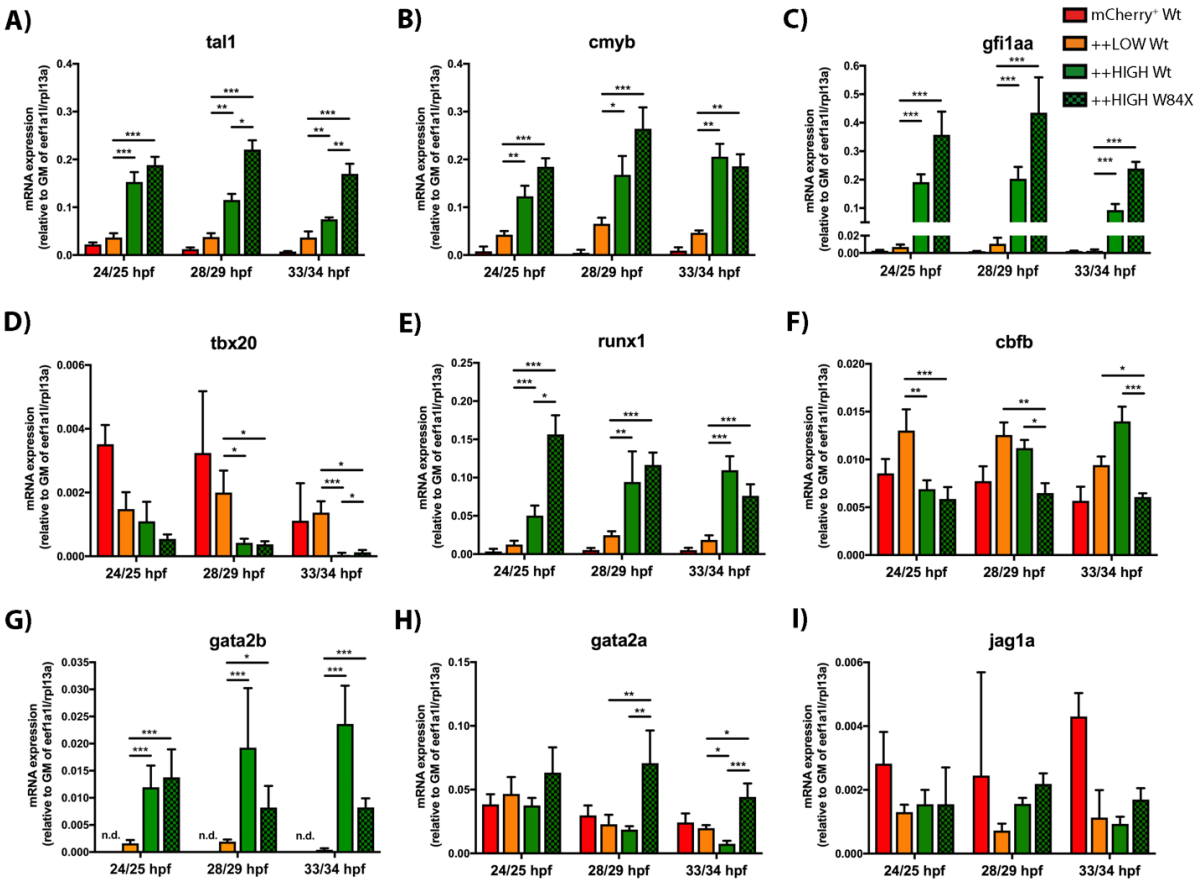


Figure 5-5: Gene expression analysis of haematopoietic and endothelial control genes.

qPCR analysis was performed using TaqMan probes on the BioMark platform (Fluidigm). Expression levels detected within the indicated cell populations sorted from *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos are presented in relation to the geometric mean (GM) of the two housekeeping genes *eefia1l* and *rpll13a*. Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and ++HIGH (Wt and W84X) populations for each independent time point: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean. An N=5 experiments were performed for Wt embryos and N=6 for mutants.

Interestingly, *tal1* expression decreased over time in the HE, but stayed up in *runx1*$^{W84X/W84X}$ mutants (Figure 5-5 A). Since *tal1* was shown to be an upstream regulator of *runx1* (Zhen et al., 2013), this finding suggests a negative regulatory loop that adjusts the expression levels of *runx1* by repression of its activator.

Unexpectedly, the HE-specific expression of *cmyb* was not reduced in *runx1*$^{W84X/W84X}$ mutants (Figure 5-5 B), which contradicts the described phenotype (Jin et al., 2009). We speculate that this

might be a timing issue and that Runx1 is not necessary for the initiation but for the maintenance of *cmyb* expression. This finding deserves a more detailed analysis in the future.

R*unx1* expression was found to be specific for the ++HIGH population and levels increased over time, thus again demonstrating the reliability of the *TgBAC(runx1P2:Citrine)* reporter line (Figure 5-5 E). Surprisingly, *runx1* expression was increased in the HE of *runx1*$^{W84X/W84X}$ mutants at around 24/25 hpf, suggesting an auto-regulatory loop that might be required to tightly regulate the timing of *runx1* up-regulation.

The *cbfb* gene, which encodes the Runx1 interaction partner, is thought to be ubiquitously expressed (Ogawa et al., 1993; Wang et al., 1993). However, the *cbfb* expression levels were initially lower in the ++HIGH population at 24/25 hpf and then increased over time in a Runx1-dependent manner (Figure 5-5 F), showing that the regulation of *cbfb* expression involves dynamic changes.

The *gata2b* gene encodes an upstream regulator of *runx1* with an HE restricted expression pattern (Butko et al., 2015). Indeed, *gata2b* expression was restricted to the ++HIGH population (Figure 5-5 G). From 24/25 to 33/34 hpf, *gata2b* expression in the HE increased up to ~2-fold, indicating an additional later requirement for *gata2b* in HSCs. In mouse, it was shown that GATA2 is essential for HSC survival (de Pater et al., 2013). Interestingly, the increased expression of *gata2b* seemed to be dependent on Runx1 (even though it was not found to be statistically significant, the reduction of *gata2b* in *runx1*$^{W84X/W84X}$ mutants was robust).

In contrast, expression of *gata2a* became depleted from the HE over time (Figure 5-5 H) and was up-regulated in the HE of *runx1*$^{W84X/W84X}$ mutants from 28/29 hpf onwards. To verify if the detected up-regulation was indeed specific to the ++HIGH population, an additional qPCR analysis was performed. Here, we used *runx1*-MO samples from all three ++LOW, ++MEDIUM and ++HIGH populations. A significant up-regulation was only detectable for the ++HIGH population (Figure 5-6 A). ISH analysis for *gata2a* on Wt and *runx1*$^{W84X/W84X}$ mutants embryos further validated these findings (Figure 5-6 B).

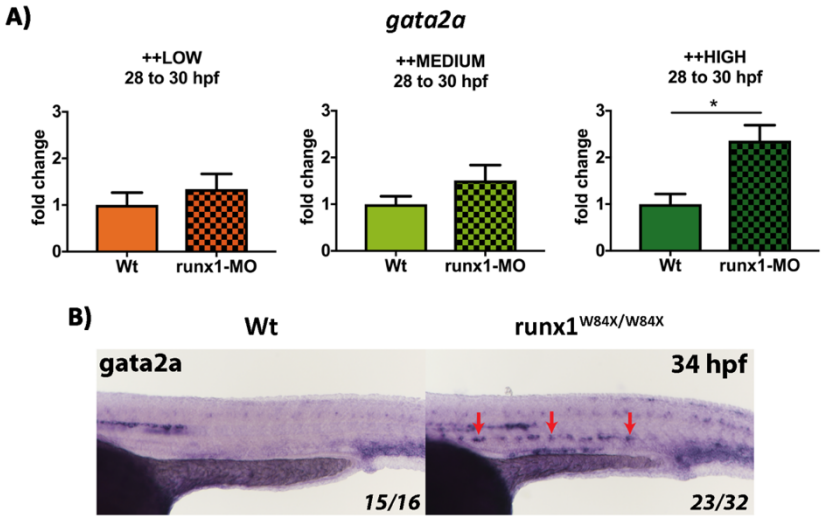


Figure 5-6: Expression of *gata2a* in the HE is repressed by Runx1.

A) qPCR analysis with left-over RNA from the RNAseq experiment. *Gata2a* expression in runx1 morphants is up-regulated specifically in the ++HIGH population (T-test). B) ISH reveals gata2a expression in the HE of *runx1*$^{W84X/W84X}$ mutants. Statistical analysis was performed with the student's t-test: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean.

Lastly, we also checked for the expression levels of *jag1a*, which we and others have previously published to be required for the programming of the HE (Espín-Palazón et al., 2014; Monteiro et al., 2016). Here, we could not detect any significant differences in the expression levels within the DA populations (++LOW and ++HIGH) (Figure 5-5 I).

5.2.3.3 Analysis of potential *runx1* MO off-targets in *runx1* mutants

Next, we aimed to validate our strategy for the identification of *runx1* MO off-targets. For that purpose, we selected 5 genes of which 3 were up-regulated and 2 were down-regulated in *runx1*-morphants in both the NEGATIVE and the ++HIGH population (Table 5-4).

Tp53 and *mdm2* showed a strong up-regulation in the MO-based RNAseq experiments. However, no real expression changes were detected in *runx1*$^{W84X/W84X}$ mutants (Figure 5-7 A-B). This agrees with findings from adult mouse HSCs where expression of *Tp53* was also not affected in *Runx1* mutants (Cai et al., 2011).

In the case of the *foxo3b* gene, we detected a strong up-regulation (up to 9-fold) in the *runx1*$^{W84X/W84X}$ mutants, even though it was not statistically significant (Figure 5-7 C). Hence, it is still possible that *foxo3b* is truly repressed by Runx1 in the HE.

Table 5-4: List of runx1-MO off-targets genes anaysed.
The table shows the detected fold changes within the respective populations (NEGATIVE and ++HIGH) of the performed RNAseq experiment.

	Gene name	**Fold change (MO over Wt) NEGATIVE**	**Fold change (MO over Wt) ++HIGH**
Up-regulated	*tp53*	12.1	5.1
	mdm2	11.2	3.4
	foxo3b	3.1	1.9
Down-regulated	*rpl22l1*	-4.3	-3.1
	pbx2	-2.6	-2.6

The two down-regulated off-target genes (*rpl22l1* and *pbx2*) were both relatively unaffected in the *runx1*$^{W84X/W84X}$ mutants (Figure 5-7 D-E). A slight reduction in expression was detectable only for the latest time point at around 33/34 hpf. However, the fold-changes we detected were minor compared to the fold-changes in the *runx1*-morphants (Table 5-4). Hence, we conclude that it is still legitimate to mark these genes as off-targets. Overall, this analysis proved that our strategy for the identification of potential off-targets in our MO-based RNAseq analysis was reliable.

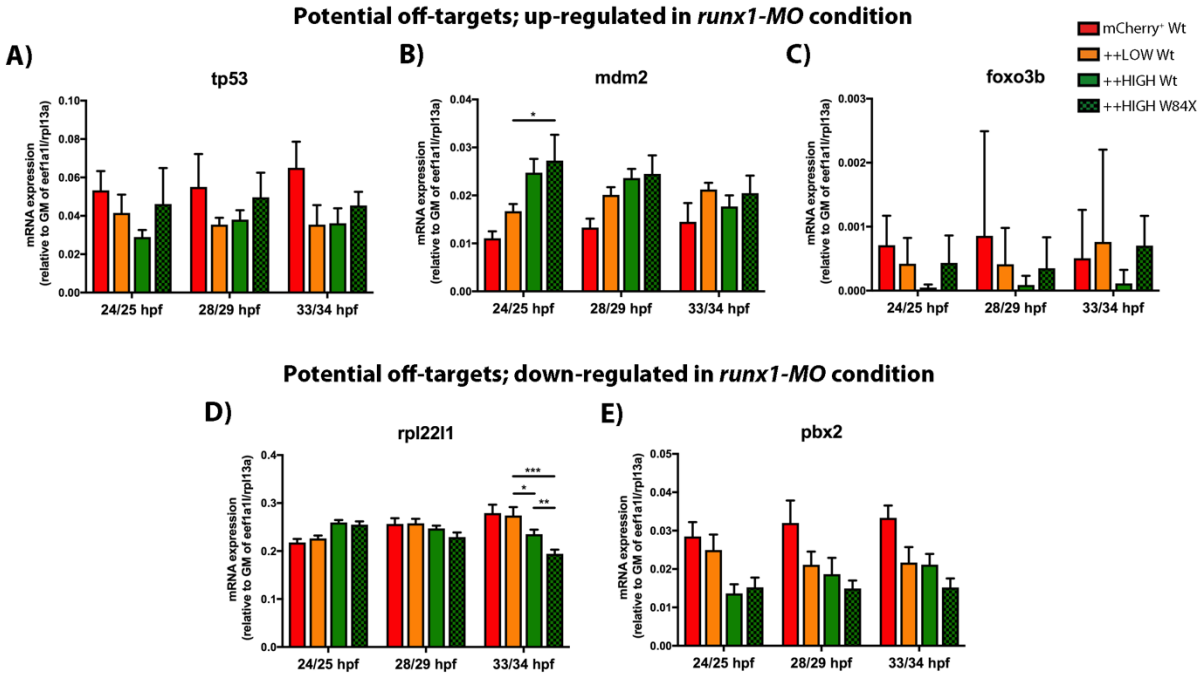


Figure 5-7: Gene expression analysis of potential *runx1* MO off-target genes.

qPCR analysis was performed using TaqMan probes on the BioMark platform (Fluidigm). Expression levels detected within the indicated cell populations sorted from *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos are presented in relation to the geometric mean (GM) of the two housekeeping genes *eefia1l* and *rpll13a*. Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and ++HIGH (Wt and W84X) populations for each independent time point: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean. An N=5 experiments were performed for Wt embryos and N=6 for mutants.

5.2.4 Validation of high-confidence potential Runx1 targets by qPCR

Here, we present the gene expression data for genes classified as high-confidence Runx1 target genes. These genes will be presented in three groups:

- Members of the *dnmt* family (*dnmt3ab*, *dnmt3ba*, *dnmt3bb.1* and *dnmt3bb.2*) including the three additional genes *dnmt3aa*, *dnmt1* and *uhrf1* (Figure 5-8).
- Genes expressed in the HE and validated in the ISH screen (*angpt1*, *gfi1ab*, *pik3cd* and *irf1b*) (Figure 5-11).
- Genes expressed in the HE but not validated in the ISH screen (*wdhd1*, *mpl* and *syk*) (Figure 5-15).

In addition, the expression data of four additional target genes (*abcg2a*, *map2k6*, *ppm1g* and *prkcbb*) are shown in the Appendix 10. Here, *abcg2a* was validated to be truly repressed by Runx1.

5.2.4.1 Expression dynamics of *dnmt* family members

The presence of most members of the zebrafish *de novo* DNA methyltransferase family in our high-confidence target list was striking. The ISH screen confirmed the HE-specific expressions not only for the already published *dnmt3bb.1* gene (Gore et al., 2016), but also for the additional isoforms *dnmt3ab, dnmt3ba* and *dnmt3bb.2* (Figure 5-2 A-D). Here, the more sensitive qPCR expression analysis also showed that all five *de novo dnmt3* genes analysed had higher expression in the ++HIGH population compared to the ++LOW population (Figure 5-8 A-E). Furthermore, expression of all five *dnmt3* isoforms was significantly reduced in the HE of *runx1*$^{W84X/W84X}$ mutants compared to the controls (Figure 5-8 A-E). ISH analysis for *dnmt3ba* and *dnmt3bb.1* on Wt embryos and *runx1*$^{W84X/W84X}$ mutants further validated that expression of these genes is Runx1-dependent (Figure 5-9 A-B).

We also checked for expression of the maintenance DNA methyl-transferase *dnmt1* and its partner *uhrf1*. Neither gene was effected by the loss of Runx1 in the ++HIGH population (Figure 5-8 F-G). This indicates that potential changes in global DNA methylation during haematopoiesis are introduced via the up-regulation of *de novo* DNA methyltransferase under the control of Runx1.

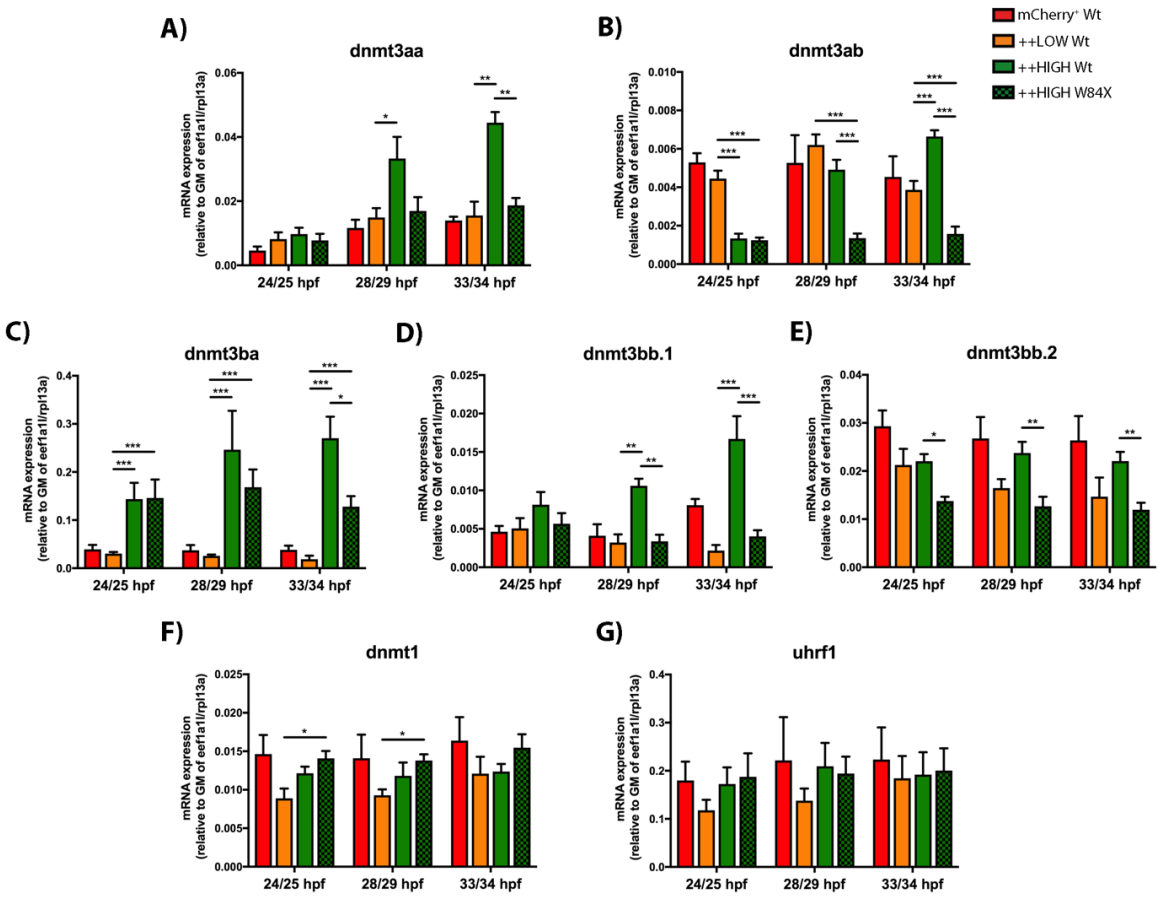


Figure 5-8: Gene expression analysis of DNA methyl-transferases.

qPCR analysis was performed using TaqMan probes on the BioMark platform (Fluidigm). Expression levels detected within the indicated cell populations sorted from *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos are presented in relation to the geometric mean (GM) of the two housekeeping genes *eefia1l* and *rpll13a*. Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and ++HIGH (Wt and W84X) populations for each independent time point: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean. An N=5 experiments were performed for Wt embryos and N=6 for mutants.

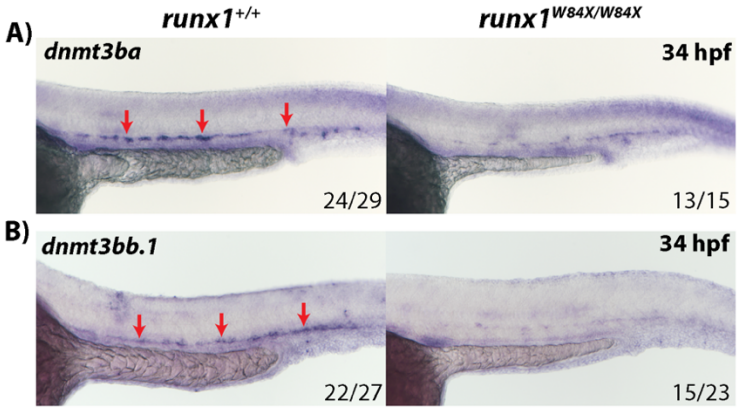


Figure 5-9: Haemogenic expression of *dnmt3* isoforms is dependent on Runx1.
ISH reveals that expression of *dnmt3ba* (A) and *dnmt3bb*.1 (B) are down-regulated in the HE of *runx1*$^{W84X/W84X}$ mutants.

We wondered if expression of the *dnmt* genes might persist into later stages of haematopoiesis. Therefore, we analysed their expression in the CHT region of 2.5 dpf embryos, where HSPCs are located. Expression in the CHT was detected for *dnmt1*, *dnmt3ba* and *dnmt3bb.1* (Figure 5-10 A, C-D). Weak expression of *dnmt3ab* was detected across the whole embryo likely including expression in HSPCs (Figure 5-10 B). Expression of *dnmt3bb.2* seems to be down-regulated in HSPCs (Figure 5-10 E)



Figure 5-10: Analysis of *dnmt* genes for expression in the CHT region.
ISH reveals expression of *dnmt1* (A), *dnmt3ba* (C) and *dnmt3bb.1* (D) within the CHT, presumably in HSPCs.

5.2.4.2 Expression dynamics of high-confidence Runx1 target genes

We analysed the expression dynamics of the 4 genes (*angpt1*, *gfi1ab*, *pik3cd* and *irf1b*), for which we detected expression in the HE by ISH (Figure 5-2 E-H). Here, we also detected enriched expression in the ++HIGH over the ++LOW population for all 4 genes at 33/34 hpf (Figure 5-11). Interestingly, *gfi1ab* was already up-regulated in the ++HIGH fraction by 24/25 hpf (Figure 5-11 B). All 4 genes were down-regulated in the ++HIGH population of *runx1*$^{W84X/W84X}$ mutants. This reduction was statistically significant for all the genes except for *irf1b*, even though the expression changes were robust. In addition, *angpt1* showed high expression in the mCherry^{+} general endothelium (Figure 5-11 A). ISH validated that expression of *angpt1* and *irf1b* in *runx1*$^{W84X/W84X}$ mutants was severely decreased (Figure 5-12). The *gfi1ab* gene expression will be analysed in detail below.

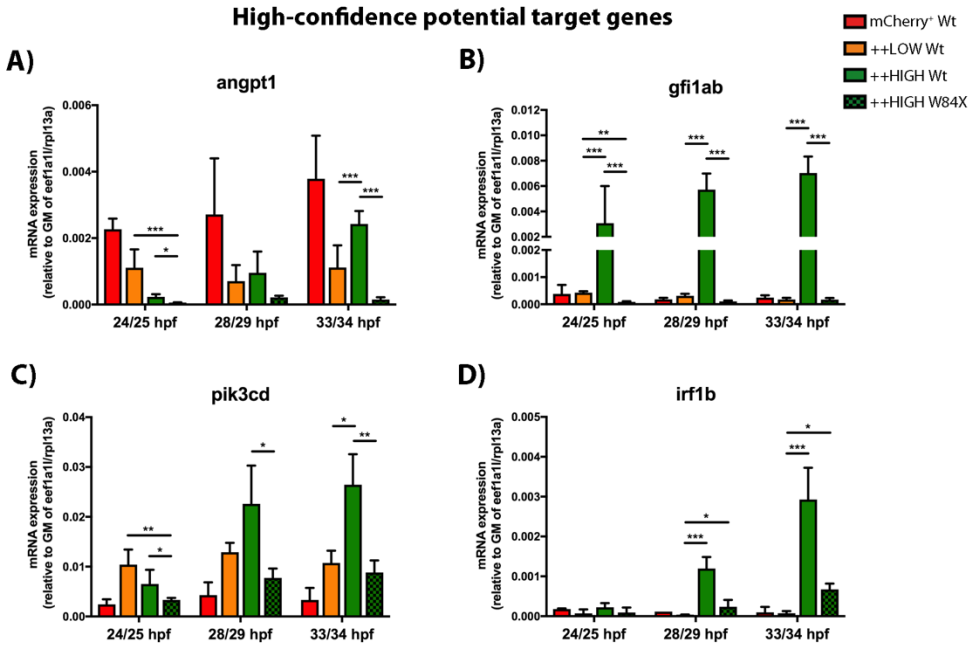


Figure 5-11: Gene expression analysis of four high-confidence potential Runx1 target genes identifies *angpt1*, *gfi1ab*, *pik3cd* and *irf1b* as bona fide Runx1 targets.

qPCR analysis was performed using TaqMan probes on the BioMark platform (Fluidigm). Expression levels detected within the indicated cell populations sorted from *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos are presented in relation to the geometric mean (GM) of the two housekeeping genes *eefia1l* and *rpll13a*. Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and ++HIGH (Wt and W84X) populations for each independent time point: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean. An N=5 experiments were performed for Wt embryos and N=6 for mutants.

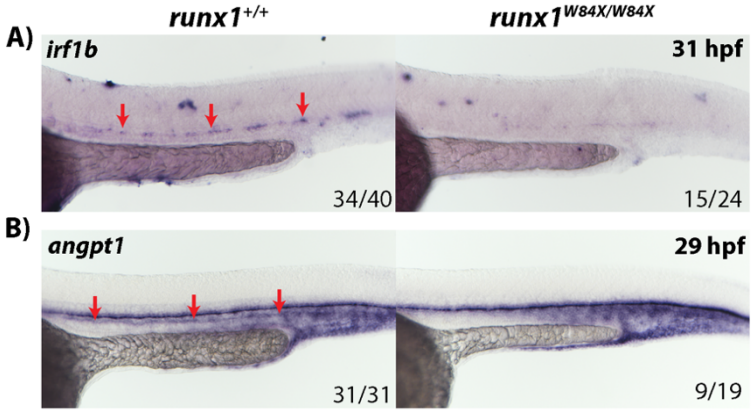


Figure 5-12: Haemogenic expression of *irf1b* and *angpt1* is dependent on Runx1.
ISH reveals that expression of *irf1b* (A) and *angpt1* (B) are down-regulated in the HE of *runx1*$^{W84X/W84X}$ mutants.

Detailed ISH analysis of *irf1b, pik3cd* and *angpt1* expression (Figure 5-13) showed similar expression dynamics as seen by qPCR (Figure 5-11 A, C-D). Here, *irf1b* expression in the HE was detected from 30 hpf onwards (Figure 5-13A-F), whereas expression of *pik3cd* (Figure 5-13H-M) and *angpt1* (Figure 5-13 O-R) in the HE were detected from 28 hpf onwards. In 2.5 dpf embryos, weak *irf1b* and *pik3cd* was also found in cells located in the CHT (Figure 5-13 G, N).

Expression analysis of *angpt1* in the CHT was not conclusive (data not shown) due to its additional expression in other endothelial cells (Figure 5-11 A) and its expression within the vascular bed of the PBI/CHT region (Figure 5-13 R). However, it has been reported that murine HSCs indeed express *Angpt1* (Takakura et al., 2000; Zhou et al., 2015).

Angpt1 encodes a growth factor that binds to the TEK (TIE2) receptor (Fukuhara et al., 2010). Interestingly, expression of *tek,* as assessed by RNAseq, was depleted from the ++HIGH population, but strongly expressed in ++LOW cells (Figure 5-13 S), and is thus complementary to *angpt1* expression.

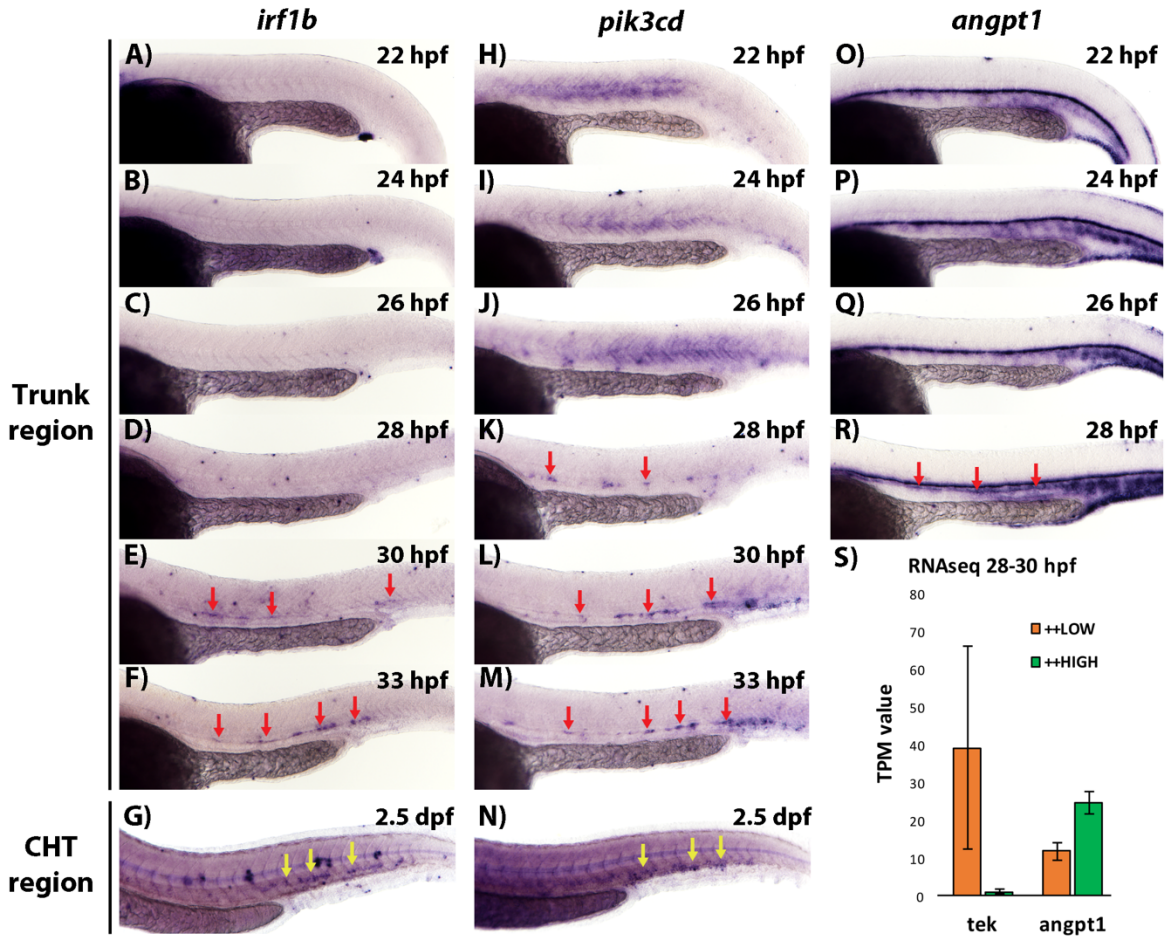


Figure 5-13: Deatailed expression dynamics of *irf1b*, *pik3cd* and *angpt1* within the HE by ISH.
A-F) ISH reveals that expression of *irf1b* in the HE becomes detectable from 30 hpf onwards. G) Weak expression of *irf1b* is also detectable within the CHT. H-M) ISH reveals that expression of *pik3cd* in the HE beomes detectable from 28-30 hpf onwards. N) Weak expression of *pik3cd* is also detectable within the CHT. O-R) ISH reveals that expression of angpt1 in the HE becomes detctable from 28 hpf onwards. S) RNAseq data show that expression of *angpt1* is higher in the ++HIGH whereas expression of its receptor tek is specific to the ++LOW and depleted from the ++HIGH. Error-bars represent the standard deviation.

To get further insights into the significance of the identified *irf1b* and *pik3cd* genes, we searched for expression of additional gene family members in the RNAseq data (Figure 5-14). The *irf* gene family is represented by 12 annotated family members for 11 of which we detected expression (Figure 5-14 A). Here, *irf1b* was the only highly expressed gene in the ++HIGH population that was also dependent on Runx1 (Figure 5-14 A; red box). Expression of *irf2* was up-regulated in the ++HIGH fraction but independent of Runx1 (Figure 5-14 A). In agreement with these findings, Irf2 was found to negatively regulate expression of *runx1* in the zebrafish HE and is thus placed upstream of *runx1* (Li et al., 2014). Of note, expression of *irf4a* (Figure 5-14 A; orange box) showed similar expression

dynamics as *irf1b*, but was expressed at a much lower level. *Irf4a* was recently found in haematopoietic cells in the CHT and to regulate the choice between T-lymphoid-primed and myeloid lineages (Wang et al., 2015). We also found that expression of *irf8* and *irf9* was repressed by Runx1 (Figure 5-14 A). It has been shown that Runx1 negatively regulates the formation of *irf8*$^{+}$ macrophages (Jin et al., 2012).

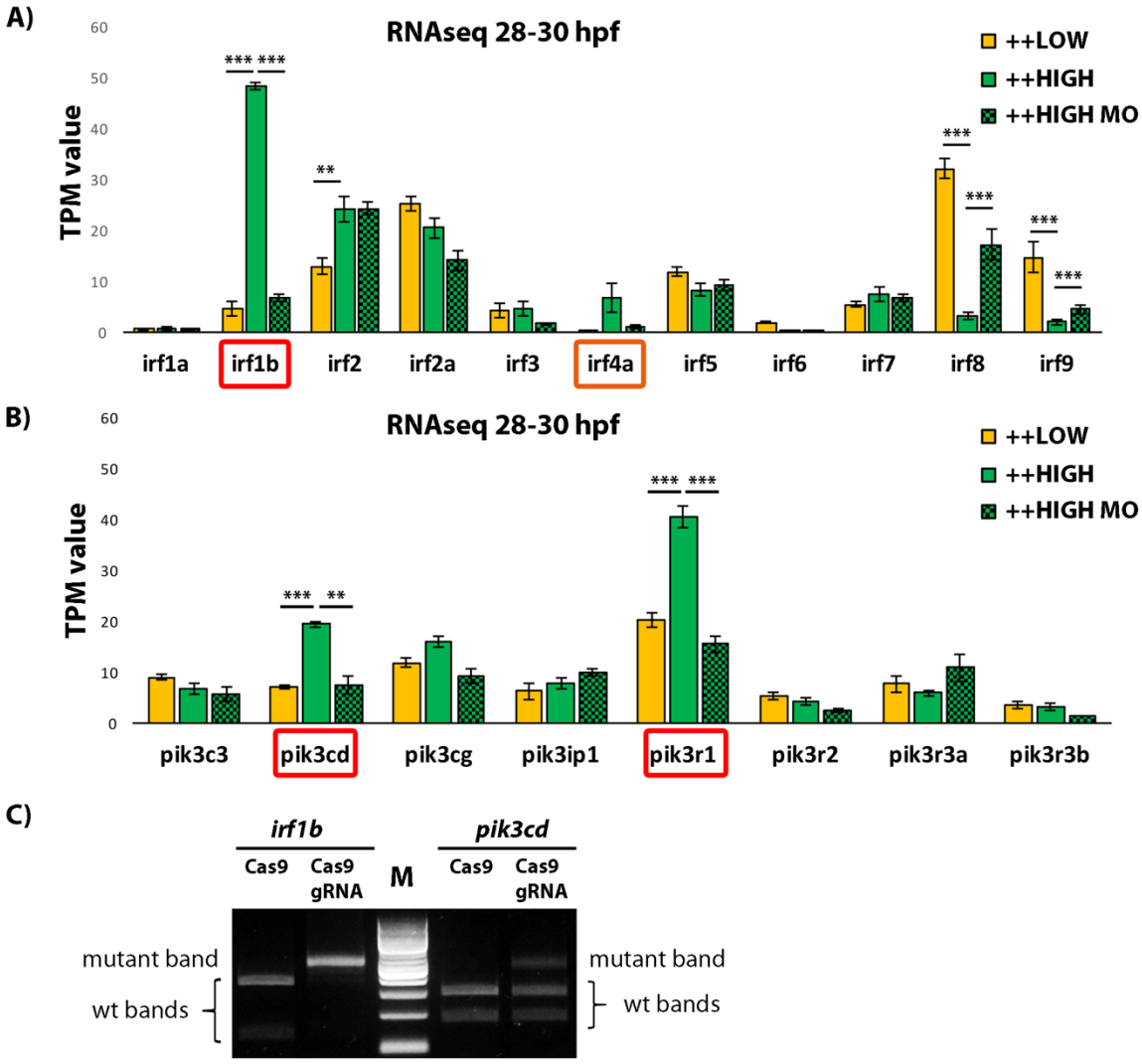


Figure 5-14: *Irf1b* and *pik3cd*/*pik3r1* are specific Runx1 targets within their respective gene families.

A) Of all annotated *irf* genes in zebrafish, only *irf1b* was significantly up-regulated in the ++HIGH population and Runx1-dependent (red box). *Irf4a* showed a similar trend, but was lower in its expression (orange box). B) Of all annotated PI3K class-I genes in zebrafish, only *pik3cd* and its binding partner *pik3r1* were significantly up-regulated in the ++HIGH population and Runx1-dependent (red boxes). * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean. C) Guide RNAs (gRNAs) were designed and tested for the generation of stable mutant lines for the genes *irf1b* and *pik3cd*.

We also found expression of 8 PI3K class-I genes in our RNAseq data-set. Here, a significant higher expression in the ++HIGH population was only detected for *pik3cd* and its interaction partner *pik3r1* (Figure 5-14 B red boxes).

Based on these results we decided to generate mutants for the *irf1b* and *pik3cd* genes using CRISPR/Cas9 technology (Varshney et al., 2015). Therefore, guide RNAs (gRNAs) were designed that target the 5' region of each gene in a location close to a restriction enzyme binding site, to allow for a simple identification of potential mutants by genotyping. gRNAs were tested for their cutting efficiency in transiently injected embryos (Figure 5-14 C). Injected embryos were raised to adulthood and will be screened for stable germ line transmission of potential mutations.

5.2.4.3 Expression dynamics of non-ISH validated high-confidence Runx1 target genes

We also analysed by qPCR expression of the three high-confidence potential target genes *wdhd1*, *mpl* and *syk* for which expression in the HE could not be proven by ISH. Here, *wdhd1* expression across the different endothelial populations was uniform, despite an earlier weaker expression in the ++LOW population at 24/25 hpf (Figure 5-15 A). This gene was not differentially expressed in the *runx1*$^{W84X/W84X}$ mutants.

In contrast, expression of *mpl* and *syk* was significantly up-regulated in ++HIGH cells at 33/34 hpf and dependent on Runx1 (Figure 5-15 B-C). For the *mpl* gene, this was also true when analysing cells at around 28/29 hpf. Of note, *mpl* is expressed by intra aortic cluster (IAC) cells and by the quiescent population of HSCs in the adult bone marrow (Petit-Cocault et al., 2007; Yoshihara et al., 2007). Its potential regulation by Runx1 has not been shown before.

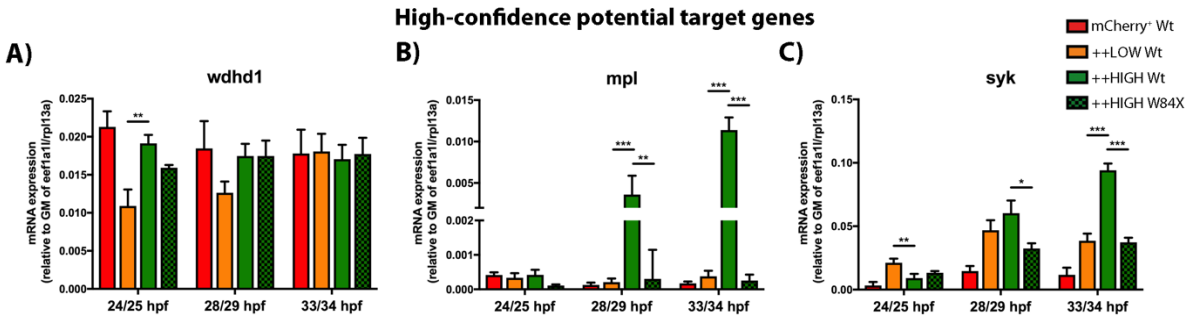


Figure 5-15: Gene expression analysis identifies *mpl* and *syk* as bona fide Runx1 targets.

qPCR analysis was performed using TaqMan probes on the BioMark platform (Fluidigm). Expression levels detected within the indicated cell populations sorted from *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos are presented in relation to the geometric mean (GM) of the two housekeeping genes *eefia1l* and *rpll13a*. Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and ++HIGH (Wt and W84X) populations for each independent time point: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean. An N=5 experiments were performed for Wt embryos and N=6 for mutants.

5.2.5 EHT in zebrafish is dependent on a conserved Runx1-Gfi1-Lsd1 axis

It has been shown that EHT in mouse is dependent on RUNX1-induced expression of *Gfi1* (Lancrin et al., 2012; Thambyrajah et al., 2015). The situation in zebrafish is less clear. *Gfi1aa* is expressed in the HE in a Notch-dependent manner, parallel to *runx1* (Thambyrajah et al., 2016). *Gfi1b* is expressed in primitive erythrocytes (Cooney et al., 2013). By contrast, the *gfi1ab* gene has not been studied in much detail.

We performed ISH for all three *gfi1* isoforms to analyse their expression between 26-36 hpf in the trunk region and at 2.5 dpf in the CHT. As mentioned above, the detection of *gfi1ab* expression by ISH in wild type embryos was mostly unsuccessful (Figure 5-16 A-D). The expression of *gfi1aa* was restricted to the HE (Figure 5-16 E-G; red arrows) but hardly detectable in the CHT, besides a few expressing cells (Figure 5-16 H; yellow arrow). Strong expression *of gfi1b* was detected in primitive erythrocytes at 26 hpf, which diminished over the next couple of hours (Figure 5-16 I-K). *Gfi1b* expression in the HE was not detectable. However, cells in the CHT were again expressing *gfi1b* (Figure 5-16 L), presumably replacing *gfi1aa* expression since a similarly dynamic expression of *Gfi1* and *Gfi1b* has been reported for the mouse (Thambyrajah et al., 2015). There, *Gfi1b* expression was

found in cells generated by the earlier waves of haematopoiesis but absent in the aortic HE, which expressed *Gfi1* instead. After EHT took place, expression of *Gfi1* in the IAC was replaced by *Gfi1b* (Thambyrajah et al., 2015).

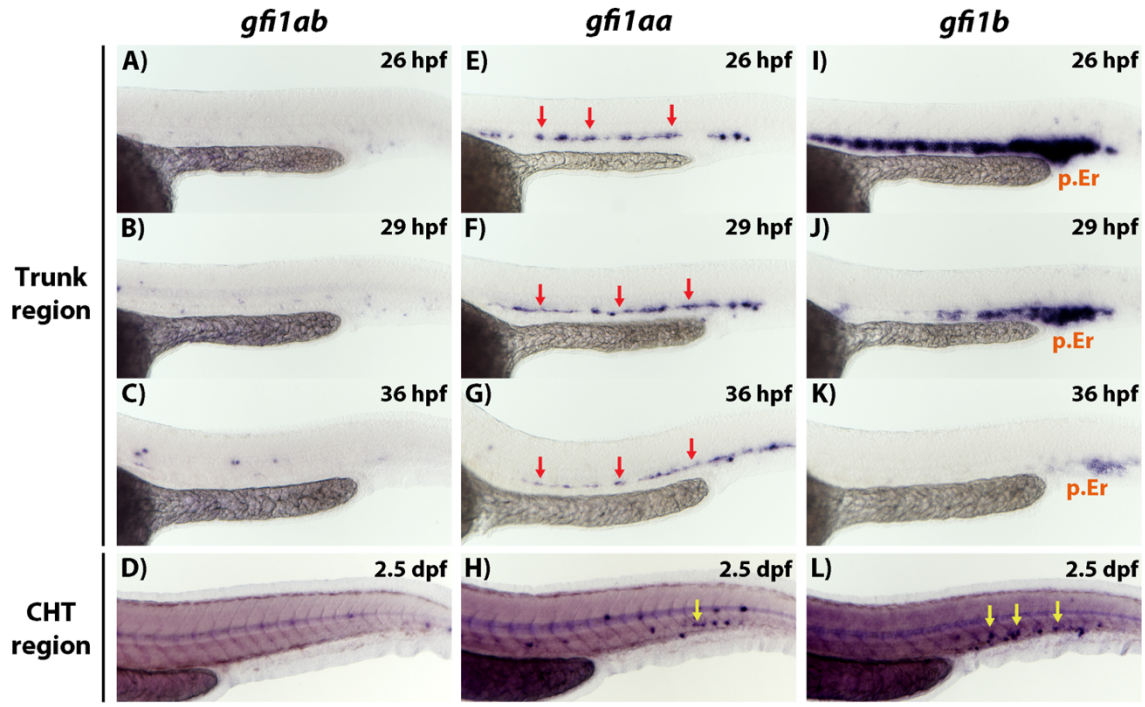


Figure 5-16: Expression analysis of *gfi1* isoforms during definitive haematopoieis in the DA and the CHT.
A-C) Expression of *gfi1ab* was barely detectable in the HE. D) No expression of gfi1ab was detected in the CHT. E-G) Expression of *gfi1aa* was detected specifically in the HE (red arrows). H) Expression of *gfi1aa* in the CHT was detected only in a few cells (yellow arrow). I-K) Expression of *gfi1b* was detected in primitive erythrocytes (p.Er) early on and gradually lost over time. L) Gfi1b is broadly expressed in the CHT.

Since expression of *gfi1ab* in the HE was dramatically increased in homozygous *gfi1aa* mutants (Thambyrajah et al., 2016), we designed a *gfi1aa* MO for functional experiments (Figure 5-17 A). This MO did not cause a loss of *gata1* expression in the primitive erythrocytes (Figure 5-17 B), as it was reported for previous *gfi1aa* MOs (Cooney et al., 2013; Wei et al., 2008), but rather phenocopied the *gfi1aa* mutant phenotype in which *gata1* expression was unaffected (Thambyrajah et al., 2016). We detected a dramatic increase in *gfi1ab* expression in *gfi1aa* morphants (Figure 5-17 C). The earliest expression of *gfi1ab* in *gfi1aa* morphants was detectable at around 24 hpf (Figure 5-17 C), which is when *gfi1ab* expression in ++HIGH cells of Wt embryos is detectable by qPCR (Figure 5-11 B). Of note, expression of *gfi1aa* was detected even earlier at around 22 hpf (Figure 5-17 D).

Finally, we injected the *gfi1aa* MO in *runx1*$^{W84X/W84X}$ embryos and analysed *gfi1ab* expression. *Gfi1aa* morphants showed a dramatic increase in *gfi1ab* expression that was lost in *runx1*$^{W84X/W84X}$ embryos, suggesting that *gfi1ab* expression in the HE is Runx1-dependent (Fig 17 E).

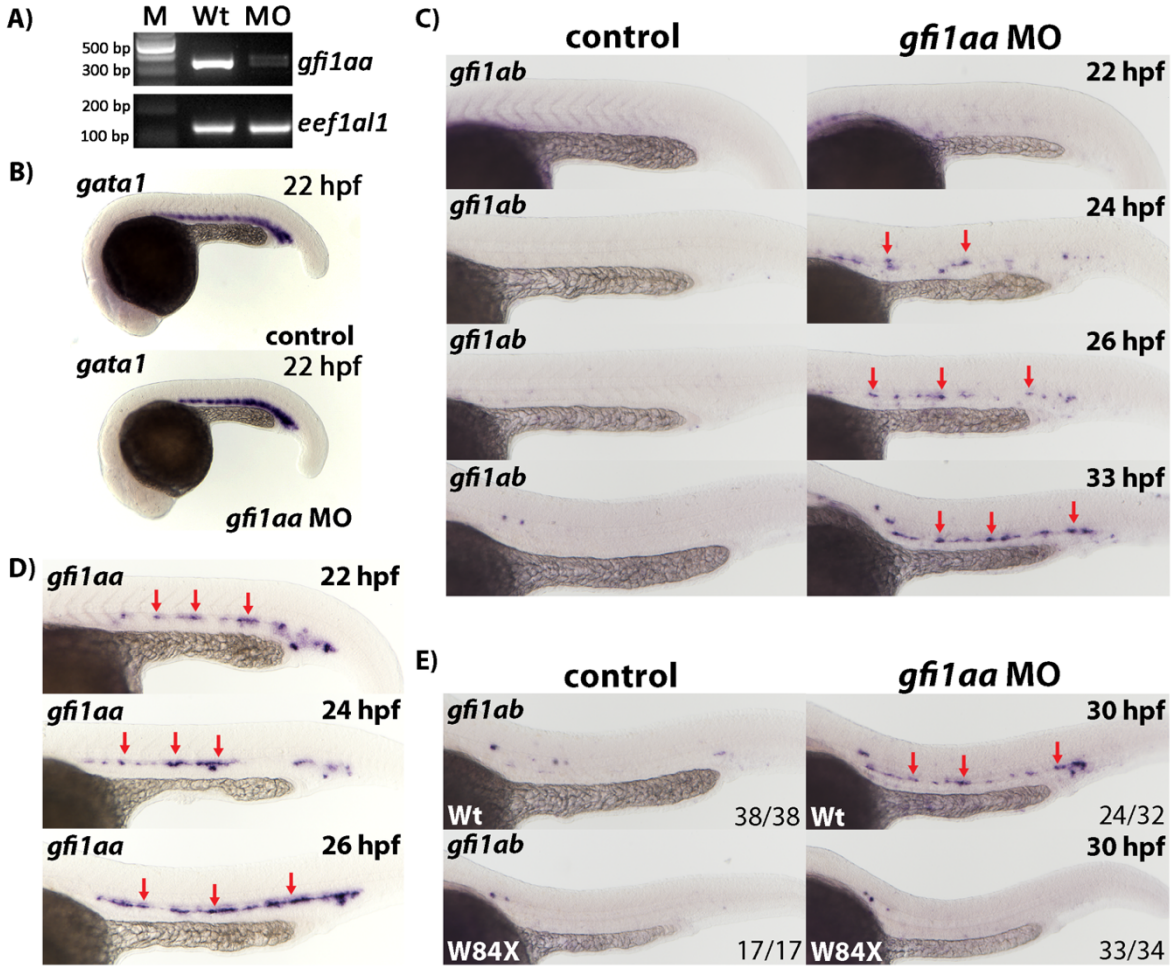


Figure 5-17: *Gfi1ab* expression in the HE can be visualized in *gfi1aa* morphants and is Runx1-dependent.
A) Validation of the splice-blocking effect of a newly designed *gfi1aa* MO. B) The new gfi1aa MO does not lead to a loss *gata1* phenotype phenocopying the reported *gfi1aa* mutant phenotype (Thambyrajah et al., 2016). C) Expression of *gfi1ab* in the HE can be visualized in *gfi1aa* morphants and becomes detectable from 24 hpf onwards. D) Expression of *gfi1aa* in the HE is already detectable at 22 hpf. E) expression of gfi1ab in the HE is lost in *runx1*$^{W84X/W84X}$ mutants.

To analyse the role of *gfi1ab* during definitive haematopoiesis in more detail, we designed a *gfi1ab* MO to perform knock-down experiments. Even though a strong splice blocking effect was detectable when injecting 7 ng of *gfi1ab* MO (Figure 5-18 A), only injections of higher concentrations (14 ng) resulted in a detectable phenotype, showing loss of *cmyb* expression in the CHT (Figure 5-18 B). This

finding suggests that even low levels of Gfi1ab can be functional. We could replicate these findings using a second *gfi1ab* MO (not shown).

It has been shown that Gfi1 proteins recruit the CoREST complex containing Lsd1, which is critical for the function of GFI1 (Figure 5-18 C) (Saleque et al., 2007; Thambyrajah et al., 2015). The interaction between GFI1 and LSD1 proteins can be inhibited with the chemical compound tranylcypromine (TC) (Takeuchi et al., 2015). A concentration of 0.3 mM of TC effects primitive haematopoiesis when administered from 12 to 24 hpf (Takeuchi et al., 2015). We treated embryos with TC only after 24 hpf to avoid early effects. For an easier screening, *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* embryos were treated and scored for Citrine^{+} cells in the CHT or in the thymus at 3 dpf (Figure 5-18 D). Expression of *kdrl:mCherry* indicated that the development of the vasculature was essentially unaffected. TC treated embryos showed reduced numbers of Citrine^{+} cells in both the CHT and the thymus (Figure 5-18 D). The phenotype was more pronounced when embryos were treated with 0.6 mM TC. Additionally, the percentage of embryos revealing a phenotype increased when treated with 0.6 mM TC (Figure 5-18 E), which we defined as the optimal concentration. Higher concentrations of TC (1.2 mM) did not result in a stronger phenotype.

Finally, we probed *rag1* expression in the thymus at 4 dpf, often used as a readout for definitive haematopoiesis. In agreement with the loss of Citrine^{+} cells we also detected a strong reduction in *rag1* staining in TC treated embryos (Figure 5-18 D), suggesting a loss of HSC-delivered T-cell progenitors. Thus we demonstrated that HSC emergence in zebrafish is conserved to mouse and also regulated by a Runx1-Gfi1-Lsd1 axis (Thambyrajah et al., 2015).

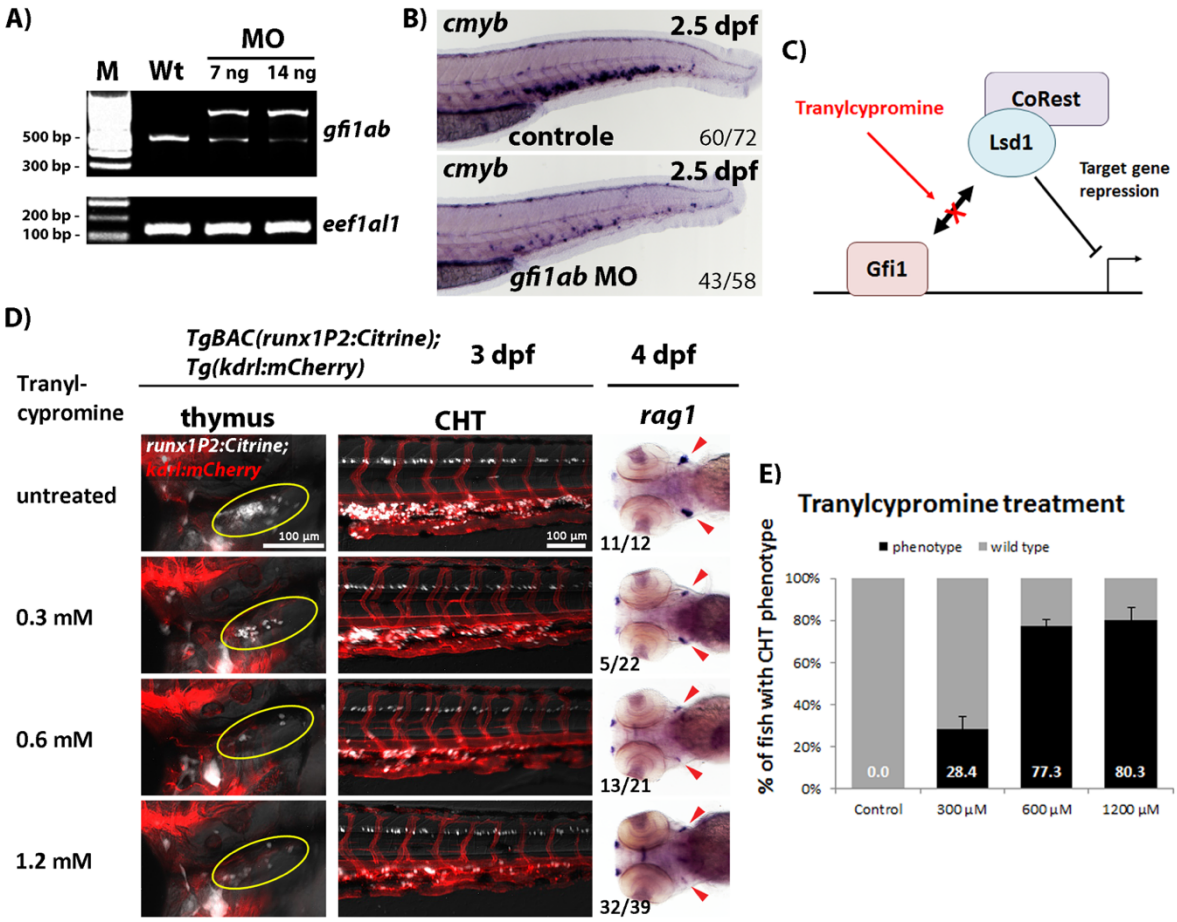


Figure 5-18: EHT in zebrafish is dependent on a conserved Runx1-Gfi1-Lsd1 axis.

A) Confirmation of the splice-blocking effect of a newly designed *gfi1ab* MO. B) *gfi1ab* morphants have a reduced number of *cmyb* expressing cells in the CHT. C) Gfi1 proteins recruit the Lsd1-CoREST complex (Saleque et al., 2007). This interaction can be blocked with the chemical compound tranylcypromine. D) Increasing concentrations of tranylcypromine lead to a reduced number of Citrine+ cells in the CHT and the thymus at 3 dpf and loss of thymic *rag1* expression at 4 dpf. Embryos were treated from 24 hpf onwards. E) The optimal concentration of tranylcypromine was defined as 600 uM. Error-bars represent the standard deviation.

5.2.6 The HE retains an arterial programme in the absence of Runx1

Several studies have pointed out that the endothelial programme is repressed in the mouse HE (Gama-Norton et al., 2015; Swiers et al., 2013b; Thambyrajah et al., 2015). Likewise, we found that the zebrafish HE at 28-30 hpf has already lost part of its endothelial signature (Chapter 4 Figure 4-8 D). To study the dynamics of endothelial gene repression in more detail, we also included the marker genes *dll4*, *efnb2a* and *ephb4b* in our qPCR analysis.

Expression levels of the arterial marker genes *dll4* and *efnb2a* at 28/29 hpf and 33/34 hpf were strongly reduced in ++HIGH cells compared to ++LOW cells (Figure 5-19 A-B). This finding agrees with a recent *in vitro* study showing that the HE is DLL4^{-} (Ditadi et al., 2015). However, in the haemogenic ++HIGH population of *runx1*$^{W84X/W84X}$ mutants, expression levels of *dll4* and *efnb2a* were significantly increased (Figure 5-19 A-B). Especially in the case of *dll4,* expression levels in *runx1*$^{W84X/W84X}$ mutants were similar to those found in ++LOW cells of the DA roof (Figure 5-19 A). These findings indicate that the loss of the arterial programme in the HE is Runx1-dependent, most likely through the activity of Gfi1 proteins.

In agreement with a Runx1-dependent repression of arterial genes, the expression levels of *dll4* and *efnb2a* were much higher in the early HE at 24/25 hpf when *runx1* expression had just initiated (Figure 5-19 A-B). Strikingly, *dll4* expression levels were identical between the ++LOW and ++HIGH populations at this time point (Figure 5-19 A). By contrast, expression of the venous marker gene *ephb4b* was always depleted from the ++HIGH population and was unchanged in *runx1*$^{W84X/W84X}$ mutants (Figure 5-19 C). These findings suggest that the arterial endothelium and the HE share a common *dll4*$^{+}$ arterial progenitor from which the HE diverges by down-regulating the arterial and up-regulating the haemogenic programme.

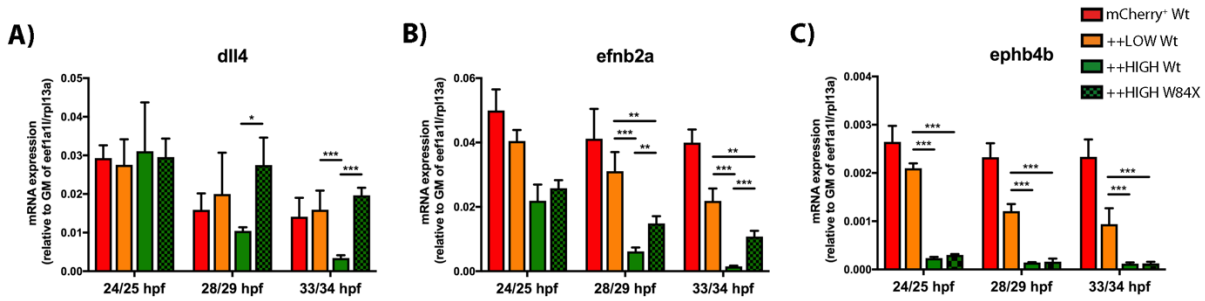


Figure 5-19: Gene expression analysis of aretrial and venous marker genes reveals an arterial origin of the HE.

A-B) Expression of arterial marker genes *dll4* and *efnb2a* was up-regulated in the HE of runx1$^{W84X/W84X}$ mutants. C) Expression of the venous marker gene *ephb4b* was unaffected in runx1$^{W84X/W84X}$ mutants. qPCR analysis was performed using TaqMan probes on the BioMark platform (Fluidigm). Expression levels detected within the indicated cell populations sorted from *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos are presented in relation to the geometric mean (GM) of the two housekeeping genes *eefia1l* and *rpll13a.* Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and ++HIGH (Wt and W84X) populations for each independent time point: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean. An N=5 experiments were performed for Wt embryos and N=6 for mutants.

To investigate these findings in greater detail, we went back and analysed expression changes of additional arterial and venous marker genes in the haemogenic ++HIGH population of control and *runx1* morphant embryos in our RNAseq data (Figure 5-20 A). Strikingly, expression of all analysed arterial marker genes was up-regulated in the MO condition, whereas all analysed venous marker genes remained relatively unchanged (Figure 5-20 A). These data highlight, that the *runx1*$^{W84X/W84X}$ HE remains largely arterial.

One of the arterial genes was *sox17,* a gene that was also found to mark the HE *in vivo* and *in vitro* (Choi et al., 2012; Clarke et al., 2013; Nakajima-Takagi et al., 2013). In mouse endothelial cells, SOX17 was shown to directly bind to and positively regulate expression of *Dll4,* (Corada et al., 2013). This and our findings suggests that Runx1 may repress *dll4* indirectly through repression of *sox17*.

In agreement with previous findings, we detected *sox17* (and *sox32,* a divergent *soxF* member found exclusively in teleosts (Chung et al., 2011)) expression in the ++HIGH population, but not within other endothelial cell types of the early embryo (Figure 5-20 B-C). Strikingly, expression of both genes was strongly up-regulated in the ++HIGH cells of *runx1*$^{W84X/W84X}$ embryos. This was validated by ISH, where *sox17* expression was only detectable in the HE of *runx1*$^{W84X/W84X}$ mutants but not in

Wt embryos (Figure 5-20 D). An inverse relationship between *Runx1* and *Sox17* expression levels in the murine HE has been shown previously (Bos et al., 2015; Chen et al., 2016). Here, we provide the first *in vivo* data indicating that Runx1 is indeed repressing *sox17* expression within the HE.

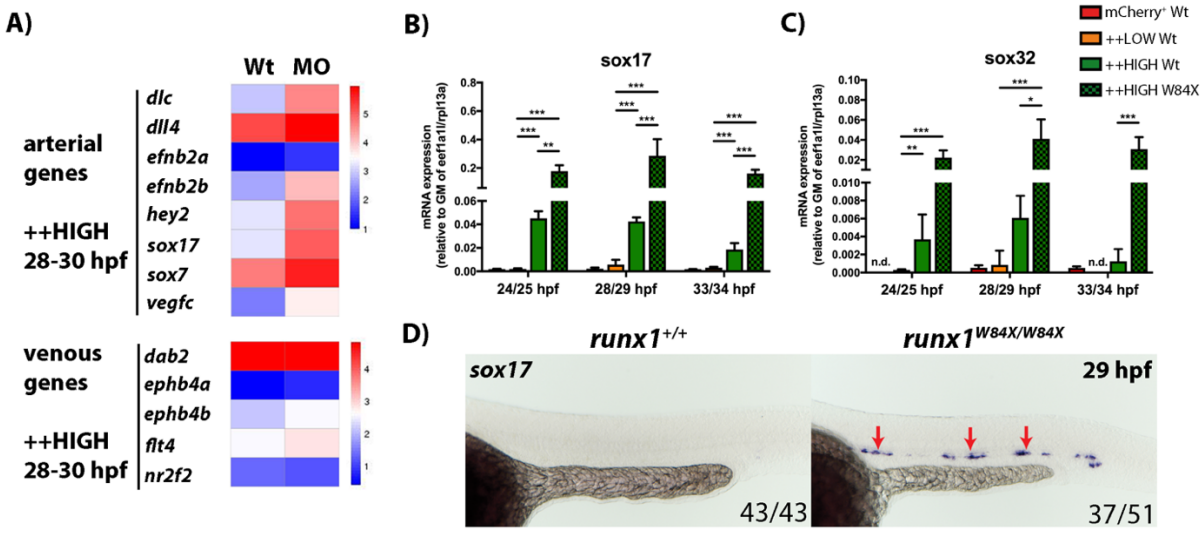


Figure 5-20: The HE of *runx1*$^{W84X/W84X}$ mutants retains an arterial programme and expression of the HE marker sox17.

A) Expression analysis of known arterial and venous genes within the ++HIGH population between the runx1 morphants (MO) and control (WT) embryos by RNAseq revealed that arterial but not venous genes are up-regulated in the loss of Runx1. B-C) Expression of *sox17* (B) and *sox32* (C) was significantly up-regulated in *runx1*$^{W84X/W84X}$ mutants. D) ISH revealed sox17 expression in the HE of runx1$^{W84X/W84X}$ mutants. qPCR analysis was performed using TaqMan probes on the BioMark platform (Fluidigm). Expression levels detected within in the indicated cell populations sorted from *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos are presented in relation to the geometric mean (GM) of the two housekeeping genes *eefia1l* and *rpll13a*. Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and ++HIGH (Wt and W84X) populations for each independent time point: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean. An N=5 experiments were performed for Wt embryos and N=6 for mutants.

We have shown that Citrine^{+} cells were still detectable in the region of the ventral lining of the DA in *runx1*$^{W84X/W84X}$ mutants at 3 dpf (Chapter 3 Figure 3-12). Because ++HIGH cells of *runx1*$^{W84X/W84X}$ mutants maintained an arterial program, we wondered if Citrine^{+} cells were still detectable in mutant embryos around 6 dpf, ~4.5 days after the HE undergoes EHT in Wt embryos.

The adult haematopoietic niches in the thymus and the kidney were completely depleted of Citrine^{+} cells in *runx1*$^{W84X/W84X}$ mutants, when compared to WT (Figure 5-21 A). Strikingly, Citrine^{+} cells with an elongated cell shape were still found in the ventral lining of the DA but were absent in Wt embryos

(Figure 5-21 B). The activity of the *kdrl:mcherry* reporter showed that the vascular system of *runx1*$^{W84X/W84X}$ mutants was intact and validated that these Citrine^{+} cells found in the DA are indeed endothelial. The numbers of embryos showing a loss of Runx1 phenotype are consistent with the expected Mendelian ratio if derived from heterozygous *runx1*$^{W84X/+}$ in-crosses (Figure 5-21 C). To make sure that no compensatory effects are introduced by the up-regulation of other *runx* family members, we checked for expression of *runx2a*, *runx2b* and *runx3* in our RNAseq data. Expression of *runx2a* and *runx2b* in the ++HIGH population was not detectable for either control or *runx1* morphant embryos and *runx3* showed a slight non-significant reduction (Figure 5-21D).

In summary, we have demonstrated that the early HE expresses an arterial programme that is down-regulated over time in a Runx1-dependent manner. In *runx1*$^{W84X/W84X}$ mutants, the HE cells continue expressing arterial genes and remain part of the DA lining.

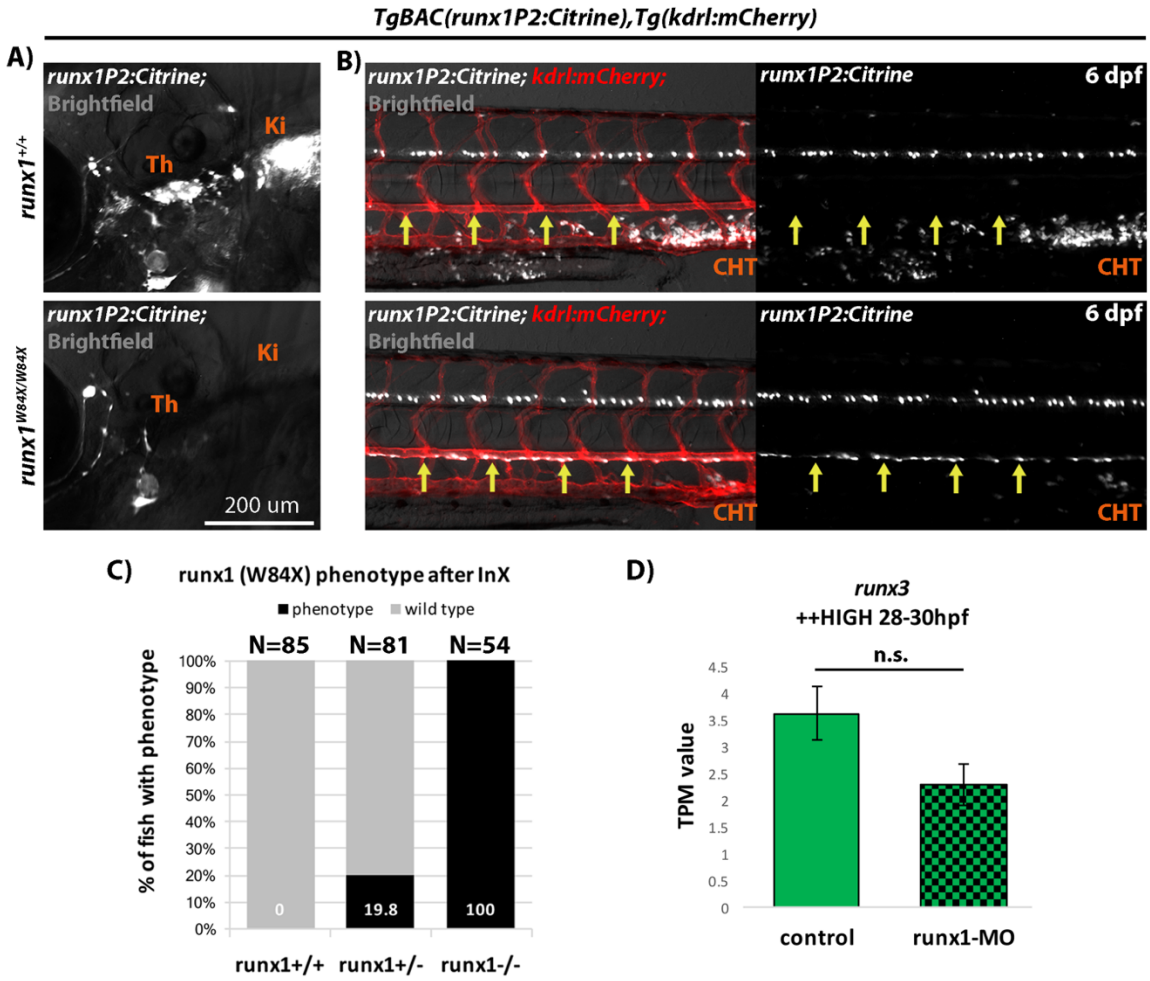


Figure 5-21: The HE in *runx1* mutants remains integrated into the DA scaffold.

A-B) *TgBAC(runx1P2:Citrine);Tg(kdrl:mCherry)* embryos on a *runx1*$^{+/+}$ or *runx1*$^{W84X/W84X}$ mutant background were analysed for transgene expression at 6 dpf. A) In *runx1*$^{+/+}$ embryos Citrine^{+} cells were detectable within the thymus (Th) and the kidney (Ki), whereas in *runx1*$^{W84X/W84X}$ mutant embryos no Citrine^{+} cell was detectable in these haematopoietic tissues. B) In the trunk region of *runx1*$^{+/+}$ embryos no Citrine^{+} cells were detected within the DA (yellow arrows) but numerous cells in the region of the CHT. In *runx1*$^{W84X/W84X}$ mutant embryos Citrine+ cells were still found in the ventral lining of the DA (yellow arrow) but strongly reduced in the CHT. C) Quantification of embryos showing a *runx1* mutant phenotype derived from in-crosses (InX) of parents with the indicated genotypes reveals a mendelian ratio. D) Expression analysis of other *runx* family members in runx1 morphants indicates the absence of any compensatory effect. Expression of *runx2a* and *runx2b* in the ++HIGH population was not detected in any case. Error-bars represent the standard deviation.

5.3 Discussion

5.3.1 Successful screening of Runx1 target genes

In our attempt to identify genes whose regulation by Runx1 during definitive haematopoiesis has not been demonstrated before, we focused mainly on those 16 genes that we defined as high-confidence Runx1 target genes (Table 5-1). Thereby, we solely focused on Runx1-activated genes. Of those, 13 genes were analysed by ISH and 11 by qPCR. Expression in the HE and a Runx1-dependency was validated for 10 genes (Table 5-5), a success rate of ~77 %. In contrast, of the 16 additional selected genes, only 2 (*dnmt3aa* and *ppm1g*) were regulated by Runx1, a success rate of ~12.5 %. This highlights the power of our selection pipeline (Figure 5-1) and emphasizes the focus on genes with an evolutionarily conserved expression. Even though such genes were shown to be bound by RUNX1 in haematopoietic cell lines and to be expressed during definitive haematopoiesis, the here identified Runx1 dependency is novel.

Additionally, we analysed 4 potential target genes negatively regulated by Runx1 (*abcg2a*, *gata2a*, *sox17* and *sox32*) which was proven to be true for all of them.

Table 5-5: Summary of the performed screens to validate HE-specific Runx1-activated genes.
For 10 of the 16 defined high-confidence Runx1 target genes a HE-specific expression and a Runx1 dependency could be validated by ISH and/or by qPCR.

	Zebrafish gene names	ISH analysis HE expression	qPCR analysis HE expression	qPCR analysis Runx1 regulated	ISH analysis Runx1 regulated
1	*angpt1*	YES	YES	YES	YES
2	*dnmt3ab*	YES	YES	YES	/
3	*dnmt3ba*	YES	YES	YES	YES
4	*dnmt3bb.1*	YES	YES	YES	YES
5	*dnmt3bb.2*	NO	YES	YES	/
6	*gfi1ab*	YES	YES	YES	YES
7	*irf1b*	YES	YES	YES	YES
8	*mpl*	YES	YES	YES	/
9	*pik3cd*	YES	YES	YES	/
10	*syk*	NO	YES	YES	/

Overall, we identified new regulatory relationships between Runx1 and its target genes that can be grouped as genes that might play a role in the transition from HE to HSPCs (*sox17*, *dll4*, *gfi1ab*) and as genes that might be required for the acquisition of the haematopoietic fate (*dnmt3b* isoforms,

dnmt3a isoforms, *pik3cd*, pik3r1, *irf1b*, *syk*, *mpl*, *angpt1, itga4*) (Figure 5-22). We also included *itga4* in this list, even though HE expression was not detectable by ISH (TaqMan probes were not available at the time), since *itga4* has been shown to be an important niche interaction factor (Craddock et al., 1997; Papayannopoulou and Craddock, 1997), bound by RUNX1 in HPC-7 cells and down-regulated in HSCs from *Runx1* mutants (Cai et al., 2011; Wilson et al., 2010a). Similarly, our RNAseq analysis revealed that *itga4* showed the highest expression in the ++HIGH population and that it was down-regulated in runx1 morphants (Appendix 9).

For all the here identified genes, direct binding by RUNX1 has been reported by others (Edwards et al., 2009; Gao et al., 2015; Tanaka et al., 2012; Wilson et al., 2010a). Expression of *dll4* is also likely regulated by Sox17 (Corada et al., 2013). Of note, the three genes *mpl*, *angpt1* and *itga4*, that all have reported function in the niche interaction, were also significantly down-regulated in HSCs from mouse *Runx1*$^{-/-}$ mutants (Cai et al., 2011). The down-regulation in *Runx1*$^{-/-}$ mutant HSCs was also true for the murine *Dnmt3b*. In the following discussion, I will focus on the individual groups of genes as indicated in Figure 5-22.

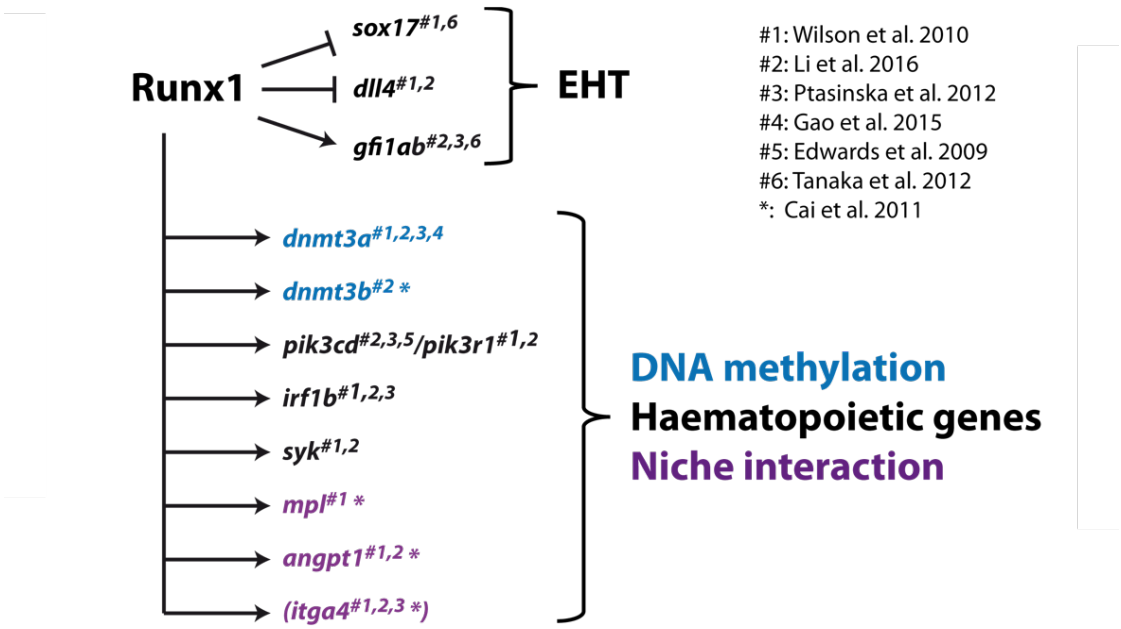


Figure 5-22: Role of Runx1 during the transition from HE to HSPCs.
Schematic highlights the identified Runx1-regulated genes and groups them according to their expected function. *: Genes that were also found to be differentially expressed in murine *Runx1* mutant HSCs. #1-5: Genes bound by RUNX1.

5.3.2 Runx1 orchestrates EHT through activation of *gfi1ab* and repression of *sox17* and Dll4-Notch signalling

In mouse, GFI1 and GFI1B have been shown to be the driving force of EHT, both *in vitro* and *in vivo* (Lancrin et al., 2012; Thambyrajah et al., 2015), mainly acting through the repression of the endothelial programme in the HE (Lancrin et al., 2012; Thambyrajah et al., 2015). They are direct transcriptional targets of RUNX1 (Lancrin et al., 2012; Wilson et al., 2010b). The situation for the *gfi1* genes in zebrafish is less clear. Here, a gene duplication event resulted in *gfi1aa* and *gfi1ab* that are both related to the mammalian *Gfi1* gene (Cooney et al., 2013). Major efforts have been centred around studying the function of *gfi1aa* and *gfi1b* (Cooney et al., 2013; Wei et al., 2008). Whereas *gfi1b* is mainly expressed in primitive blood, *gfi1aa* showed an expression in the HE at least from 22 hpf onwards (Figure 5-16), but was found to be controlled by Notch signalling independent of Runx1 (Figure 5-23 A) (Thambyrajah et al., 2016). Also, *gfi1aa*$^{-/-}$ mutants have no definitive haematopoietic defects. It is not clear if compensatory effects are involved as seen for the functionally interchangeable murine *Gfi1* and *Gfi1b* isoforms, whereby in *Gfi1*$^{-/-}$ mutants, *Gfi1b* expression can be found in the HE instead of *Gfi1* (Fiolka et al., 2006; Thambyrajah et al., 2015).

Through our genome wide expression screen, we identified *gfi1ab* as regulated by Runx1. MO-mediated knock-down of *gfi1ab* resulted in a reduced number of *cmyb*$^{+}$ HSPC in the CHT, most likely caused by a block in EHT (Figure 5-18 B). However, we take the displayed phenotype in *gfi1ab* morphants with caution, since compensatory effects are described for that gene family. Studies in genetic knockouts might therefore be mandatory (Fiolka et al., 2006). On the other hand, it has also been described that a compensatory network can buffer against genetic mutations, a phenomenon that is not observed after MO-induced knockdown (Rossi et al., 2015).

GFI1 proteins act through recruitment of the chromatin-modifying protein LSD1, a member of the CoREST complex (Saleque et al., 2007). A RUNX1-GFI1-LSD1 axis was shown to epigenetically silence the endothelial programme in the HE (Thambyrajah et al., 2015). To determine if EHT in zebrafish is controlled by similar mechanisms to mouse, we abolished Lsd1 activity and showed that zebrafish EHT is also orchestrated by a Runx1-Gfi1-Lsd1 axis (Figure 5-18 C-E, Figure 5-23 B).

Making use of a newly designed *gfi1aa* MO (Figure 5-17 A-B), we showed that expression of *gfi1ab* is already induced at 24 hpf (Figure 5-17 C), which is shortly after *runx1* expression becomes detectable by ISH and at least 4 hr earlier then the up-regulation of any other Runx1 target we identified (Figure 5-13). Early HE-specific expression of *gfi1ab* was also seen by qPCR at 24/25 hpf (Figure 5-11 B). For most other Runx1 targets, up-regulated expression in the HE was detected from 28/29 hpf onwards (Figure 5-8, Figure 5-11, Figure 5-15). 28/29 hpf is also the time when haemogenic expression of *cbfb* was up-regulated (Figure 5-5 F). These findings indicate that, whereas most Runx1 targets might be regulated in a Cbfb-dependent manner, the control of *gfi1ab* expression by Runx1 likely does not require Cbfb (Figure 5-23 B). In fact, zebrafish embryos mutant for *cbfb* do not show a block in EHT; cells of the HE still emerge into the sub-aortic space but subsequently fail to enter circulation (Bresciani et al., 2014). Furthermore, murine RUNX1 was shown to bind DNA even in the absence of CBFb, but in a less stable manner (Hoogenkamp et al., 2009). These findings support the notion that Gfi1ab might indeed be required for EHT. We plan to perform gene expression experiments in *cbfb* mutants to show that Cbfb is dispensable for *gfi1ab* expression, but potentially required for all the other identified target genes.

Whereas expression of *gfi1ab* in the HE increased over time between 24/25 hpf and 33/34 hpf (Figure 5-11 B, Figure 5-17 C), expression of *gfi1aa* decreased (Figure 5-5 C, Figure 5-16 A-C). Since *gfi1aa* expression was reported to be controlled by Notch signalling (Thambyrajah et al., 2016) and we detected a down-regulation of *dll4* in the HE (Figure 5-19 A), we propose that Runx1 indirectly down-regulates *gfi1aa* by preventing Dll4 induced Notch signalling (Figure 5-23 B). A direct binding of RUNX1 to *Dll4* is seen in ChIPseq studies (Figure 5-22), but is not reported otherwise. We speculate that RUNX1 mainly blocks *Dll4* expression indirectly by repressing *Sox17*, which was shown to directly regulate transcriptional activation of *Dll4* in endothelial cells (Corada et al., 2013).

In mouse embryos, expression of *Sox17* can be detected in the HE (Choi et al., 2012; Lizama et al., 2015). Furthermore, an inverse relationship between *Sox17* and *Runx1* expression levels in the HE has been described (Bos et al., 2015; Chen et al., 2016; Lizama et al., 2015). SOX17 was shown to be required for priming the haemogenic potential and to “lock” the HE state (Clarke et al., 2013;

Nakajima-Takagi et al., 2013); overexpression of *Sox17* reprogrammed haematopoietic progenitors into HE-like cells (Nakajima-Takagi et al., 2013). Direct repression of *Sox17* by RUNX1 was also suggested previously *in vitro* during differentiation culture of early blood progenitors (EBPs) or in human umbilical artery endothelial cell (HUAEC) culture (Lizama et al., 2015; Tanaka et al., 2012). Here, we provide the first functional *in vivo* data indicating that Runx1 is indeed repressing *sox17* expression within the aortic HE.

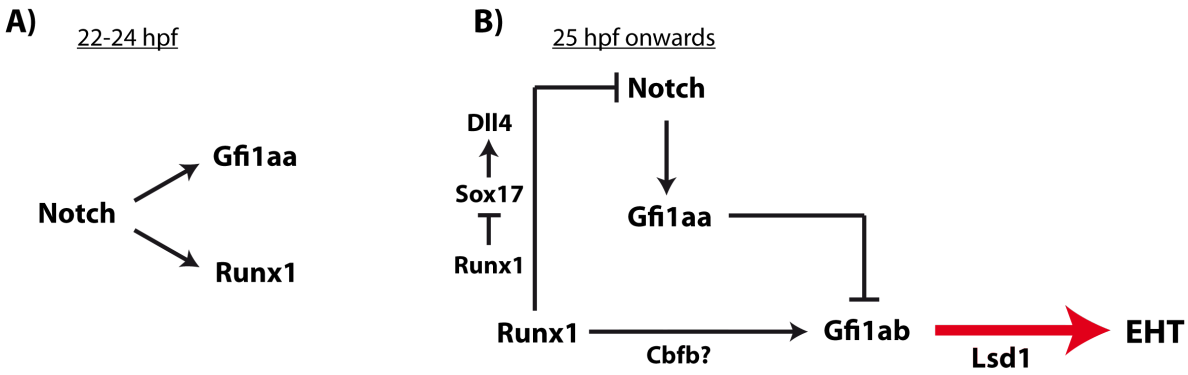


Figure 5-23: Regulation of *gfi1aa* and *gfi1ab* in the HE.

A) Expression of gfi1aa and runx1 in the early HE around 22-24 hpf is induced by Notch signalling (Thambyrajah et al., 2016). Expression of gfi1aa is independent of Runx1. B) Expression of gfi1ab might be induced by Runx1 via two independent mechanisms: First, via a direct regulation (here it remains to be defined if this regulation is dependent on Cbfb). Second, via an indirect regulation by a reduction of Notch signalling upstream of Gfi1aa which represses expression of gfi1ab. Notch signalling might be repressed by repressing Sox17 which is a positive input for Dll4. EHT is then orchestrated via a conserved Gfi1-Lsd1 axis.

The down-regulation of *dll4* in the HE (Figure 5-19 A) might not only have implications for controlling expression levels of *gfi1aa* and consequently of *gfi1ab*. Different ratios between DLL4 and JAG1 mediated Notch signalling have been shown to define arterial and haemogenic fate outcome (Gama-Norton et al., 2015). JAG1 is needed for HSC specification, whereas high DLL4 signalling drives arterial specification (Espín-Palazón et al., 2014; Gama-Norton et al., 2015; Monteiro et al., 2016; Robert-Moreno et al., 2008). Interestingly, we did not detect significant differences in the expression levels of *jag1a* between the HE and the DA roof cells (Figure 5-5 I). The different ratios between Jag1a and Dll4 were rather established by a specific down-regulation of *dll4* expression in the HE (Figure 5-19 A). Therefore we propose a model in which cells of the early DA around 24/25 hpf are expressing *dll4* at high levels, thus driving an arterial programme in all cells (Figure 5-24 A).

Jag1 expression levels remain uniform and are required in the HE to drive expression of *runx1,* together with additional localized signals (Espín-Palazón et al., 2014; Monteiro et al., 2016; Robert-Moreno et al., 2008; Wilkinson et al., 2009). Activity of Runx1 in the HE down-regulates *dll4* and thus Notch signalling, which is necessary for further haematopoietic commitment (Zhang et al., 2015). In agreement with this model, a mouse study utilizing transgenic Notch reporter lines with different sensitivity suggested that haematopoietic cells originate from cells with a low Notch signal (Gama-Norton et al., 2015). However, due to the expression dynamics of the transgene it could not be resolved if arterial cells and HSCs originate from distinct or a common precursor. Our model proposes that HE cells are originally derived from precursors with high Notch signalling, similar to arterial cells.

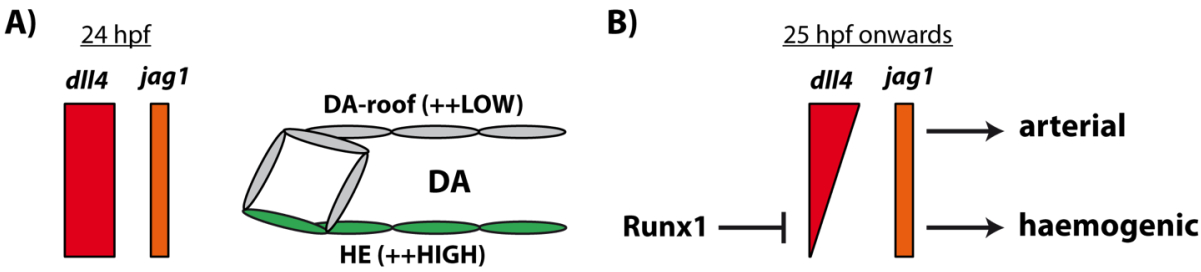


Figure 5-24: A higher Jag1/Dll4 signalling ratio in the HE is accomplished by a Runx1-dependent down-regulation of *dll4*.

A) The early DA around 24 hpf shows similar expression levels of dll4 and jag1 in both cells of the DA roof and the HE. B) With the up-regulation of runx1 in the HE population expression of dll4 gets repressed leading to an increased Jag1/Dll4 ratio. A lower Jag1/Dll4 ratio maintains an arterial programme whereas a higher ratio is required for haematopoietic differentiation.

5.3.3 Runx1 establishes an epigenetic state in haematopoietic progenitor cells through regulation of *dnmt3* genes

RUNX1 is pivotal during the time window in which the HE undergoes EHT but not thereafter (Chen et al., 2009). Conditional knock-out of RUNX1 in the FL or in bone-marrow HSCs leads to minor haematopoietic defects but does not abolish HSCs (Cai et al., 2011; Chen et al., 2009; Growney et al., 2005; Ichikawa et al., 2004; Ichikawa et al., 2008). A more detailed analysis of the temporal requirements for RUNX1 in the HE and early progenitors revealed a requirement in at least a subset of forming HSCs, after most HE cells have already differentiated into IACs (Tober et al., 2013). It is

speculated that RUNX1, among other activities, establishes an epigenetic state that will eventually stabilise and thus become RUNX1 independent. In that sense, a global analysis of RUNX1 binding within *in vitro* generated HE-like cells depicted RUNX1 as a pioneer factor that induced acetylation of H3K9 at a multitude of chromatin sites (Lichtinger et al., 2012). Furthermore, induction of RUNX1 expression led to a global redistribution of TAL1 and FLI1 binding sites (Lichtinger et al., 2012).

Additional epigenetic modifications include the methylation of DNA, introduced and maintained through DNMT enzymes (Jones, 2012; Robertson, 2005). DNA methylation controls the activated or repressed state of genes, depending on context (Jones, 2012). Mechanistically, DNA methylation is established via the activity of the *de novo* DNMT3 isoforms and maintained through cell division by the function of the DNMT1 enzyme. A recent study highlighted that the zebrafish *dnmt3bb.1* gene is regulated by Runx1 and important for the gene body methylation of *cmyb* and thus required to maintain expression of *cmyb* in HSPCs (Gore et al., 2016). Similarly, expression of *Dnmt3* in mouse was also detected in IACs (Watanabe et al., 2004) and *Dnmt3a* was directly bound by RUNX1/ETO in AML cells (Gao et al., 2015). Here, we have provided evidence that at least 5 of the 6 zebrafish *dnmt3* isoforms are regulated by Runx1 within the HE (Table 5-5). Furthermore, sustained expression in the region of the CHT was detected for both *dnmt3ba* and *dnmt3bb.1* (Figure 5-10 C-D) as well as for *dnmt1* (Figure 5-10 A), indicating a requirement for the maintenance of introduced methylation marks. Such a requirement was shown in *dnmt1*$^{-/-}$ mutant zebrafish, where *cebpa* regulatory regions were hypomethylated, causing *cebpa* up-regulation that resulted in an HSPC proliferation block (Liu et al., 2015). It is furthermore tempting to speculate, that DNA methylation is involved in permanently silencing the endothelial programme in developing HSCs, initially repressed by GFI1 and LSD1 (Thambyrajah et al., 2015). An recruitment between DNMT3a and LSD1 that regulates DNA methylation has been described (Petell et al., 2016).

5.3.4 Runx1 regulates genes with reported functions in the haematopoietic system already during HE stages

We identified 3 Runx1-activated genes (*irf1b*, *pik3cd* and *syk*) with reported roles in mature immune cells. Members of the IRF family of TFs function in immunity but are also involved in the cellular differentiation of haematopoietic cells (Tamura et al., 2008). A role for inflammatory signalling during early definitive haematopoiesis has been discovered recently (Espín-Palazón et al., 2014; He et al., 2015; Li et al., 2014; Lu et al., 2013; Sawamiphak et al., 2014). Here, all identified processes and the activity of two zebrafish Irf factors, Irf7 and Irf2, were thought to be involved in the programming of the HE upstream of *runx1* (Li et al., 2014; Lu et al., 2013). With *irf1b*, we found for the first time an inflammatory gene downstream of Runx1 with a HE-specific expression (Figure 5-12 A).

In mouse, IRF1 is involved in the regulation of myeloid differentiation as well as the development of $CD8^+$ T-cells, at the expense of CD4+ regulatory T-cells (Abdollahi et al., 1991; Fragale et al., 2008; Matsuyama et al., 1993; Penninger et al., 1997; Tamura et al., 2008; Testa et al., 2004). Strikingly, RUNX1 is also required for $CD8^+$ T-cell development during thymopoiesis and for silencing *CD4* expression (Taniuchi et al., 2002; Woolf et al., 2003), which suggests that RUNX1 might regulate expression of *Irf1* also during the development of T-cells. A functional role for IRF1 has not been analysed during embryonic haematopoiesis. We plan to study the role of Irf1b by generating CRISPR/Cas9 mutants.

The *Pik3cd* gene is predominantly expressed in leukocytes and is a member of Class I of the phosphoinositide 3-kinases (PI3K) (Chantry et al., 1997; Deane and Fruman, 2004). The catalytic PIK3CD isoform builds a functional PI3Kd complex together with its regulatory binding partner PIK3R1. PI3Ks are part of the mTOR pathway and catalyse the phosphorylation of phosphatidylinositol-4,5-bisphosphate (PIP_2) to phosphatidylinositol-3,4,5-trisphosphate (PIP_3), antagonized by PTEN (Deane and Fruman, 2004; Lucas et al., 2016). In zebrafish Pten is required to prevent hyperproliferation of HSPCs in the CHT (Choorapoikayil et al., 2014; Dong et al., 2014). These effects can be rescued by a pan chemical inhibitor LY294002 affecting all PI3Ks. Furthermore,

the zebrafish Pik3cg isoform has been reported to act upstream of *runx1* expression (Li et al., 2015). A role for the *Pik3cd* and *Pik3r1* genes for development of HSCs has not been described yet.

The specific expression of *pik3cd* in the zebrafish HE was intriguing (Figure 5-2 H). Interestingly, the *Pik3cd* promoter is directly bound to by RUNX1 in the acute megakaryocytic leukaemia (AMKL) cell line, Meg-01 (Edwards et al., 2009). In AML, the PIK3CD isoform was shown to be essential for proliferation (Martelli et al., 2006; Sujobert et al., 2005). PIK3CD is also involved in natural killer (NK) cell development, in addition to its function in the secretion of cytokines by different types of immune cells (Kim et al., 2007; Soond et al., 2010). It remains to be analysed if *pik3cd* expression in the HE and during HSC development has implications for proliferation, or if it might be involved in the inflammatory signalling that was recently discovered to be required for definitive haematopoiesis (Espín-Palazón et al., 2014; He et al., 2015; Li et al., 2014; Lu et al., 2013; Sawamiphak et al., 2014). We are generating CRISPR/Cas9 *pik3cd* mutant lines to study its role in HSPCs.

Syk encodes a non-receptor tyrosine kinase which is widely expressed in haematopoietic cells and transmits signals from B- and T-cell receptors as well as from integrins (Seda and Mraz, 2015; Turner et al., 2000). During embryogenesis, *syk* expression in myeloid cells is involved in the proper separation of the lymphatic and blood system, and *Syk*$^{-/-}$ mutant mice die at midgestation due to blood-lymphatic shunts (Abtahian, 2003; Böhmer et al., 2010; Sebzda et al., 2006). Whether expression of *syk* has any implications for the development of the HSC system is not known. Of note, blood-lymphatic shunts lead to embryonic haemorrhages, which are also observed in *Runx1*$^{-/-}$ mutant mice (Okuda et al., 1996).

In summary, the three genes *syk*, *irf1b* and *pik3cd* have known roles in mature blood cells but might have additional roles in embryonic haematopoiesis. Further studies are needed to elucidate their specific functions

5.3.5 Runx1 regulates expression of stem cell niche interaction factors

We identified three Runx1-activated genes (*mpl*, *angpt1* and *itga4*) with reported roles in the HSC niche. Several data-sets indicate that MPL might be an important niche factor throughout embryonic haematopoiesis. The thrombopoietin (TPO) receptor encoded by the *Mpl* gene is expressed by all HSCs and in the adult bone marrow niche. Long-term HSCs are closely associated with TPO-producing osteoblastic cells, a critical association for the maintenance of the quiescent HSC population (Yoshihara et al., 2007). *Mpl* mutant mice have reduced HSC repopulating activity (Kimura et al., 1998; Solar et al., 1998). Moreover, murine *Mpl* is expressed by HSCs within IACs and in the FL as early as E10.5 (Harada et al., 2017; Petit-Cocault et al., 2007). It was suggested that Mpl contributes to the formation and the maintenance of IACs (Harada et al., 2017). Functional consequences of the loss of *Mpl* during mouse haematopoiesis are a delayed production of HSCs by the AGM region, accompanied by a reduced amplification and survival/self-renewal of HSCs during their passage to the FL (Petit-Cocault et al., 2007). Our ISH was not sensitive enough to detect expression of *mpl* in the zebrafish embryo (Figure 5-2 I). However, both, the RNAseq and qPCR analysis detected specific *mpl* expression in the HE that was ~7 to 20-times higher than in other endothelial populations and totally dependent on Runx1 (Figure 5-15 B).

Angpt1 encodes a ligand for the endothelium-specific TEK (TIE2) receptor and is mainly involved in angiogenesis (Fukuhara et al., 2010). In a study by Takakura et al. (2000) it was shown that embryonic HSCs are a source of ANGPT1 and potentially involved in angiogenesis (Takakura et al., 2000). Vascular systems in yolk sac, heart and head of *Runx1*$^{-/-}$ embryos are defective in the development of branched small vessels from arteries (Okuda et al., 1996). It is speculated that the haemorrhages seen in *Runx1*$^{-/-}$ mutant mouse embryos (Okuda et al., 1996) are secondary effects due to the lack of ANGPT1-producing HSCs. Our findings show that the expression of *angpt1* is indeed regulated by Runx1 (Figure 5-11 A, Figure 5-12), This interaction might be direct since in the mouse endothelial cell line Mss31, expression of a CBFb-MYH11 fusion protein, which interferes with the biological activity of RUNX1, inhibited the expression of *Angpt1* (Namba et al., 2000). Within the

perivascular bone marrow niche, *Angpt1* was found most highly expressed by HSCs, where it reduced vascular leakiness under niche regeneration conditions (Zhou et al., 2015). Even though expression of *Angpt1* seems to be dispensable for adult HSCs under steady-state conditions, the data indicate that ANGPT1 might be an additional niche interaction factor.

Interestingly, we detected high expression levels of *angpt1* in the HE and high levels of *tek* in the DA roof (Figure 5-13 S). The absence of tek expression seems to be a defining feature of HE that can generate HSC, distinguishing it from the HE found in the YS (Chen et al., 2011). Since ANGPT1 can also act as a chemoattractant, so that endothelial cells sprout towards ANGPT1 producing tissues (Witzenbichler et al., 1998), we hypothesize that one function of Angpt1 expression in the zebrafish HE might be to attract lateral DA cells to allow a fast closure of the potential lesion introduced by the budding of HE cells out of the DA lining. Such rearrangements of endothelial cells during the time-window of EHT have been observed in zebrafish, where cells of the DA lateral or even dorsal wall move to the DA floor (Kissa and Herbomel, 2010).

The *Itga4* gene is expressed by HSCs and is involved in their migration (Craddock et al., 1997; Hirsch et al., 1996). The use of adhesion-blocking antibodies against ITGA4 led to increased HSC frequency in the blood (Craddock et al., 1997; Papayannopoulou and Craddock, 1997). Expression of *Itga4* was also found to define embryonic HSCs from E11.5 onwards, and was critical for HSCs to be able to develop through the foetal stage (Gribi et al., 2006). Because *Itga4* can be directly bound by RUNX1 (Wilson et al., 2010a) and our RNAseq showed that in the zebrafish HE *itga4* is down-regulated in the absence of Runx1 (Appendix 9), we hypothesize that RUNX1 initiates *Itga4* expression in HSPC to allow for a proper interaction with the haematopoietic niche in the fetal liver during development.

In summary, the 3 genes *Mpl*, *Angpt1* and *Itga4* have been shown to be important niche interaction factors. Expression of these genes seems to be dependent on RUNX1. We speculate that part of the mechanism, through which RUNX1 facilitates the full commitment to a haematopoietic fate, is to control expression of such factors required for an optimal interaction with the subsequent

haematopoietic niches to allow full maturation into adult HSCs. The importance for correct niche interaction has also been highlighted in recent *in vitro* differentiation protocols (Lis et al., 2017; Sandler et al., 2014).

5.3.6 The HE is initially *Dll4*$^{+}$ and maintains an arterial programme in *runx1* mutants

To date, it remains a controversial question whether the HE represents an independent lineage or if it diverges from an arterial progenitor. In all vertebrate model organisms the HE is always found in close association with the arterial vasculature which led to the hypothesis that the HE and the arterial endothelium share a common lineage path (Ciau-Uitz et al., 2013; Clements and Traver, 2013; Kalev-Zylinska et al., 2002; Richard et al., 2013). In support of this, the initiation and progression of *Runx1* expression in the DA is detected in endothelial cells only after vessel structures have been established (Gering and Patient, 2005; North et al., 1999; Richard et al., 2013).

However, this view was challenged in a recent *in vitro* differentiation study. In that study by Ditadi et al. (2015), following an initial 3 days of mesodermal induction and a subsequent 5 days long haematopoietic and endothelial specification phase, a $CD34^{+}CD43^{-}$ cell fraction was identified that exhibited the capacity to generate T-lymphoid, erythroid and myeloid cells (Ditadi et al., 2015; Sturgeon et al., 2014). This population could be further subdivided into cells with different endothelial potentials: $CD73^{-}CD184^{-}$ (haemogenic), $CD73^{med}CD184^{+}$ (arterial) and $CD73^{hi}CD184^{-}$ (venous). Within the haemogenic fraction, lineage-restricted HE-like cells were identified that never gave rise to both endothelial and haematopoietic progenies. Interestingly, the haemogenic potential of the $CD73^{-}CD184^{-}$ cell fraction was further enriched if sorted for $Dll4^{-}$ cells. By contrast, $CD73^{-}CD184^{-}$ $DLL4^{+}$ cells were enriched for vascular potential. These data indeed indicate that the generated HE-like cells are separated from arterial cells at the time point of analysis. In agreement, another study detected that prospective HE cells from the murine AGM around E8.5 to E10.5, defined by the expression of the +23GFP reporter, only gave rise to endothelial or haematopoietic lineages, but never to both (Swiers et al., 2013b). Most of such HE cells of an E10.5 embryo lacked endothelial potential.

Consistent with these findings, we also observed a global down-regulation of the endothelial programme in the zebrafish HE already around 28-30 hpf (Chapter 4 Figure 4-8 D). Further, we found that the HE is indeed depleted of *dll4* expression at 33/34 hpf, while the DA roof cells maintain high expression of *dll4* (Figure 5-19 A). However, when analysing *runx1*$^{W84X/W84X}$ mutant embryos, expression levels for *dll4* in both the HE and the DA roof cells were similar (Figure 5-19 A), indicating that the loss of *dll4* in the HE is Runx1 dependent. Consequently, when we analysed *dll4* expression in a less mature HE around 24/25 hpf, just when *runx1* expression was initiated, we detected high levels of *dll4* in the HE, suggesting that the HE is derived from a *dll4*$^{+}$ progenitors. Since we also showed that the HE of *runx1* morphants maintained an arterial programme (Figure 5-20 A) and that the HE in *runx1*$^{W84X/W84X}$ mutants remained integrated into the DA lining, even 4.5 days after EHT occurred in Wt embryos (Figure 5-21 A), we speculate that the HE is indeed derived from *dll4*$^{+}$ arterial progenitors (Figure 5-25). In agreement, +23GFP^{+} HE cells from the murine AGM of an E8.5 embryo revealed an equal potential to form vascular tubules compared to +23GFP^{-} endothelial cells and, in addition, single E7.5 23GFP^{+} mesodermal cells gave rise to both, haematopoietic and endothelial cells (Swiers et al., 2013b). We speculate that such dynamic changes in *Dll4* expression were missed in the Ditadi et al. (2015) study, since cells were analysed at the end of a 5 days long specification phase which potentially represent a more mature HE-like state.

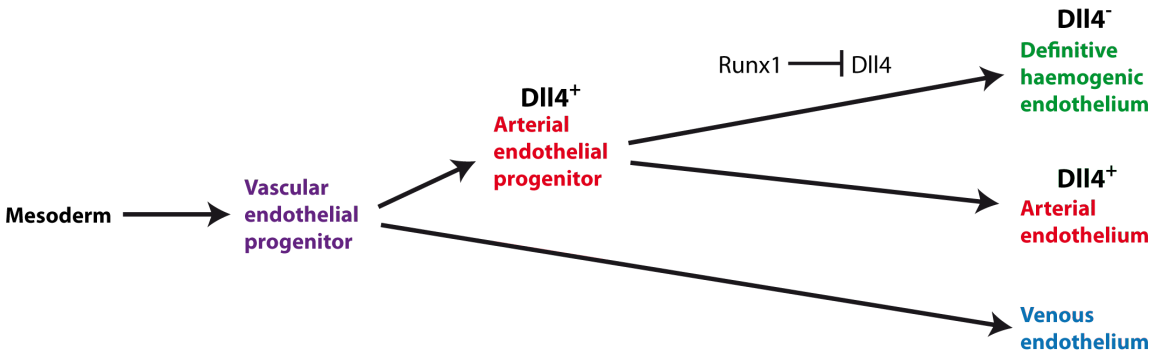


Figure 5-25: Model depicting the lineage relationship of the arterial and haemogenic endothelium.
Mesoderm-derived vascular endothelial progenitors are specified relatively early into arterial or venous progenitor cells (Zhong et al., 2001). The definitive HE is specified from Dll4+ arterial endothelial progenitors and down-regulates Dll4 in a Runx1-dependent manner.

It was also reported that genes required for arterial identity repress the haematopoietic fate (Lizama et al., 2015). A loss of such arterial genes in the HE promoted a haematopoietic fate conversion, indicating that such cells can switch between an arterial or haemogenic fate, again suggesting a common lineage path. Indeed, it is reported that *Dll4*$^{-/-}$ mice lack *Runx1* expression in the HE at E9.5 highlighting that an arterial identity is necessary to establish aortic HE (Robert-Moreno et al., 2008). Our laboratory generated similar data in which *runx1* expression is absent in the HE of *dll4* morphants (unpublished). Similarly, the arterial gene *EfnB2* was recently shown to be required for DA haematopoiesis (Chen et al., 2016).

It remains to be determined at which time point the HE diverges from the arterial lineage. RUNX1 is initially not needed for HE differentiation as haemogenic potential can already be detected in murine +23GFP^{+} cells before *Runx1* is expressed (Swiers et al., 2013b). Similarly, we detected HE-specific expression of *gfi1aa* already at 22 hpf (Figure 5-17 D), prior to *runx1* expression. Expression of *gata2b* in the zebrafish DA can be detected already around 18-20 hpf (Butko et al., 2015) and a recent study suggested that the HSC fate is partly established around 15 to 16 hpf during the axial migration of PLM cells, well before the formation of the DA (Kobayashi et al., 2014), suggesting a stepwise establishment of the haematopoietic fate. These findings still agree with our proposed model (Figure 5-25) since arterial and venous precursors in zebrafish were found to be separated in PLM cells already between the 7 to 12 somite stage (13 to 15 hpf) (Zhong et al., 2001).

In summary, we hypothesize that in zebrafish, mesoderm-derived vascular endothelial progenitors first diverge into a venous and an arterial lineage (Figure 5-25). Cells of the arterial lineage that migrate to the midline and form the DA will be exposed to extrinsic signals from various sources, including Dlc and Dld dependent Notch signalling from the somites or localized BMP signalling from mesenchyme ventral to the DA (Kobayashi et al., 2014; Pouget et al., 2014; Wilkinson et al., 2009). These signals will stepwise establish a haemogenic programme that finally involves the up-regulation of *runx1*. In turn, Runx1 then orchestrates the differentiation from the HE to HSCs by down-regulating the endothelial/arterial programme, by initiating EHT and by establishing a full commitment to the haematopoietic fate. This involves, among others, the establishment of a specific epigenetic state

involving DNA methylation and the expression of niche interaction factors that allows HSCs to fully mature within the subsequent haematopoietic niches.

5.4 Conclusion

Here, we identified several new RUNX1-regulated genes that might help to understand its specific role during HSC specification in greater depth. The diversity of such genes and their respective roles imply that a full maturation into fully functional HSCs does not depend on a few critical genes but on a whole set of factors that act in different ways. While some of the more global activities might be conducted by DNMT enzymes, other factors might allow the developing HSC to respond to extrinsic cues from haematopoietic niches. Lastly, it became evident that in the absence of RUNX1 the HE remains essentially arterial and stays an integrated part of the DA. Our data strengthens the hypothesis that the HE and the aortic endothelium share a common arterial lineage during early development

- Chapter 6 -

Final conclusion

The critical role of RUNX1 for the *de novo* generation of HSCs during embryogenesis involves orchestrating the EHT process that allows developing HSCs to bud out of the DA lining. Consequently, in the absence of RUNX1, the embryo simply fails to generate HSCs. However, *in vitro* and *in vivo* studies suggest that RUNX1 facilitates additional tasks to establish full haematopoietic commitment (Lancrin et al., 2012; Tober et al., 2013). This project aimed to identify novel RUNX1-regulated genes within the HE and to extend our knowledge about the mechanisms underlying *de novo* generation of HSCs.

Here I generated and characterized a novel zebrafish *runx1P2* reporter line based on a zebrafish BAC. Transgenic *TgBAC(runx1P2:Citrine)* embryos reliably reported *runx1* expression throughout the haematopoietic system during development. We exploited this reporter line to isolate pure HE from the other endothelial cells of the DA by FACS. Analysis of genome-wide expression within the respective tissues and upon Runx1 loss of function enabled us to identify Runx1-regulated genes within the HE. To select the most promising target genes, we applied a selection logic that identified genes with an evolutionarily conserved expression during embryonic haematopoiesis and that also have reported RUNX1 binding peaks by ChIPseq analysis *in vitro*. These high-confidence Runx1 target genes were verified by two complementary screens that validated their correct spatial expression and Runx1 dependency in mutant embryos. Overall, we verified 10 Runx1-activated genes (plus *itga4*; only identified by RNAseq). These target genes can be divided into 4 separate groups depending on their most likely function.

The first group contains the *gfi1ab* gene, that we speculate is the functional equivalent to the murine *Gfi1* genes and required for EHT. Here, we have presented data that indicate that a conserved Runx1-Gfi1-Lsd1 axis orchestrates EHT in zebrafish.

The second group contains the *dnmt3* isoforms *dnmt3ab*, *dnmt3ba*, *dnmt3bb.1* and *dnmt3bb.2,* genes that are involved in the *de novo* methylation of DNA. It is speculated that RUNX1 can also fulfil epigenetic functions on its own by either acting as a pioneering factor introducing H3K9Ac marks and by rearranging binding sites of TFs including FLI1 and TAL1 (Lichtinger et al., 2012). By controlling expression of *Dnmt3* genes, we propose that RUNX1 is further indirectly involved in regulating epigenetic modification involving DNA methylation in a broader sense than previously

reported (Gore et al., 2016). It is tempting to speculate that one potential role of DNA methylation during definitive haematopoiesis is to maintain silencing of the endothelial genes that were initially repressed by GFI1-LSD1.

The third group contains the three genes *syk*, *irf1b* and *pik3cd* (together with its partner *pik3r1*), all with reported functions in adult immune cells. Overall, not much is known about the role of *syk*, *irf1b* and *pik3cd* during definitive haematopoiesis. We plan to investigate *irf1b* and *pik3cd* function in more detail by generating CRISPR/Cas9 mutants. Since the zebrafish HE was shown to be a heterogeneous tissue that also gives rise to adult microglia (Xu et al., 2015), it remains to be analysed if the here identified Runx1 targets are actually involved in the stem cell programme or rather in the lineage determination of restricted progenitors. ISH analysis for various specific lineage markers will be informative in generated mutants.

The fourth group contains the genes *mpl* and *angpt1* (plus *itga4*), genes that are reported to be stem cell niche interaction factors in adult HSCs and that were found to be down-regulated in murine HSCs from *Runx1*$^{-/-}$ mutants. Our discovery that their expression in zebrafish is Runx1-dependent already in the HE suggests that these genes may be required for the maturation of HSCs in different niches, including the IAC or the FL(/CHT).

In summary, we speculate that once HSCs have emerged from the DA, their specification and maturation into fully committed HSCs must integrate several intrinsic and extrinsic mechanisms, including epigenetic regulation and interaction with the correct stem cell niches, and that RUNX1 is on top of these programmes. No severe HSC-related phenotype was reported in knockout mice for any of the target genes we identified, suggesting that each single factor can be substituted for in single knock-outs. This contrasts with the extreme phenotype of *Gfi1*$^{-/-}$ *Gfi1b*$^{-/-}$ mutant mice in which HE cells cannot initiate EHT and thus HSCs do not form (Thambyrajah et al., 2015). However, if *Runx1* is conditionally deleted shortly after EHT, the combined misexpression of the here identified RUNX1-targets might result in the failure of inducing the full HSC programme, established in subsequent stem cell niches (Chen et al., 2009; Tober et al., 2013). The exact mechanisms remain to be investigated.

As s second outcome of this study we could highlight a remarkably close relationship between the arterial and the haemogenic lineage, that has recently been under debate (Ditadi et al., 2015). Despite being always located in the lining of the major embryonic arteries, it has never been undisputedly resolved whether the HE represents an independent lineage generated from its own mesodermal precursor, or if the HE diverges from an arterial progenitor. Here, we have been able to show that the less mature HE expresses the arterial marker *dll4* at similar levels to arterial cells in the DA roof. To validate the co-expression of *runx1* and *dll4* within the HE on the single-cell level, double fluorescent ISH experiments will be performed. Furthermore, we found the arterial programme to be down-regulated in the HE in a Runx1-dependent manner, whereas expression of venous genes was unaffected and generally low. Strikingly, in *TgBAC(runx1P2:Citrine)* embryos, before/around the time of DA lumenization when all DA cells are exposed to the same localized signals to a similar extent, all DA cells expressed weak *Citrine* under the control of the endogenous *runx1P2* promoter. Consistent with this, studies have shown that when ventral localized BMP signalling was increased, *runx1* expression extended dorsally within the DA (Pouget et al., 2014). We speculate that the HE and aortic endothelium share a common arterial endothelial cell ancestor and that specifically aortic cells of the early embryo have the potential to be reprogrammed into haematopoietic cells. These findings have implications for the generation of transplantable HSCs *in vitro* from ESCs or patient-specific iPS cells, which is so far dependent on genetic manipulations (Lis et al., 2017; Sugimura et al., 2017). Even though HE-like cells can be generated *in vitro*, we propose that one will need first to identify conditions that specifically generate endothelial cells of the aortic lineage of the trunk region (Ciau-Uitz and Patient, 2016) and then to induce differentiation into embryonic HE-like cells that can give rise to HSCs.

- Chapter 7 -

Appendix

Appendix 1: Sequence alignment between zebrafish runx1 cDNA sequences.

Clustal Omega *runx1* cDNA alignment for GeneBank files AF391125 and AB043787.

```
AF391125        ------------------------------------------------------------
AB043787        CGGCAGCCCCGGTACATCTAAGGGCATCACAGACCTGTTATTGCTCCATCTCGCGTGGCT 60

AF391125        ------------------------------------------------------------
AB043787        GAACGCCACTTGTCCCTCTAAGAAGTTGGGACGCCGACCGCGCGGGGCCGCGTAACTATT 120

AF391125        ------------------------------------------------------------
AB043787        TAGCATGCCGGAGATTCATTCACAGCGCATCATCTGATCCAGGAACAGCACTTTCTCTCA 180

AF391125        ------------------------------------------------------------
AB043787        CTGGTCCTTCTTTTTTAGTCTTTTTAACCCAGCTAGTACTTTTGTAGTCAGTTAAAGAAA 240

AF391125        ------------------------------------------------------------
AB043787        GCGAAAGTGTTTGCGTCAGGTGCGCGCTCTAACTTTCTAAAGGATACGCGCGTGGCCGAT 300

AF391125        ------------------------------------------------------------
AB043787        GATGATTTCTGCTGAAAGAGGAACAGCATCTGCGTAAAAACTGATCTGAGTTCAGCCCAC 360

AF391125        ----------------GG----------------------CACGAGCCGAAACTCACGGAGA 24
AB043787        TTCACACCGAACAGCAGGGTTGTTAAGAGCTGTGGATCCCCATCCCGAAACTCACGGAGA 420
                              **                    *.*.: ***************

AF391125        CAAAAACATACCAAATGGTTTTTCTTTGGGACGCCAAATACGAACCTGCGCCCGGGCGAC 84
AB043787        CAAAAACATACCAAATGGTTTTTCTTTGGGACGCCAAATACGAACCTGCGCCCGGGCGAC 480
                ************************************************************

AF391125        GCTTCACGCCTCCCTCCACCACCCTGAGCTCGGGGAAGATGAGCGAGGGTTTGCCTCTGG 144
AB043787        GCTTCACGCCTCCCTCCACCACCCTGAGCTCGGGGAAGATGAGCGAGGGTTTGCCCCTGG 540
                ************************************************************ ****

AF391125        GGGCTCAGGAGAGCACCGGCGCCGCTCTGGTGGGCAAACTGCGCATGGCAGACCGCAGCA 204
AB043787        GGGCTCAGGAGAGCACCGGCGCCGCTCTGGTGGGCAAACTGCGCATGGCAGACCGCAGCA 600
                ************************************************************

AF391125        TGGTGGAGGTGCTGTCCGACCACCCGGGGGAACTGGTGCGCACCGACAGCCCGAACTTCC 264
AB043787        TGGTGGAAGTACTGTCCGACCACCCAGGGGAACTGGTGCGCACCGACAGCCCGAACTTCC 660
                *******.**.**************.**********************************

AF391125        TCTGCTCCGTCCTGCCGACACACTGGCGCTGCAACAAGACCCTGCCCATCGCCTTCAAGG 324
AB043787        TCTGCTCCGTCCTGCCGACACACTGGCGCTGCAACAAGACCCTGCCCATCGCCTTCAAGG 720
                ************************************************************

AF391125        TGGTTGCCCTTGGTGATATCCCAGATGGAACACTGGTCACAGTGATGGCGGGAAATGATG 384
AB043787        TGGTTGCCCTTGGTGATATCCCAGACGGGACACTGGTCACAGTGATGGCGGGAAATGATG 780
                ************************* **.******************************

AF391125        AGAATTACTCCGCTGAGCTCCGCAACGCTACAGCCGCCATCAAGAACCAAGTGGCTCGTT 444
AB043787        AGAATTACTCCGCTGAGCTCCGCAACGCTACAGCCGCCATCAAGAACCAAGTGGCTCGTT 840
                ************************************************************

AF391125        TCAACGACCTGCGTTTCGTGGGCCGCAGCGGACGAGGGAAGAGCTTCACTCTCACCATCA 504
AB043787        TCAACGACCTGCGTTTCGTGGGCCGCAGCGGACGAGGGAAGAGCTTCACTCTCACCATCA 900
                ************************************************************

AF391125        CCGTCTTCACAAACCCTCCTCAAGTGGCCACATATCAGAGAGCCATAAAGATCACAGTGG 564
AB043787        CCGTCTTCACAAACCCTCCTCAAGTGGCCACATATCAGAGAGCCATAAAGATCACAGTGG 960
                ************************************************************

AF391125        ACGGACCCCGAGAGCCACGACGCCACCGACAGAAGCCGGATGAAGCAGTAAAGCCAGGCG 624
AB043787        ACGGACCCCGAGAGCCACGACGCCACCGACAGAAGCCGGATGAAGCAGTAAAGCCAGGCG 1020
                ************************************************************

AF391125        CGTTAGCTTTCTCTGAGCAGTTGAGGCGAAGTGCCATGCGGTGCAGCCCACACCACGGCC 684
AB043787        CGTTAGCTTTCTCTGAGCAGTTGAGGCGAAGTGCCATGCGGTGCAGCCCACACCACGGCC 1080
                ************************************************************

AF391125        CCGCCCCCAACACACGCCCCACCCTCAACACGCCCCCCTTCGGCAGCCCCGCCCACAGCC 744
AB043787        CCGCCCCCAACACACGCCCCACCCTCAACACGCCCCCCTTCGGCAGCCCCGCCCACAGCC 1140
```

```
                ************************************************************

AF391125        AGATTCCCGACTCCCGGCAGATGCAGACGTCTCCATCCTGGTCGTATGAGCAGTCGTATC 804
AB043787        AGATTCCCGACTCCCGGCAGATGCAGACGTCTCCATCCTGGTCGTATGAGCAGTCGTACC 1200
                ************************************************************ *

AF391125        CATATCTGGGCCCGATCTCCACCCCAGCGGTCCACCCTACAACACCAATCTCACCCAACC 864
AB043787        CATATCTGGGCCCGATCTCCACCCCAGCGGTCCACCCTACAACACCAATCTCACCCAACC 1260
                ************************************************************

AF391125        GGACCGCTTTACACTGTCCAGAGCTGACGGCGTTCACAGATCCCCGTGTGGGTTTGGAGC 924
AB043787        GGACCGCTTTACACTGTCCAGAGCTGACGGCGTTCACAGATCCCCGTGTGGGTTTGGAGC 1320
                ************************************************************

AF391125        GCTCCTTCCCATCACTCCCATCTCTCCCAGATGGGCGTTTCTCTGACCCGCGGGTGCCGT 984
AB043787        GCTCCTTCCCATCACTCCCATCTCTCCCAGATGGGCGTTTCTCTGACCCGCGGGTGCCGT 1380
                ************************************************************

AF391125        ACCCCACCGGCGCCTTCACCTACACCCCCACGCCCGTCACCAGCGCCATCGGTATTGGCA 1044
AB043787        ACCCCACCGGCGCCTTCACCTACACCCCCACGCCCGTCACCAGCGCCATCGGCATTGGCA 1440
                ************************************************** *******

AF391125        TGTCAGCCATGAGCAGCCCTGCGGGACGCTACCACACATACCTGCCGCCGGCGTACCCCG 1104
AB043787        TGTCAGCCATGAGCAGCCCTGCGGGACGCTACCACACATACCTGCCGCCGGCGTACCCCG 1500
                ************************************************************

AF391125        CGGGCTCCTCGCAGGCCCAGGCTGGAGCCTTCCAGGCGAGCTCGTCCCCATATCACCTGT 1164
AB043787        CGGGCTCCTCGCAGGCCCAGGCTGGAGCCTTCCAGGCGAGCTCGTCCCCATATCACCTGT 1560
                ************************************************************

AF391125        ACTACAGCAGCGCCGCCGGCTCCTACCAGTTCTCCATGATGCCCAGCGGAGGGGCGGCAG 1224
AB043787        ACTACAGCAGCGCCGCCGGCTCCTACCAGTTCTCCATGATGCCCAGCGGAGGGGCGGCAG 1620
                ************************************************************

AF391125        CAGGGGAGCGCTCGCCGCCCCGAATCCTACCCTGCACCAACGCCTCCACAGGCTCCGCCC 1284
AB043787        CAGGGGAGCGCTCGCCGCCCCGAATCCTACCCTGCACCAACGCCTCCACAGGCTCCGCCC 1680
                ************************************************************

AF391125        TCCTGCACCCCTCGCTCCCCAATCAGAGCGAAGGAGTGGTGGAGGCGGAGGGAAGCCACA 1344
AB043787        TCCTGCACCCCTCGCTCCCCAATCAGAGCGAAGGAGTGGTGGAGGCGGAGGGAAGCCACA 1740
                ************************************************************

AF391125        GCAGTTCGCCAACCAGCATGTCTGTAGAGGCCGTCTGGAGGCCATACTGACAGCCAATAG 1404
AB043787        GCAGTTCGCCAACCAGCATGTCTGTAGAGGCCGTCTGGAGGCCATACTGACAGCCAATAG 1800
                ************************************************************

AF391125        CACGCCGGCAGCTCATTTAAAGGGACAGTTCACCCAAAAATGTACTTACTATTTTCAAAT 1464
AB043787        CACGCCGGCAGCTCATTTAAAGGGACAGTTCACCCAAAAATGTACTTACTATTTCCAAAT 1860
                ************************************************** *****

AF391125        CTGTGAGTTTCTTTCTTCAGTTGAACACTAGAGAAGATAAACTAAAGAAAGCAGAAATCT 1524
AB043787        CTGTGAGTTTCTTTCTTCAGTTGAACACTAGAGGAGATATACTAAAGAAAGCAGAGATCT 1920
                ********************************.*****:***************.****

AF391125        GTTGACGTCCATAGTAG-AAAAACACATAAAAGTCAATTGATTTGACATGCTTTAGACTT 1583
AB043787        GTTGAGTTCCATAGTAGAAAAAACACATGAAAGTCAATTGATTTGACATGCTTTAGACTT 1980
                *****  ********** **********.******************************

AF391125        TATTTTGAAACAAATTAAATTAAATCAAATTATAAAAACAGCCAGAAAAGAGAAACTGAT 1643
AB043787        TATTTTG-AACAAATTAAATTAAATCAAAAAAAAAAAAA---------------------- 2018
                ******* **********************::*:*****.

AF391125        GTGTTACTGTCATCATTAATGCCCCATCAGTGCATTTCTATGCCTTTTCTATGGTTTCTT 1703

AB043787        ------------------------------------------------------------


AF391125        TTCTAATTTGCATGTTAAGTGCAATTAAATTCTAAAAATAAAATGTAAAGTCAAAGCATA 1763
AB043787        ------------------------------------------------------------


AF391125        AATAGGATATATGCACAGCTAGTGTTTCCTCCCATACTGATAATCTTCCGGTGAACGGTT 1823
AB043787        ------------------------------------------------------------


AF391125        AATATGACTTTTTTCAGATCTATACAGTTTATTTTACGGCATTAATGTTGTAATGCAATT 1883
AB043787        ------------------------------------------------------------
```

```
AF391125        AAAATACAATCAGGTAAACAGATTTAAGCATTTATTTAGTTAATCAAGCATAAACCGTGA 1943
AB043787        ------------------------------------------------------------

AF391125        TGAAAAGTCACGCACATGCACGTCACAATCAGCGGGCGAGCGGAGAAACTCCATTGAAAA 2003
AB043787        ------------------------------------------------------------

AF391125        CACTGGGGGAAAATAATGTTTATATTTTAAAGACATGATGTGGAAAAATGTCATTTAATG 2063
AB043787        ------------------------------------------------------------

AF391125        CAGTGCTTCTTGTAGATTGCGAGACACACTTTATATAAATATAAAAGCAGACACTTATGC 2123
AB043787        ------------------------------------------------------------

AF391125        TCTAAAATGTCGTCACTAATGACTAAATCACTGATGAATATCCACAAGATAAAGGTAAAC 2183
AB043787        ------------------------------------------------------------

AF391125        ATCGCTGAAGTGTAGCTCTGACAGGAACTTGAGTTATTCTTTTCCTTAGTTTTGCTTAAT 2243
AB043787        ------------------------------------------------------------

AF391125        ATTAATATTGCTGGCAGCGCGGTGGCGCAGGGGGTAGCAAAGTCACCTTACAGCAAGATG 2303
AB043787        ------------------------------------------------------------

AF391125        GTCGCTGGTTCGAGCCTCCGCTGGGTCAGTTGGCATTTCTGTGTGGAGTTTGCATTTGGA 2363
AB043787        ------------------------------------------------------------

AF391125        CCCCACAGTCCAAACACATGCGCTATAGGGGAATTGGGTAAGCTAGATTGGCCATGGTGT 2423
AB043787        ------------------------------------------------------------

AF391125        ATGAGGGAGTTTCCCAGCGATGGGTTGTAGCTGGAAGGACATCCGCTGCGTAAAACATAT 2483
AB043787        ------------------------------------------------------------

AF391125        GCTGGATAAGTTGACCCCAAAATAAAAAGAAACAAAATGCCAAAAAAAAAAAAAAAAAAA 2543
AB043787        ------------------------------------------------------------
```

Appendix 2: Sequence alignments between zebrafish Runx1 protein sequences.

Clustal Omega Runx1 protein sequence alignment encoded by the GeneBank files AF391125 and AB043787

```
runx1_AF391125.1        MVFLWDAKYEPAPGRRFTPPSTTLSSGKMSEGLPLGAQESTGAALVGKLRMADRSMVEVL 60
runx1_AB043787.1        MVFLWDAKYEPAPGRRFTPPSTTLSSGKMSEGLPLGAQESTGAALVGKLRMADRSMVEVL 60
                        ************************************************************

runx1_AF391125.1        SDHPGELVRTDSPNFLCSVLPTHWRCNKTLPIAFKVVALGDIPDGTLVTVMAGNDENYSA 120
runx1_AB043787.1        SDHPGELVRTDSPNFLCSVLPTHWRCNKTLPIAFKVVALGDIPDGTLVTVMAGNDENYSA 120
                        ************************************************************

runx1_AF391125.1        ELRNATAAIKNQVARFNDLRFVGRSGRGKSFTLTITVFTNPPQVATYQRAIKITVDGPRE 180
runx1_AB043787.1        ELRNATAAIKNQVARFNDLRFVGRSGRGKSFTLTITVFTNPPQVATYQRAIKITVDGPRE 180
                        ************************************************************

runx1_AF391125.1        PRRHRQKPDEAVKPGALAFSEQLRRSAMRCSPHHGPAPNTRPTLNTPPFGSPAHSQIPDS 240
runx1_AB043787.1        PRRHRQKPDEAVKPGALAFSEQLRRSAMRCSPHHGPAPNTRPTLNTPPFGSPAHSQIPDS 240
                        ************************************************************

runx1_AF391125.1        RQMQTSPSWSYEQSYPYLGPISTPAVHPTTPISPNRTALHCPELTAFTDPRVGLERSFPS 300
runx1_AB043787.1        RQMQTSPSWSYEQSYPYLGPISTPAVHPTTPISPNRTALHCPELTAFTDPRVGLERSFPS 300
                        ************************************************************

runx1_AF391125.1        LPSLPDGRFSDPRVPYPTGAFTYTPTPVTSAIGIGMSAMSSPAGRYHTYLPPAYPAGSSQ 360
runx1_AB043787.1        LPSLPDGRFSDPRVPYPTGAFTYTPTPVTSAIGIGMSAMSSPAGRYHTYLPPAYPAGSSQ 360
                        ************************************************************

runx1_AF391125.1        AQAGAFQASSSPYHLYYSSAAGSYQFSMMPSGGAAAGERSPPRILPCTNASTGSALLHPS 420
runx1_AB043787.1        AQAGAFQASSSPYHLYYSSAAGSYQFSMMPSGGAAAGERSPPRILPCTNASTGSALLHPS 420
                        ************************************************************

runx1_AF391125.1        LPNQSEGVVEAEGSHSSSPTSMSVEAVWRPY 451
runx1_AB043787.1        LPNQSEGVVEAEGSHSSSPTSMSVEAVWRPY 451
                        ********************************
```

Appendix 3: Runx1 cDNA sequence comparison across species

Shown are the coding parts of the cDNA sequences of the P2 driven *Runx1b* transcript from human (*RUNX1* 201 transcript; Ensembl), mouse (*Runx1* 201 transcript, Ensembl) and zebrafish zebrafish *runx1* DNA sequence (RefSeq:NM_131603). Red: P2-specific sequence. Yellow: in-exon splice site. Green: Runt domain.

human *RUNX1* DNA sequence (1362 bp) - (*RUNX1* 201 transcript; Ensembl)

ATGCGTATCCCCGTAGATGCCAGCACGAGCCGCCGCTTCACGCCGCCTTCCACCGCGCTGAGCCCAGGCAAGATGAGCGAGGCGTTGCCGCT
GGGCGCCCCGGACGCCGGCGCTGCCCTGGCCGGCAAGCTGAGGAGCGGCGACCGCAGCATGGTGGAGGTGCTGGCCGACCACCCGGGCGAGC
TGGTGCGCACCGACAGCCCCAACTTCCTCTGCTCCGTGCTGCCTACGCACTGGCGCTGCAACAAGACCCTGCCCATCGCTTTCAAGGTGGTG
GCCCTAGGGGATGTTCCAGATGGCACTCTGGTCACTGTGATGGCTGGCAATGATGAAAACTACTCGGCTGAGCTGAGAAATGCTACCGCAGC
CATGAAGAACCAGGTTGCAAGATTTAATGACCTCAGGTTTGTCGGTCGAAGTGGAAGAGGGAAAAGCTTCACTCTGACCATCACTGTCTTCA
CAAACCCACCGCAAGTCGCCACCTACCACAGAGCCATCAAAATCACAGTGGATGGGCCCCGAGAACCTCGAAGACATCGGCAGAAACTAGAT
GATCAGACCAAGCCCGGGAGCTTGTCCTTTTCCGAGCGGCTCAGTGAACTGGAGCAGCTGCGGCGCACAGCCATGAGGGTCAGCCCACACCA
CCCAGCCCCCACGCCCAACCCTCGTGCCTCCCTGAACCACTCCACTGCCTTTAACCCTCAGCCTCAGAGTCAGATGCAGGATACAAGGCAGA
TCCAACCATCCCCACCGTGGTCCTACGATCAGTCCTACCAATACCTGGGATCCATTGCCTCTCCTTCTGTGCACCCAGCAACGCCCATTTCA
CCTGGACGTGCCAGCGGCATGACAACCCTCTCTGCAGAACTTTCCAGTCGACTCTCAACGGCACCCGACCTGACAGCGTTCAGCGACCCGCG
CCAGTTCCCCGCGCTGCCCTCCATCTCCGACCCCCGCATGCACTATCCAGGCGCCTTCACCTACTCCCCGACGCCGGTCACCTCGGGCATCG
GCATCGGCATGTCGGCCATGGGCTCGGCCACGCGCTACCACACCTACCTGCCGCCGCCCTACCCCGGCTCGTCGCAAGCGCAGGGAGGCCCG
TTCCAAGCCAGCTCGCCCTCCTACCACCTGTACTACGGCGCCTCGGCCGGCTCCTACCAGTTCTCCATGGTGGGCGGCGAGCGCTCGCCGCC
GCGCATCCTGCCGCCCTGCACCAACGCCTCCACCGGCTCCGCGCTGCTCAACCCCAGCCTCCCGAACCAGAGCGACGTGGTGGAGGCCGAGG
GCAGCCACAGCAACTCCCCCACCAACATGGCGCCCTCCGCGCGCCTGGAGGAGGCCGTGTGGAGGCCCTACTGA

mouse *Runx1* DNA sequence (1356 bp) - (*Runx1* 201 transcript, Ensembl)

ATGCGTATCCCCGTAGATGCCAGCACGAGCCGCCGCTTCACGCCGCCTTCCACCGCGCTGAGCCCCGGCAAGATGAGCGAGGCGCTGCCGCT
GGGCGCCCCGGATGGCGGCGCCGCCCTGGCCAGCAAGCTGAGGAGCGGCGACCGCAGCATGGTGGAGGTACTAGCTGACCACCCTGGCGAGC
TAGTGCGCACCGACAGCCCCAACTTCCTCTGCTCCGTGCTACCCACTCACTGGCGCTGCAACAAGACCCTGCCCATCGCTTTCAAGGTGGTG
GCACTGGGGGACGTCCCGGATGGCACTCTGGTCACCGTCATGGCAGGCAACGATGAAAACTACTCGGCAGAACTGAGAAATGCTACCGCGGC
CATGAAGAACCAGGTAGCGAGATTCAACGACCTCAGGTTTGTCGGGCGGAGCGGTAGAGGCAAGAGCTTCACTCTGACCATCACCGTCTTTA
CAAATCCGCCACAAGTTGCCACCTACCATAGAGCCATCAAAATCACAGTGGACGGCCCCCGAGAACCCCGAAGACATCGGCAGAAACTAGAT
GATCAGACCAAGCCCGGGAGTTTGTCCTTTTCCGAGCGGCTCAGTGAATTGGAGCAGCTGCGGCGCACGGCCATGAGGGTCAGCCCGCACCA
CCCAGCCCCCACGCCCAACCCTCGGGCCTCCTTGAACCACTCCACTGCCTTTAACCCTCAGCCTCAAAGTCAGATGCAGGATGCCAGGCAGA
TCCAGCCATCCCCACCGTGGTCCTATGACCAGTCCTACCAGTACCTGGGATCCATCACCTCTTCCTCTGTCCACCCAGCGACACCCATTTCA
CCCGGCCGTGCCAGCGGCATGACCAGCCTCTCTGCAGAACTTTCCAGTCGACTCTCAACGGCTCCGGACCTGACCGCCTTCGGCGACCCACG
CCAGTTCCCTACTCTGCCGTCCATCTCCGACCCGCGCATGCACTACCCAGGCGCCTTCACCTACTCGCCGCCCGTCACGTCGGGCATCGGCA
TCGGCATGTCAGCCATGAGCTCGGCCTCTCGCTACCACACCTACCTGCCGCCGCCCTACCCCGGCTCATCACAGGCGCAGGCCGGGCCCTTC
CAGACCGGCTCGCCCTCCTACCATCTATACTACGGCGCCTCGGCCGGTTCCTACCAGTTCTCCATGGTGGGCGGAGAGAGATCGCCCCCGCG
CATCCTGCCGCCCTGCACCAACGCATCCACCGGCGCCGCGCTGCTCAACCCCAGCCTCCCCAGCCAGAGCGACGTGGTGGAGACCGAGGGCA
GCCATAGCAACTCGCCCACCAACATGCCCCCCGCGCGCCTGGAGGAGGCCGTGTGGCGGCCCTACTGA

zebrafish *runx1* DNA sequence (1356 bp) - (RefSeq:NM_131603)

ATGGTTTTTCTTTGGGACGCCAAATACGAACCTGCGCCCGGGCGACGCTTCACGCCTCCCTCCACCACCCTGAGCTCGGGGAAGATGAGCGA
GGGTTTGCCTCTGGGGGCTCAGGAGAGCACCGGCGCCGCTCTGGTGGGCAAACTGCGCATGGCAGACCGCAGCATGGTGGAGGTGCTGTCCG
ACCACCCGGGGGAACTGGTGCGCACCGACAGCCCGAACTTCCTCTGCTCCGTCCTGCCGACACACTGGCGCTGCAACAAGACCCTGCCCATC
GCCTTCAAGGTGGTTGCCCTTGGTGATATCCCAGATGGAACACTGGTCACAGTGATGGCGGGAAATGATGAGAATTACTCCGCTGAGCTCCG
CAACGCTACAGCCGCCATCAAGAACCAAGTGGCTCGTTTCAACGACCTGCGTTTCGTGGGCCGCAGCGGACGAGGGAAGAGCTTCACTCTCA
CCATCACCGTCTTCACAAACCCTCCTCAAGTGGCCACATATCAGAGAGCCATAAAGATCACAGTGGACGGACCCCGAGAGCCACGACGCCAC
CGACAGAAGCCGGATGAAGCAGTAAAGCCAGGCGCGTTAGCTTTCTCTGAGCAGTTGAGGCGAAGTGCCATGCGGTGCAGCCCACACCACGG
CCCCGCCCCCAACACACGCCCCACCCTCAACACGCCCCCCTTCGGCAGCCCCGCCCACAGCCAGATTCCCGACTCCCGGCAGATGCAGACGT
CTCCATCCTGGTCGTATGAGCAGTCGTATCCATATCTGGGCCCGATCTCCACCCCAGCGGTCCACCCTACAACACCAATCTCACCCAACCGG
ACCGCTTTACACTGTCCAGAGCTGACGGCGTTCACAGATCCCCGTGTGGGTTTGGAGCGCTCCTTCCCATCACTCCCATCTCTCCCAGATGG
GCGTTTCTCTGACCCGCGGGTGCCGTACCCCACCGGCGCCTTCACCTACACCCCCACGCCCGTCACCAGCGCCATCGGTATTGGCATGTCAG
CCATGAGCAGCCCTGCGGGACGCTACCACACATACCTGCCGCCGGCGTACCCCGCGGGCTCCTCGCAGGCCCAGGCTGGAGCCTTCCAGGCG
AGCTCGTCCCCATATCACCTGTACTACAGCAGCGCCGCCGGCTCCTACCAGTTCTCCATGATGCCCAGCGGAGGGGCGGCAGCAGGGGAGCG
CTCGCCGCCCCGAATCCTACCCTGCACCAACGCCTCCACAGGCTCCGCCCTCCTGCACCCCTCGCTCCCCAATCAGAGCGAAGGAGTGGTGG
AGGCGGAGGGAAGCCACAGCAGTTCGCCAACCAGCATGTCTGTAGAGGCCGTCTGGAGGCCATACTGA

Percent Identity Matrix - created by Clustal2.1 - complete sequence

1: Zebrafish	100.00	69.86	68.72
2: Human	69.86	100.00	91.37
3: Mouse	68.72	91.37	100.00

Percent Identity Matrix - created by Clustal2.1 - Runt domain

1: Zebrafish	100.00	84.42	81.82
2: Human	84.42	100.00	89.61
3: Mouse	81.82	89.61	100.00

Appendix 4: RUNX1 protein sequence comparison across species

Shown are the amino-acid sequences of the P2 driven *Runx1b* transcript from human (*RUNX1* 201 transcript; Ensembl), mouse (*Runx1* 201 transcript, Ensembl) and zebrafish zebrafish *runx1* DNA sequence (RefSeq:NM_131603). Green: Runt domain.

human RUNX1 aa sequence (453 aa)

```
M R I P V D A S T S R R F T P P S T A L S P G K M S E A L P L G A P D A G A A L A G K L R S G
D R S M V E V L A D H P G E L V R T D S P N F L C S V L P T H W R C N K T L P I A F K V V A L
G D V P D G T L V T V M A G N D E N Y S A E L R N A T A A M K N Q V A R F N D L R F V G R S G
R G K S F T L T I T V F T N P P Q V A T Y H R A I K I T V D G P R E P R R H R Q K L D D Q T K
P G S L S F S E R L S E L E Q L R R T A M R V S P H H P A P T P N P R A S L N H S T A F N P Q
P Q S Q M Q D T R Q I Q P S P P W S Y D Q S Y Q Y L G S I A S P S V H P A T P I S P G R A S G
M T T L S A E L S S R L S T A P D L T A F S D P R Q F P A L P S I S D P R M H Y P G A F T Y S
P T P V T S G I G I G M S A M G S A T R Y H T Y L P P P Y P G S S Q A Q G G P F Q A S S P S Y
H L Y Y G A S A G S Y Q F S M V G G E R S P P R I L P P C T N A S T G S A L L N P S L P N Q S
D V V E A E G S H S N S P T N M A P S A R L E E A V W R P Y
```

mouse RUNX1 aa sequence (451 aa)

```
M R I P V D A S T S R R F T P P S T A L S P G K M S E A L P L G A P D G G A A L A S K L R S G
D R S M V E V L A D H P G E L V R T D S P N F L C S V L P T H W R C N K T L P I A F K V V A L
G D V P D G T L V T V M A G N D E N Y S A E L R N A T A A M K N Q V A R F N D L R F V G R S G
R G K S F T L T I T V F T N P P Q V A T Y H R A I K I T V D G P R E P R R H R Q K L D D Q T K
P G S L S F S E R L S E L E Q L R R T A M R V S P H H P A P T P N P R A S L N H S T A F N P Q
P Q S Q M Q D A R Q I Q P S P P W S Y D Q S Y Q Y L G S I T S S S V H P A T P I S P G R A S G
M T S L S A E L S S R L S T A P D L T A F G D P R Q F P T L P S I S D P R M H Y P G A F T Y S
P P V T S G I G I G M S A M S S A S R Y H T Y L P P P Y P G S S Q A Q A G P F Q T G S P S Y H
L Y Y G A S A G S Y Q F S M V G G E R S P P R I L P P C T N A S T G A A L L N P S L P S Q S D
V V E T E G S H S N S P T N M P P A R L E E A V W R P Y
```

zebrafish Runx1 aa sequence (451 aa)

```
M V F L W D A K Y E P A P G R R F T P P S T T L S S G K M S E G L P L G A Q E S T G A A L V G
K L R M A D R S M V E V L S D H P G E L V R T D S P N F L C S V L P T H W R C N K T L P I A F
K V V A L G D I P D G T L V T V M A G N D E N Y S A E L R N A T A A I K N Q V A R F N D L R F
V G R S G R G K S F T L T I T V F T N P P Q V A T Y Q R A I K I T V D G P R E P R R H R Q K P
D E A V K P G A L A F S E Q L R R S A M R C S P H H G P A P N T R P T L N T P P F G S P A H S
Q I P D S R Q M Q T S P S W S Y E Q S Y P Y L G P I S T P A V H P T T P I S P N R T A L H C P
E L T A F T D P R V G L E R S F P S L P S L P D G R F S D P R V P Y P T G A F T Y T P T P V T
S A I G I G M S A M S S P A G R Y H T Y L P P A Y P A G S S Q A Q A G A F Q A S S S P Y H L Y
Y S S A A G S Y Q F S M M P S G G A A A G E R S P P R I L P C T N A S T G S A L L H P S L P N
Q S E G V V E A E G S H S S S P T S M S V E A V W R P Y
```

Percent Identity Matrix - created by Clustal2.1 – complete sequence

1: Zebrafish	100.00	78.64	77.41
2: Human	78.64	100.00	96.23
3: Mouse	77.41	96.23	100.00

Percent Identity Matrix - created by Clustal2.1 – Runt domain

1: Human	100.00	100.00	96.88
2: Mouse	100.00	100.00	96.88
3: Zebrafish	96.88	96.88	100.00

Appendix 5: HE-specific genes with a log2FC > 1.7

Gene ID	Log2FC	Gene ID	Log2FC	Gene ID	Log2FC
gfi1ab	5.51	si:dkey-22a1.3	2.18	nle1	1.84
gfi1aa	5.29	polr1b	2.17	scg3	1.84
prkcha	4.37	hemk1	2.17	nars2	1.84
vwf	4.16	msto1	2.17	mettl20	1.83
dnmt3ba	4.13	pus3	2.15	helt	1.83
dachb	3.89	pvalb6	2.15	ttc27	1.83
zgc:198241	3.83	shdb	2.14	tyw3	1.82
cox8b	3.81	si:dkey-191g15.12	2.13	si:dkey-281i8.1	1.81
rom1b	3.72	ampd1	2.12	elp2	1.81
pmp22b	3.72	mybbp1a	2.11	amacr	1.80
irf1b	3.56	selp	2.10	pex11b	1.80
si:ch211-214p16.2	3.45	ptgdsb	2.10	trmt61a	1.80
zgc:171686	3.41	shq1	2.09	slc5a6b	1.80
si:dkey-253d23.9	3.41	itm2cb	2.09	creg2	1.79
kng1	3.36	zgc:101000	2.08	prepl	1.79
mpl	3.33	zgc:113162	2.07	znf296	1.78
blf	3.21	sardh	2.07	fth1b	1.78
zgc:194665	3.18	thop1	2.06	slc6a6b	1.77
mst1	3.09	c10h21orf59	2.04	zgc:194224	1.76
pygma	3.04	cyp46a1.2	2.04	mocs3	1.76
b2m	2.99	klf9	2.03	taf1b	1.76
klhl24b	2.92	si:ch73-208g10.1	2.03	oclna	1.76
skap1	2.91	gstcd	2.02	prmt3	1.74
si:ch211-214p16.1	2.85	garnl3	2.00	apex2	1.74
esr1	2.82	ppp1r1c	2.00	tfb1m	1.73
tnni1al	2.78	eif4a2	2.00	wbscr22	1.72
zgc:175248	2.77	fgf12b	1.99	dcaf17	1.71
lcmt2	2.76	mcoln2	1.98	pdcd2l	1.71
zgc:136858	2.64	akt3a	1.98	tyw5	1.71
psme2	2.64	mettl8	1.96	rtn4ip1	1.70
zgc:92880	2.62	ugp2a	1.96	rbm38	1.70
zgc:195081	2.61	akap1b	1.96		
sox17	2.61	rab3da	1.96		
dnmt3bb.1	2.60	znhit3	1.96		
qrsl1	2.44	c19h1orf109	1.95		
cbsa	2.42	bcat1	1.94		
gata2b	2.38	diexf	1.94		
alpl	2.38	pfas	1.94		
polr3gla	2.35	fam210a	1.91		
myb	2.34	gart	1.91		
zgc:103678	2.31	cox6b1	1.89		
ngrn	2.29	lyve1a	1.89		
zgc:153521	2.26	hgh1	1.89		
tnnt2d	2.24	pwp2h	1.88		
zcchc4	2.23	trmt5	1.88		
si:dkey-188i13.10	2.22	smyd3	1.87		
fam212ab	2.20	npy8br	1.86		
pax5	2.19	prr5	1.85		
fam167ab	2.19	hbaa1	1.85		
fzd8b	2.19	enpp6	1.85		

Appendix 6: Top 100 *runx1* MO down-regulated (Runx1-activated) genes

Gene ID	Log2FC	Gene ID	Log2FC
angpt1	4.36	rhbdl3	1.70
olfm2a	3.73	b3glcta	1.69
mpl	3.61	rrbp1b	1.69
crfb15	3.44	agap2	1.68
gfi1ab	3.18	rapgef1b	1.66
myhz1.1	3.01	zgc:195282	1.65
ctssa	2.84	rorab	1.63
mdm1	2.80	ift80	1.63
col5a3a	2.79	ccr9a	1.62
cpa5	2.72	notch1b	1.61
zgc:158868	2.58	gpm6aa	1.61
mapk4	2.54	si:ch211-222k6.3	1.59
pigrl2.1	2.53	mfsd2aa	1.59
irf1b	2.52	vegfab	1.58
selp	2.43	etnk2	1.58
zgc:113209	2.40	scg3	1.57
atp1b2a	2.39	rab3da	1.57
zgc:152774	2.36	si:ch211-214p16.1	1.53
psma6l	2.34	nucb2b	1.52
cfh	2.31	tnnt1	1.52
eomesa	2.28	syk	1.52
mrpl28	2.21	aqp3a	1.51
trove2	2.16	fam117bb	1.51
slc16a3	2.13	hsd17b12b	1.48
mtmr1a	2.13	cpm	1.47
ugt8	2.08	tox	1.46
slc25a47a	2.06	sox11a	1.46
lmf2a	2.04	si:rp71-1g18.13	1.45
pygmb	1.98	coro1b	1.45
marcksa	1.97	hps5	1.45
klhl31	1.91	zgc:162707	1.45
camk1db	1.91	acsl1a	1.44
scgn	1.90	smad3b	1.44
tnnc2	1.90	hacd3	1.42
pnp5b	1.88	slc1a4	1.42
si:dkey-97m3.1	1.87	ptn	1.40
bcl6a	1.86	atp1a3a	1.37
aldoab	1.85	lin7b	1.36
rab3ab	1.83	mdm4	1.36
bcat1	1.82	nme2b.2	1.36
dnmt3bb.1	1.81	bag5	1.35
aldh3a2a	1.80	si:dkey-10p5.7	1.35
dpf1	1.79	mat2ab	1.33
nat8l	1.78	myh11a	1.32
pigrl2.4	1.78	lef1	1.31
ppp2r2bb	1.76	brf1a	1.31
prkcbb	1.75	psme2	1.30
ptpro	1.75	skap1	1.30
dnmt3ab	1.75	znf395	1.30
mst1	1.72	capn2b	1.30

Appendix 7: Top 100 *runx1* MO up-regulated (Runx1-repressed) genes

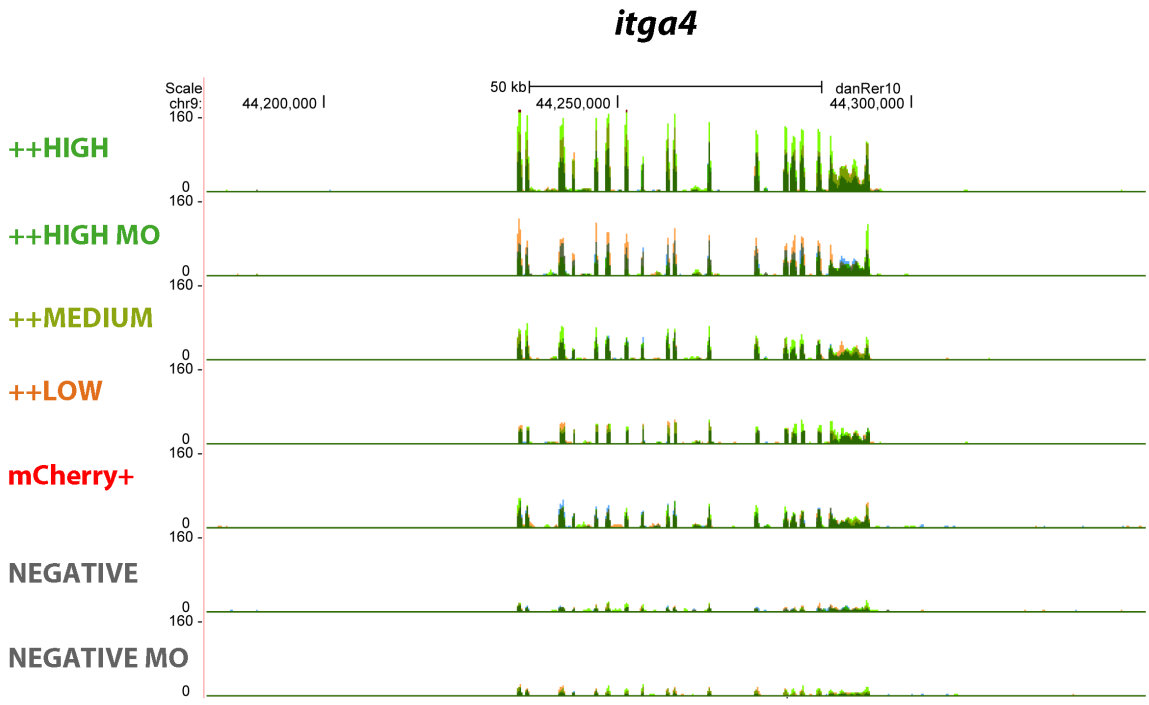
Gene ID	Log2FC	Gene ID	Log2FC
cyt1l	-6.58	LOC100149618	-2.54
anxa1c	-6.17	zgc:153184	-2.53
krt4	-5.81	pdgfab	-2.51
hamp	-4.97	zgc:92066	-2.51
anxa2a	-4.66	ctssb.2	-2.50
trpv1	-4.44	si:dkey-226m8.7	-2.50
si:ch211-117m20.5	-4.14	mir451a	-2.48
ccl34a.4	-4.07	si:ch211-79k12.1	-2.47
irg1l	-4.04	stk25a	-2.46
mafba	-3.92	si:ch211-106a19.1	-2.46
slc43a2b	-3.88	fam26f	-2.44
havcr1	-3.82	oclnb	-2.43
agr1	-3.81	dennd5b	-2.42
mafbb	-3.76	lygl1	-2.42
ch25h	-3.54	ptgs2a	-2.40
ccr12a	-3.52	cyp1d1	-2.39
zgc:173915	-3.49	zgc:77517	-2.36
mfap4	-3.41	gsk3aa	-2.36
lgals2a	-3.36	slc30a1a	-2.35
sema6dl	-3.33	pld1a	-2.33
cd9b	-3.32	arid3c	-2.31
cyp2k16	-3.16	apela	-2.27
zgc:175128	-3.16	zgc:109934	-2.26
ms4a17a.9	-3.13	lmo1	-2.24
crhbp	-3.12	cxcl12b	-2.21
entpd2a.1	-3.08	meox1	-2.21
csf1ra	-3.04	zgc:153284	-2.21
hmox1a	-2.97	gp1bb	-2.21
zgc:136870	-2.95	sox32	-2.19
zgc:158343	-2.94	prickle1a	-2.19
penka	-2.94	tmem11	-2.18
mir144	-2.93	cabp2a	-2.18
irf8	-2.90	si:dkey-33i11.3	-2.18
abcg2a	-2.88	zgc:173594	-2.18
tfec	-2.85	tal2	-2.18
cx32.2	-2.76	penkb	-2.17
spint2	-2.74	extl2	-2.14
s100a10b	-2.74	eya2	-2.14
il6	-2.70	atoh8	-2.13
mpeg1.1	-2.65	zgc:101810	-2.13
zgc:194125	-2.65	hbbe2	-2.13
cmklr1	-2.65	zgc:198419	-2.13
angptl7	-2.64	ppp1r3cb	-2.12
zgc:171687	-2.64	glcci1	-2.12
zgc:110340	-2.63	abhd17c	-2.11
icn	-2.62	hspb8	-2.10
hoxd3a	-2.62	ca2	-2.09
ptgdsb	-2.58	syngr1a	-2.09
hoxb10a	-2.57	col1a2	-2.07
plk3	-2.55	aqp8a.1	-2.06

Appendix 8: List of 85 HE-specific Runx1-activated genes

abcf2a	hacd3	pik3r1	tp53i11a
acsl1a	hmga1a	pmp22b	uhrf1
adh5	hps5	ppp5c	uspl1
agap2	impdh1b	prkab1a	wdhd1
akt1	iqgap1	prkcbb	xbp1
angpt1	irf1b	psme2	xpc
ankrd10a	itm2cb	ptdss1a	zgc:152774
arhgap11a	kdm6bb	pygma	zgc:158604
atad5b	lpar5a	rab3da	zgc:92791
b3gnt2b	mcm5	rgs13	znf296
bcat1	mdm4	rhbdl3	
bcl6a	mki67	rrm1	
brf1a	mknk2b	scg3	
coro1b	mpl	selp	
cpm	mrpl28	sept9b	
cyfip2	mst1	si:ch211-214p16.1	
dnmt3ab	mthfr	si:ch211-226m16.2	
dnmt3ba	mybbp1a	si:ch73-208g10.1	
dnmt3bb.1	nat8l	skap1	
dnmt3bb.2	nme2b.2	slc27a4	
dpf1	parp1	stmn1a	
esr1	pbrm1l	syk	
ezh2	pdk2b	tagapb	
ezrb	pdxka	tfr1b	
gfi1ab	pik3cd	tkfc	

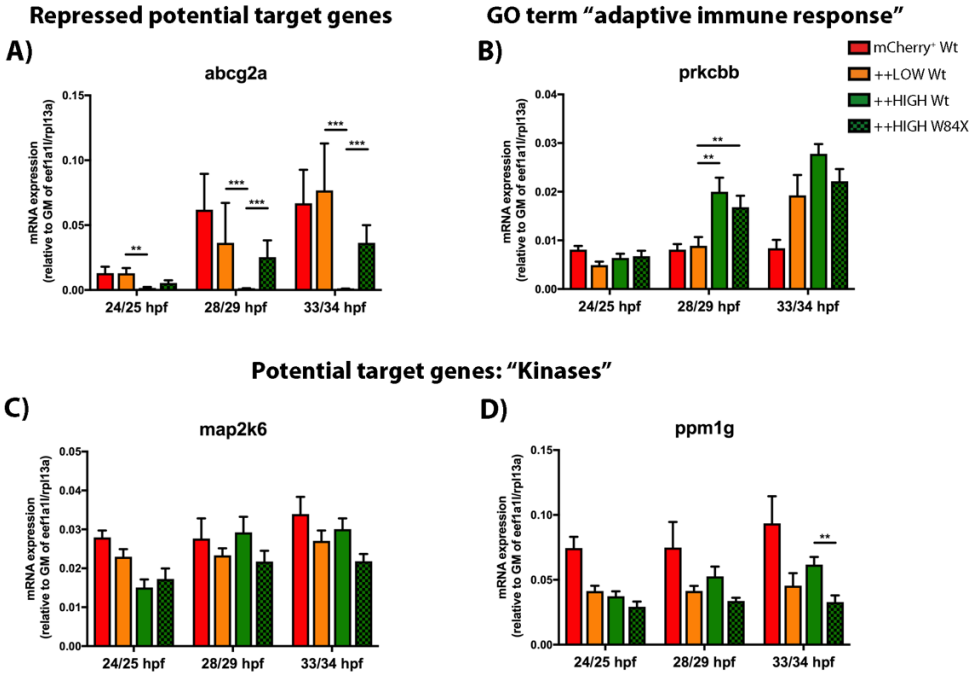
Appendix 9: UCSC gene tracks for *itga4*.

Gene track shows the region encoding the zebrafish *itga4* gene on Chromosome 9, together with the mapped reads of the RNAseq experiment. For each population, the triplicates are overlaid.



Appendix 10:Gene expression analysis of additional potential Runx1 target genes abcg2a, prkcbb, map2k6 and ppm1g.

qPCR analysis was performed using TaqMan probes on the BioMark platform (Fluidigm). Expression levels detected within the indicated cell populations sorted from *TgBAC(runx1P2:Citrine); Tg(kdrl:mCherry)* embryos are presented in relation to the geometric mean (GM) of the two housekeeping genes *eefia1l* and *rpll13a*. Statistical analysis (2-way ANOVA) was performed between the ++LOW (Wt) and ++HIGH (Wt and W84X) populations for each independent time point: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. Error-bars represent the standard error of the mean.



- Chapter 8 -

Bibliography

Abdollahi, A., Lord, K. A., Hoffman-Liebermann, B. and Liebermann, D. A. (1991). Interferon regulatory factor 1 is a myeloid differentiation primary response gene induced by interleukin 6 and leukemia inhibitory factor: role in growth inhibition. *Cell Growth Differ.* **2**, 401–407.

Abtahian, F. (2003). Regulation of Blood and Lymphatic Vascular Separation by Signaling Proteins SLP-76 and Syk. *Science (80-.).* **299**, 247–251.

Adamo, L. and García-Cardeña, G. (2012). The vascular origin of hematopoietic cells. *Dev. Biol.* **362**, 1–10.

Adamo, L., Naveiras, O., Wenzel, P. L., McKinney-Freeman, S., Mack, P. J., Gracia-Sancho, J., Suchy-Dicey, A., Yoshimoto, M., Lensch, M. W., Yoder, M. C., et al. (2009). Biomechanical forces promote embryonic haematopoiesis. *Nature* **459**, 1131–1135.

Adolfsson, J., Borge, O. J., Bryder, D., Theilgaard-Mönch, K., Åstrand-Grundström, I., Sitnicka, E., Sasaki, Y. and Jacobsen, S. E. W. (2001). Upregulation of Flt3 expression within the bone marrow Lin-Sca1+c-kit+ stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity* **15**, 659–669.

Ahn, D. G., Ruvinsky, I., Oates, A. C., Silver, L. M. and Ho, R. K. (2000). tbx20, a new vertebrate T-box gene expressed in the cranial motor neurons and developing cardiovascular structures in zebrafish. *Mech. Dev.* **95**, 253–258.

Aho, T. L. T., Sandholm, J., Peltola, K. J., Ito, Y. and Koskinen, P. J. (2006). Pim-1 kinase phosphorylates RUNX family transcription factors and enhances their activity. *BMC Cell Biol.* **7**, 21.

Alliot, F., Godin, I. and Pessac, B. (1999). Microglia derive from progenitors, originating from the yolk sac, and which proliferate in the brain. *Dev. Brain Res.* **117**, 145–152.

Allis, C. D. and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500.

Antas, V. I., Al-Drees, M. A., Prudence, A. J. A., Sugiyama, D. and Fraser, S. T. (2013). Hemogenic endothelium: A vessel for blood production. *Int. J. Biochem. Cell Biol.* **45**, 692–695.

Asou, H., Tashiro, S., Hamamoto, K., Otsuji, A., Kita, K. and Kamada, N. (1991). Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood* **77**, 2031–2036.

Bae, S. C. and Lee, Y. H. (2006). Phosphorylation, acetylation and ubiquitination: The molecular basis of RUNX regulation. *Gene* **366**, 58–66.

Bee, T., Ashley, E. L. K., Bickley, S. R. B., Jarratt, A., Li, P. S., Sloane-Stanley, J., Göttgens, B. and De Bruijn, M. F. T. R. (2009). The mouse Runx1 +23 hematopoietic stem cell enhancer confers hematopoietic specificity to both Runx1 promoters. *Blood* **113**, 5121–5214.

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

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.* **102**, 134–139.

Bertrand, J. Y., Kim, A. D., Violette, E. P., Stachura, D. L., Cisson, J. L. and Traver, D. (2007). Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. *Development* **134**, 4147–4156.

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

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

Bertrand, J. Y., Cisson, J. L., Stachura, D. L. and Traver, D. (2010b). Notch signaling distinguishes 2 waves of definitive hematopoiesis in the zebrafish embryo. *Blood* **115**, 2777–2783.

Biggs, J. R., Peterson, L. F., Zhang, Y., Kraft, A. S. and Zhang, D.-E. (2006). AML1/RUNX1 Phosphorylation by Cyclin-Dependent Kinases Regulates the Degradation of AML1/RUNX1 by the Anaphase-Promoting Complex. *Mol. Cell. Biol.* **26**, 7420–7429.

Böhmer, R., Neuhaus, B., Bühren, S., Zhang, D., Stehling, M., Böck, B. and Kiefer, F. (2010). Regulation of developmental lymphangiogenesis by Syk(+) leukocytes. *Dev. Cell* **18**, 437–49.

Böiers, C., Carrelha, J., Lutteropp, M., Luc, S., Green, J. C. C. A., Azzoni, E., Woll, P. S. S., Mead, A. J. J., Hultquist, A., Swiers, G., et al. (2013). Lymphomyeloid Contribution of an Immune-Restricted Progenitor Emerging Prior to Definitive Hematopoietic Stem Cells. *Cell Stem Cell*.

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.

Bollerot, K., Romero, S., Dunon, D. and Jaffredo, T. (2005). Core binding factor in the early avian embryo: cloning of Cbfbeta and combinatorial expression patterns with Runx1. *Gene Expr. Patterns* **6**, 29–39.

Bos, F. L., Hawkins, J. S. and Zovein, A. C. (2015). Single-cell resolution of morphological changes in hemogenic endothelium. *Development* **142**, 2719–2724.

Brend, T. and Holley, S. A. (2009). Zebrafish Whole Mount High-Resolution Double Fluorescent In Situ Hybridization. *JoVE* **25**, 3–5.

Bresciani, E., Carrington, B., Wincovitch, S., Jones, M., Gore, A. V., Weinstein, B. M., Sood, R. and Liu, P. P. (2014). CBFβ and RUNX1 are required at 2 different steps during the development of hematopoietic stem cells in zebrafish. *Blood* **124**, 70–78.

Burns, C. E., DeBlasio, T., Zhou, Y., Zhang, J., Zon, L. and Nimer, S. D. (2002). Isolation and characterization of runxa and runxb, zebrafish members of the runt family of transcriptional regulators. *Exp. Hematol.* **30**, 1381–1389.

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.

Burns, C. E., Galloway, J. L., Smith, A. C. H., Keefe, M. D., Cashman, T. J., Paik, E. J., Mayhall, E. A., Amsterdam, A. H. and Zon, L. I. (2009). A genetic screen in zebrafish defines a hierarchical network of pathways required for hematopoietic stem cell emergence. *Blood* **113**, 5776–5782.

Bussmann, J. and Schulte-Merker, S. (2011). Rapid BAC selection for tol2-mediated transgenesis in zebrafish. *Development* **138**, 4327–4332.

Butko, E., Distel, M., Pouget, C., Weijts, B., Kobayashi, I., Ng, K., Mosimann, C., Poulain, F. E., McPherson, A., Ni, C.-W., et al. (2015). Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. *Development* **142**, 1050–1061.

Butko, E., Pouget, C. and Traver, D. (2016). Complex regulation of HSC emergence by the Notch signaling pathway. *Dev. Biol.* **409**, 129–138.

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.

Cai, X., Gaudet, J. J., Mangan, J. K., Chen, M. J., De Obaldia, M. E., Oo, Z., Ernst, P. and Speck, N. A. (2011). Runx1 loss minimally impacts long-term hematopoietic stem cells. *PLoS One* **6**,.

Cai, X., Gao, L., Teng, L., Ge, J., Oo, Z. M., Kumar, A. R., Gilliland, D. G., Mason, P. J., Tan, K. and Speck, N. A. (2015). Runx1 Deficiency Decreases Ribosome Biogenesis and Confers Stress Resistance to Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell* **17**, 165–177.

Cameron, E. R. and Neil, J. C. (2004). The Runx genes: lineage-specific oncogenes and tumor suppressors. *Oncogene* **23**, 4308–4314.

Challen, G. A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J. S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2011). Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31.

Chantry, D., Vojtek, A., Kashishian, A., Holtzman, D. A., Wood, C., Gray, P. W., Cooper, J. A. and Hoekstra, M. F. (1997). p110d, a novel phosphatidylinositol 3-kinase catalytic subunit that associates with p85 and is expressed predominantly in leukocytes. *J. Biol. Chem.* **272**, 19236–19241.

Chen, A. T. and Zon, L. I. (2009). Zebrafish blood stem cells. *J. Cell. Biochem.* **108**, 35–42.

Chen, C. L., Broom, D. C., Liu, Y., De Nooij, J. C., Li, Z., Cen, C., Samad, O. A., Jessell, T. M., Woolf, C. J. and Ma, Q. (2006). Runx1 determines nociceptive sensory neuron phenotype and is required for thermal and neuropathic pain. *Neuron* **49**, 365–377.

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

Chen, M. J., Li, Y., De Obaldia, M. E., Yang, Q., Yzaguirre, A. D., Yamada-Inagawa, T., Vink, C. S., Bhandoola, A., Dzierzak, E. and Speck, N. A. (2011). Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell* **9**, 541–552.

Chen, I.-I., Caprioli, A., Ohnuki, H., Kwak, H., Porcher, C. and Tosato, G. (2016). EphrinB2 regulates the emergence of a hemogenic endothelium from the aorta. *Sci. Rep.* **6**, 27195–27211.

Chi, N. C., Shaw, R. M., De Val, S., Kang, G., Jan, L. Y., Black, B. L. and Stainier, D. Y. R. (2008). Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. *Genes Dev.* **22**, 734–739.

Choi, E., Kraus, M. R. C., Lemaire, L. A., Yoshimoto, M., Vemula, S., Potter, L. A., Manduchi, E., Stoeckert, C. J., Grapin-Botton, A. and Magnuson, M. A. (2012). Dual lineage-specific expression of Sox17 during mouse embryogenesis. *Stem Cells* **30**, 2297–2308.

Choorapoikayil, S., Kers, R., Herbomel, P., Kissa, K. and Den Hertog, J. (2014). Pivotal role of Pten in the balance between proliferation and differentiation of hematopoietic stem cells in zebrafish. *Blood* **123**, 184–190.

Chuang, L. S. H., Ito, K. and Ito, Y. (2013). RUNX family: Regulation and diversification of roles through interacting proteins. *Int. J. Cancer* **132**, 1260–1271.

Chung, M. I. S., Ma, A. C. H., Fung, T. K. and Leung, A. Y. H. (2011). Characterization of sry-related hmg box group f genes in zebrafish hematopoiesis. *Exp. Hematol.* **39**,.

Ciau-Uitz, A. and Patient, R. (2016). The embryonic origins and genetic programming of emerging haematopoietic stem cells. *FEBS Lett.* **590**, 4002–4015.

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

Ciau-Uitz, A., Liu, F. and Patient, R. (2010). Genetic control of hematopoietic development in Xenopus and zebrafish. *Int. J. Dev. Biol.* **54**, 1139–1149.

Ciau-Uitz, A., Wang, L., Patient, R. and Liu, F. (2013). ETS transcription factors in hematopoietic stem cell development. *Blood Cells. Mol. Dis.* **51**, 248–55.

Clark, K. J., Urban, M. D., Skuster, K. J. and Ekker, S. C. (2011). Transgenic zebrafish using transposable elements. *Methods Cell Biol.* **104**, 137–149.

Clarke, R. L., Yzaguirre, A. D., Yashiro-Ohtani, Y., Bondue, A., Blanpain, C., Pear, W. S., Speck, N. A. and Keller, G. (2013). The expression of Sox17 identifies and regulates haemogenic endothelium. *Nat. Cell Biol.* **15**, 502–510.

Clements, W. K. and Traver, D. (2013). Signalling pathways that control vertebrate haematopoietic stem cell specification. *Nat. Rev. Immunol.* **13**, 336–348.

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

Cooney, J. D., Hildick-Smith, G. J., Shafizadeh, E., McBride, P. F., Carroll, K. J., Anderson, H., Shaw, G. C., Tamplin, O. J., Branco, D. S., Dalton, A. J., et al. (2013). Teleost growth factor independence (gfi) genes differentially regulate successive waves of hematopoiesis. *Dev. Biol.* **373**, 431–441.

Corada, M., Orsenigo, F., Morini, M. F., Pitulescu, M. E., Bhat, G., Nyqvist, D., Breviario, F., Conti, V., Briot, A., Iruela-Arispe, M. L., et al. (2013). Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nat. Commun.* **4**,.

Corbel, C. and Salaün, J. (2002). AlphaIIb integrin expression during development of the murine hemopoietic system. *Dev. Biol.* **243**, 301–311.

Coskun, S. and Hirschi, K. K. (2010). Establishment and regulation of the HSC niche: Roles of osteoblastic and vascular compartments. *Birth Defects Res. Part C - Embryo Today Rev.* **90**, 229–242.

Craddock, C. F., Nakamoto, B., Andrews, R. G., Priestley, G. V and Papayannopoulou, T. (1997). Antibodies to VLA4 integrin mobilize long-term repopulating cells and augment cytokine-induced mobilization in primates and mice. *Blood* **90**, 4779–88.

Crane, G. M., Jeffery, E. and Morrison, S. J. (2017). Adult haematopoietic stem cell niches. *Nat. Rev. Immunol.*

Cumano, A. and Godin, I. (2007). Ontogeny of the hematopoietic system.pdf.

D'Alonzo, R. C., Selvamurugan, N., Karsenty, G. and Partridge, N. C. (2002). Physical interaction of the activator protein-1 factors c-Fos and c-Jun with Cbfa1 for collagenase-3 promoter activation. *J. Biol. Chem.* **277**, 816–822.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S. M., Phair, R. D. and Singer, R. H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806.

Davidson, A. J. and Zon, L. I. (2004). The "definitive" (and "primitive") guide to zebrafish hematopoiesis. *Oncogene* **23**, 7233–7246.

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

de Bruijn, M. and Dzierzak, E. (2017). Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood* **129**, 2061–2069.

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

de Jong, J. L. O. and Zon, L. I. (2005). Use of the Zebrafish System to Study Primitive and Definitive Hematopoiesis. *Annu. Rev. Genet.* **39**, 481–501.

Deane, J. A. and Fruman, D. A. (2004). P hosphoinositide 3-K inase : Diverse Roles in Immune Cell Activation. *Annu. Rev. Immunol.* **22**, 563–598.

de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P. S., Camper, S. A., Speck, N. and Dzierzak, E. (2013). *Gata2* is required for HSC generation and survival. *J. Exp. Med.* **210**, 2843–2850.

Detrich, H. W., Kieran, M. W., Chan, F. Y., Barone, L. M., Yee, K., Rundstadler, J. A., Pratt, S., Ransom, D. and Zon, L. I. (1995). Intraembryonic hematopoietic cell migration during vertebrate development. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 10713–7.

Ditadi, A., Sturgeon, C. M., Tober, J., Awong, G., Kennedy, M., Yzaguirre, A. D., Azzola, L., Ng, E. S., Stanley, E. G., French, D. L., et al. (2015). Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat. Cell Biol.* **17**, 580–591.

Dong, Z.-W., Ren, C.-G., Xia, Y., Su, D., Du, T.-T., Fan, H.-B., Yuan, H., Wang, L., Dong, M., Li, W.-C., et al. (2014). Pten regulates homeostasis and inflammation-induced migration of myelocytes in zebrafish. *J. Hematol. Oncol.* **7**, 17.

Dooley, K. and Zon, L. I. (2000). Zebrafish: a model system for the study of human disease. *Curr. Opin. Genet. Dev.* **10**, 252–256.

Dooley, K. A., Davidson, A. J. and Zon, L. I. (2005). Zebrafish scl functions independently in hematopoietic and endothelial development. *Dev. Biol.* **277**, 522–536.

Doulatov, S., Notta, F., Laurenti, E. and Dick, J. E. (2012). Hematopoiesis: A human perspective. *Cell Stem Cell* **10**, 120–136.

Dunne, J., Cullmann, C., Ritter, M., Soria, N. M., Drescher, B., Debernardi, S., Skoulakis, S., Hartmann, O., Krause, M., Krauter, J., et al. (2006). siRNA-mediated AML1/MTG8 depletion affects differentiation and proliferation-associated gene expression in t(8;21)-positive cell lines and primary AML blasts. *Oncogene* **25**, 6067–6078.

Durand, C., Robin, C., Bollerot, K., Baron, M. H., Ottersbach, K. and Dzierzak, E. (2007). Embryonic stromal clones reveal developmental regulators of definitive hematopoietic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20838–43.

Dussault, A.-A. and Pouliot, M. (2006). Rapid and simple comparison of messenger rna levels using real-time PCR. *Biol. Proced. Online* **8**, 1–10.

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

Edwards, H., Xie, C., LaFiura, K. M., Dombkowski, A. A., Buck, S. A., Boerner, J. L., Taub, J. W., Matherly, L. H. and Ge, Y. (2009). RUNX1 regulates phosphoinositide 3-kinase/AKT pathway: Role in chemotherapy sensitivity in acute megakaryocytic leukemia. *Blood* **114**, 2744–2752.

Eggers, J. H., Stock, M., Fliegauf, M., Vonderstrass, B. and Otto, F. (2002). Genomic characterization of the RUNX2 gene of Fugu rubripes. *Gene* **291**, 159–167.

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

Elagib, K. E., Racke, F. K., Mogass, M., Khetawat, R., Delehanty, L. L. and Goldfarb, A. N. (2003). RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood* **101**, 4333–4341.

Elcheva, I., Brok-Volchanskaya, V., Kumar, A., Liu, P., Lee, J.-H. H., Tong, L., Vodyanik, M., Swanson, S., Stewart, R., Kyba, M., et al. (2014). Direct induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators. *Nat Commun* **5**, 4372.

Ellertsdóttir, E., Lenard, A., Blum, Y., Krudewig, A., Herwig, L., Affolter, M. and Belting, H. G. (2010). Vascular morphogenesis in the zebrafish embryo. *Dev. Biol.* **341**, 56–65.

Espín-Palazón, R., Stachura, D. L., Campbell, C. A., García-Moreno, D., Del Cid, N., Kim, A. D., Candel, S., Meseguer, J., Mulero, V. and Traver, D. (2014). Proinflammatory signaling regulates hematopoietic stem cell emergence. *Cell* **159**, 1070–1085.

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

Ferreras, C., Lancrin, C., Lie-A-Ling, M., Kouskoff, V. and Lacaud, G. (2011). Identification and characterization of a novel transcriptional target of RUNX1/AML1 at the onset of hematopoietic development. *Blood* **118**, 594–597.

Fiolka, K., Hertzano, R., Vassen, L., Zeng, H., Hermesh, O., Avraham, K. B., Dührsen, U. and Möröy, T. (2006). Gfi1 and Gfi1b act equivalently in haematopoiesis, but have distinct, non-overlapping functions in inner ear development. *EMBO Rep.* **7**, 326–333.

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

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., et al. (2012). Ensembl 2012. *Nucleic Acids Res.* **40**,.

Fouquet, B., Weinstein, B. M., Serluca, F. C. and Fishman, M. C. (1997). Vessel patterning in the embryo of the zebrafish: guidance by notochord. *Dev. Biol.* **183**, 37–48.

Fragale, A., Gabriele, L., Stellacci, E., Borghi, P., Perrotti, E., Ilari, R., Lanciotti, A., Remoli, A. L., Venditti, M., Belardelli, F., et al. (2008). IFN regulatory factor-1 negatively regulates CD4+ CD25+ regulatory T cell differentiation by repressing Foxp3 expression. *J. Immunol.* **181**, 1673–82.

Frame, J. M., McGrath, K. E. and Palis, J. (2013). Erythro-myeloid progenitors: "Definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. *Blood Cells, Mol. Dis.* **51**, 220–225.

Frame, J. M., Fegan, K. H., Conway, S. J., McGrath, K. E. and Palis, J. (2016). Definitive Hematopoiesis in the Yolk Sac Emerges from Wnt-Responsive Hemogenic Endothelium Independently of Circulation and Arterial Identity. *Stem Cells* **34**, 431–444.

Fuentes, F., Reynolds, E., Lewellis, S. W., Venkiteswaran, G. and Knaut, H. (2016). A Plasmid Set for Efficient Bacterial Artificial Chromosome (BAC) Transgenesis in Zebrafish. *G3: Genes|Genomes|Genetics* **6**, 829–834.

Fukuhara, S., Sako, K., Noda, K., Zhang, J., Minami, M. and Mochizuki, N. (2010). Angiopoietin-1/Tie2 receptor signaling in vascular quiescence and angiogenesis. *Histol. Histopathol.* **25**, 387–396.

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

Gao, X. N., Yan, F., Lin, J., Gao, L., Lu, X. L., Wei, S. C., Shen, N., Pang, J. X., Ning, Q. Y., Komeno, Y., et al. (2015). AML1/ETO cooperates with HIF1α to promote leukemogenesis through DNMT3a transactivation. *Leukemia* **29**, 1730–1740.

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

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

Gering, M., Rodaway, A. R. F., Göttgens, B., Patient, R. K. and Green, A. R. (1998). The SCL gene specifies haemangioblast development from early mesoderm. *EMBO J.* **17**, 4029–4045.

Geudens, I., Herpers, R., Hermans, K., Segura, I., Ruiz De Almodovar, C., Bussmann, J., De Smet, F., Vandevelde, W., Hogan, B. M., Siekmann, A., et al. (2010). Role of delta-like-4/notch in the formation and wiring of the lymphatic network in zebrafish. *Arterioscler. Thromb. Vasc. Biol.* **30**, 1695–1702.

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**, 1935–40.

Giraldo, P. and Montoliu, L. (2001). Size matters: Use of YACs, BACs and PACs in transgenic animals. *Transgenic Res.* **10**, 83–103.

Goetz, T. L., Gu, T. L., Speck, N. A. and Graves, B. J. (2000). Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Mol. Cell. Biol.* **20**, 81–90.

Gomez Perdiguero, E., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M. F., Geissmann, F., et al. (2014). Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* **518**, 547–551.

Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie-A-Ling, M., Lilly, A. J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., et al. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev. Cell* **36**, 572–587.

Gordon-Keylock, S., Sobiesiak, M., Rybtsov, S., Moore, K. and Medvinsky, A. (2013). Mouse extraembryonic arterial vessels harbor precursors capable of maturing into definitive HSCs. *Blood* **122**, 2338–2345.

Gore, A. V., Athans, B., Iben, J. R., Johnson, K., Russanova, V., Castranova, D., Pham, V. N., Butler, M. G., Williams-Simons, L., Nichols, J. T., et al. (2016). Epigenetic regulation of hematopoiesis by DNA methylation. *Elife* **5**,.

Göttgens, B., Nastos, A., Kinston, S., Piltz, S., Delabesse, E. C. M., Stanley, M., Sanchez, M. J., Ciau-Uitz, A., Patient, R. and Green, A. R. (2002). Establishing the transcriptional programme for blood: The SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. *EMBO J.* **21**, 3039–3050.

Goyama, S., Yamaguchi, Y., Imai, Y., Kawazu, M., Nakagawa, M., Asai, T., Kumano, K., Mitani, K., Ogawa, S., Chiba, S., et al. (2004). The transcriptionally active form of AML1 is required for hematopoietic rescue of the AML1-deficient embryonic para-aortic splanchnopleural (P-Sp) region. *Blood* **104**, 3558–3564.

Goyama, S., Huang, G., Kurokawa, M. and Mulloy, J. C. (2015). Posttranslational modifications of RUNX1 as potential anticancer targets. *Oncogene* **34**, 3483–92.

Gratwohl, A., Baldomero, H., Aljurf, M., Marcelo, C., Bouzas, L. F., Yoshimi, A., Szer, J., Schwendener, A., Gratwohl, M., Ph, D., et al. (2011). Hematopoietic stem cell transplantation: A global perspective. *JAMA* **303**, 1617–1624.

Gribi, R., Hook, L., Ure, J. and Medvinsky, A. (2006). The differentiation program of embryonic definitive hematopoietic stem cells is largely α4 integrin independent. *Blood* **108**, 501–509.

Growney, J. D., Shigematsu, H., Li, Z., Lee, B. H., Adelsperger, J., Rowan, R., Curley, D. P., Kutok, J. L., Akashi, K., Williams, I. R., et al. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood* **106**, 494–504.

Gu, T. L., Goetz, T. L., Graves, B. J. and Speck, N. A. (2000). Auto-inhibition and partner proteins, core-binding factor beta (CBFbeta) and Ets-1, modulate DNA binding by CBFalpha2 (AML1). *Mol. Cell. Biol.* **20**, 91–103.

Guo, H. and Friedman, A. D. (2011). Phosphorylation of RUNX1 by cyclin-dependent kinase reduces direct interaction with HDAC1 and HDAC3. *J. Biol. Chem.* **286**, 208–215.

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.

Hale, G., Zhang, M. J., Bunjes, D., Prentice, H. G., Spence, D., Horowitz, M. M., Barrett, A. J. and Waldmann, H. (1998). Improving the outcome of bone marrow transplantation by using CD52 monoclonal antibodies to prevent graft-versus-host disease and graft rejection. *Blood* **92**, 4581–4590.

Hanai, J. I., Chen, L. F., Kanno, T., Ohtani-Fujita, N., Kim, W. Y., Guo, W. H., Imamura, T., Ishidou, Y., Fukuchi, M., Shi, M. J., et al. (1999). Interaction and functional cooperation of PEBP2/CBF with Smads. Synergistic induction of the immunoglobulin germline Cα promoter. *J. Biol. Chem.* **274**, 31577–31582.

Hannah, R., Joshi, A., Wilson, N. K., Kinston, S. and Gttgens, B. (2011). A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms. *Exp. Hematol.* **39**, 531–541.

Harada, K., Nobuhisa, I., Anani, M., Saito, K. and Taga, T. (2017). Thrombopoietin contributes to the formation and the maintenance of hematopoietic progenitor-containing cell clusters in the aorta-gonad-mesonephros region. *Cytokine* **95**, 35–42.

Harris, J. M., Esain, V., Frechette, G. M., Harris, L. J., Cox, A. G., Cortes, M., Garnaas, M. K., Carroll, K. J., Cutting, C. C., Khan, T., et al. (2013). Glucose metabolism impacts the spatiotemporal onset and magnitude of HSC induction in vivo. *Blood* **121**, 2483–2493.

He, Q., Zhang, C., Wang, L., Zhang, P., Ma, D., Lv, J. and Liu, F. (2015). Inflammatory signaling regulates hematopoietic stem and progenitor cell emergence in vertebrates. *Blood* **125**, 1098–1106.

Heintz, N. (2001). BAC to the future: the use of bac transgenic mice for neuroscience research. *Nat. Rev. Neurosci.* **2**, 861–870.

Herbomel, P., Thisse, B. and Thisse, C. (1999). Ontogeny and behaviour of early macrophages in the zebrafish embryo.

Development **126**, 3735–45.

Herbomel, P., Thisse, B. and Thisse, C. (2001). Zebrafish Early Macrophages Colonize Cephalic Mesenchyme and Developing Brain, Retina, and Epidermis through a M-CSF Receptor-Dependent Invasive Process. *Dev. Biol.* **238**, 274–288.

Hirai, H., Ogawa, M., Suzuki, N., Yamamoto, M., Breier, G., Mazda, O., Imanishi, J. and Nishikawa, S. I. (2003). *Hemogenic and nonhemogenic endothelium can be distinguished by the activity of fetal liver kinase (Flk)-1 promoter/enhancer during mouse embryogenesis*.

Hirai, H., Samokhvalov, I. M., Fujimoto, T., Nishikawa, S., Imanishi, J. and Nishikawa, S. I. (2005). Involvement of Runx1 in the down-regulation of fetal liver kinase-1 expression during transition of endothelial cells to hematopoietic cells. *Blood* **106**, 1948–1955.

Hirsch, E., Iglesias, a, Potocnik, a J., Hartmann, U. and Fässler, R. (1996). Impaired migration but not differentiation of haematopoietic stem cells in the absence of beta1 integrins. *Nature* **380**, 171–175.

Hirschi, K. K. (2012). Hemogenic endothelium during development and beyond. *Blood* **119**, 4823–4837.

Hofmeister, C. C., Zhang, J., Knight, K. L., Le, P. and Stiff, P. J. (2007). Ex vivo expansion of umbilical cord blood stem cells for transplantation: growing knowledge from the hematopoietic niche. *Bone Marrow Transpl.* **39**, 11–23.

Holz-Schietinger, C., Matje, D. M. and Reich, N. O. (2012). Mutations in DNA methyltransferase (DNMT3A) observed in acute myeloid leukemia patients disrupt processive methylation. *J. Biol. Chem.* **287**, 30941–30951.

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

Hsu, K., Traver, D., Kutok, J. L., Hagen, A., Liu, T. X., Paw, B. H., Rhodes, J., Berman, J. N., Zon, L. I., Kanki, J. P., et al. (2004). The pu.1 promoter drives myeloid gene expression in zebrafish. *Blood* **104**, 1291–1297.

Hu, W., Feng, Z. and Levine, A. J. (2012). The Regulation of Multiple p53 Stress Responses is Mediated through MDM2. *Genes Cancer* **3**, 199–208.

Huang, G., Shigesada, K., Ito, K., Wee, H. J., Yokomizo, T. and Ito, Y. (2001). Dimerization with PEBP2β protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. *EMBO J.* **20**, 723–733.

Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., Okuno, Y., Dayaram, T., Growney, J. D., et al. (2008a). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nat. Genet.* **40**, 51–60.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2008b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.

Huang, H.-T., Kathrein, K. L., Barton, A., Gitlin, Z., Huang, Y.-H., Ward, T. P., Hofmann, O., Dibiase, A., Song, A., Tyekucheva, S., et al. (2013). A network of epigenetic regulators guides developmental haematopoiesis in vivo. *Nat. Cell Biol.* **15**, 1516–1525.

Hug, B. A., Ahmed, N., Robbins, J. A. and Lazar, M. A. (2004). A chromatin immunoprecipitation screen reveals protein kinase Cbeta as a direct RUNX1 target gene. *J Biol Chem* **279**, 825–830.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J.-R. J. and Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229.

Ichikawa, M., Asai, T., Saito, T., Yamamoto, G., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Hirai, H., et al. (2004). AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat. Med.* **10**, 299–304.

Ichikawa, M., Goyama, S., Asai, T., Kawazu, M., Nakagawa, M., Takeshita, M., Chiba, S., Ogawa, S. and Kurokawa, M. (2008). AML1/Runx1 negatively regulates quiescent hematopoietic stem cells in adult hematopoiesis. *J. Immunol.* **180**, 4402–8.

Irion, S., Clarke, R. L., Luche, H., Kim, I., Morrison, S. J., Fehling, H.-J. and Keller, G. M. (2010). Temporal specification of blood progenitors from mouse embryonic stem cells and induced pluripotent stem cells. *Development* **137**, 2829–2839.

Ito, Y. (1999). Molecular basis of tissue-specific gene expression mediated by the runt domain transcription factor PEBP2/CBF. *Genes Cells* **4**, 685–96.

Ito, K., Lim, A. C. B., Salto-Tellez, M., Motoda, L., Osato, M., Chuang, L. S. H., Lee, C. W. L., Voon, D. C. C., Koo, J. K. W., Wang, H., et al. (2008). RUNX3 Attenuates β-Catenin/T Cell Factors in Intestinal Tumorigenesis. *Cancer*

Cell **14**, 226–237.

Ito, Y., Bae, S.-C. and Chuang, L. S. H. (2015). The RUNX family: developmental regulators in cancer. *Nat. Rev. Cancer* **15**, 81–95.

Jacob, B., Osato, M., Yamashita, N., Wang, C. Q., Taniuchi, I., Littman, D. R., Asou, N. and Ito, Y. (2010). Stem cell exhaustion due to Runx1 deficiency is prevented by Evi5 activation in leukemogenesis. *Blood* **115**, 1610–1620.

Jaffredo, T., Bollerot, K., Sugiyama, D., Gautier, R. and Drevon, C. (2005). Tracing the hemangioblast during embryogenesis: Developmental relationships between endothelial and hematopoietic cells. *Int. J. Dev. Biol.* **49**, 269–277.

Jin, S.-W. (2005). Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development* **132**, 5199–5209.

Jin, H., Sood, R., Xu, J., Zhen, F., English, M. a, Liu, P. P. and Wen, Z. (2009). Definitive hematopoietic stem/progenitor cells manifest distinct differentiation output in the zebrafish VDA and PBI. *Development* **136**, 647–654.

Jin, H., Li, L., Xu, J., Zhen, F., Zhu, L., Liu, P. P., Zhang, M., Zhang, W. and Wen, Z. (2012). Runx1 regulates embryonic myeloid fate choice in zebrafish through a negative feedback loop inhibiting Pu.1 expression. *Blood* **119**, 5239–5249.

Jing, L. and Zon, L. I. (2011). Zebrafish as a model for normal and malignant hematopoiesis. *Dis. Model. Mech.* **4**, 433–438.

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492.

Jonsen, M. D., Petersen, J. M., Xu, Q. P. and Graves, B. J. (1996). Characterization of the cooperative function of inhibitory sequences in Ets-1. *Mol. Cell. Biol.* **16**, 2065–2073.

Kalev-Zylinska, M. L., Horsfield, J. A., Flores, M. V. C., Postlethwait, J. H., Vitas, M. R., Baas, A. M., Crosier, P. S. and Crosier, K. E. (2002). Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. *Development* **129**, 2015–30.

Kamachi, Y., Ogawa, E., Asano, M., Ishida, S., Murakami, Y., Satake, M., Ito, Y. and Shigesada, K. (1990). Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer. *J Virol* **64**, 4808–4819.

Kanno, T., Kanno, Y., Chen, L. F., Ogawa, E., Kim, W. Y. and Ito, Y. (1998). Intrinsic transcriptional activation-inhibition domains of the polyomavirus enhancer binding protein 2/core binding factor alpha subunit revealed in the presence of the beta subunit. *Mol. Cell. Biol.* **18**, 2444–54.

Kartalaei, P. S., Yamada-Inagawa, T., Vink, C. S., de Pater, E., van der Linden, R., Marks-Bluth, J., van der Sloot, A., van den Hout, M., Yokomizo, T., van Schaick-Solernó, M. L., et al. (2015). Whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation. *J. Exp. Med.* **212**, 93–106.

Kataoka, H., Ochi, M., Enomoto, K. I. and Yamaguchi, A. (2000). Cloning and embryonic expression patterns of the zebrafish Runt domain genes, runxa and runxb. *Mech. Dev.* **98**, 139–143.

Kaufman, D. S., Hanson, E. T., Lewis, R. L., Auerbach, R. and Thomson, J. A. (2001). Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10716–21.

Kennedy, M., D'Souza, S. L., Lynch-Kattman, M., Schwantz, S. and Keller, G. (2007). Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. In *Blood*, pp. 2679–2687.

Kennedy, M., Awong, G., Sturgeon, C. M., Ditadi, A., LaMotte-Mohs, R., Zúñiga-Pflücker, J. C. and Keller, G. (2012). T Lymphocyte Potential Marks the Emergence of Definitive Hematopoietic Progenitors in Human Pluripotent Stem Cell Differentiation Cultures. *Cell Rep.* **2**, 1722–1735.

Kiel, M. J. and Morrison, S. J. (2006). Maintaining Hematopoietic Stem Cells in the Vascular Niche. *Immunity* **25**, 862–864.

Kiel, M. J., Yilmaz, Ö. H., Iwashita, T., Yilmaz, O. H., Terhorst, C. and Morrison, S. J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121.

Kierdorf, K., Erny, D., Goldmann, T., Sander, V., Schulz, C., Perdiguero, E. G., Wieghofer, P., Heinrich, A., Riemke, P., Hölscher, C., et al. (2013). Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat. Neurosci.* **16**, 273–280.

Kim, W. Y., Sieweke, M., Ogawa, E., Wee, H. J., Englmeier, U., Graf, T. and Ito, Y. (1999). Mutual activation of Ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. *EMBO J.* **18**, 1609–20.

Kim, N., Saudemont, A., Webb, L., Camps, M., Ruckle, T., Hirsch, E., Turner, M. and Colucci, F. (2007). The p110delta catalytic isoform of PI3K is a key player in NK-cell development and cytokine secretion. *Blood* **110**, 3202–3208.

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

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. and Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn. an Off. public* **203**, 253–310.

Kimura, S., Roberts, A. W., Metcalf, D. and Alexander, W. S. (1998). Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1195–1200.

Kina, T., Ikuta, K., Takayama, E., Wada, K., Majumdar, A. Sen, Weissman, I. L. and Katsura, Y. (2000). The monoclonal antibody TER-119 recognizes a molecule associated with glycophorin A and specifically marks the late stages of murine erythroid lineage. *Br. J. Haematol.* **109**, 280–287.

Kissa, K. and Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112–5.

Kissa, K., Murayama, E., Zapata, A., Cortés, A., Perret, E., Machu, C. and Herbomel, P. (2008). Live imaging of emerging hematopoietic stem cells and early thymus colonization. *Blood* **111**, 1147–1156.

Kitabayashi, I., Yokoyama, A., Shimizu, K. and Ohki, M. (1998). Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *EMBO J.* **17**, 2994–3004.

Kitabayashi, I., Aikawa, Y., Nguyen, L. A., Yokoyama, A. and Ohki, M. (2002). Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. *EMBO J.* **20**, 7184–7196.

Kobayashi, I., Kobayashi-Sun, J., Kim, A. D., Pouget, C., Fujita, N., Suda, T. and Traver, D. (2014). Jam1a–Jam2a interactions regulate haematopoietic stem cell fate through Notch signalling. *Nature* **512**, 319–323.

Komori, T. and Kishimoto, T. (1998). Cbfa1 in bone development. *Curr. Opin. Genet. Dev.* **8**, 494–499.

Koressaar, T. and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291.

Kramer, I., Sigrist, M., De Nooij, J. C., Taniuchi, I., Jessell, T. M. and Arber, S. (2006). A role for Runx transcription factor signaling in dorsal root ganglion sensory neuron diversification. *Neuron* **49**, 379–393.

Kraus, P., Winata, C. L. and Lufkin, T. (2014). BAC transgenic zebrafish for transcriptional promoter and enhancer studies. In *Bacterial Artificial Chromosomes: Second Edition*, pp. 245–258.

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.

Kurokawa, M. (2006). AML1/Runx1 as a versatile regulator of hematopoiesis: regulation of its function and a role in adult hematopoiesis. *Int. J. Hematol.* **84**, 136–42.

Kurokawa, M., Yamaguchi, Y., Imai, Y., Izutsu, K., Nitta, E., Mitani, K. and Masanobu Satake, 3 Tetsuo Noda, 4 Yoshiaki Ito, 5 And Hisamaru Hirai1 (2004). The corepressor mSin3A regulates phosphorylation-induced activation, intranuclear location, and stability of AML1. *Mol. Cell. Biol.* **24**, 1033.

Kuvardina, O. N., Herglotz, J., Kolodziej, S., Kohrs, N., Herkt, S., Wojcik, B., Oellerich, T., Corso, J., Behrens, K., Kumar, A., et al. (2015). RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation. *Blood* **125**, 3570–3579.

Lacaud, G., Gore, L., Kennedy, M., Kouskoff, V., Kingsley, P., Hogan, C., Carlsson, L., Speck, N., Palis, J. and Keller, G. (2002). Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro. *Blood* **100**, 458–466.

Lam, E. Y. N., Chau, J. Y. M., Kalev-Zylinska, M. L., Fountaine, T. M., Mead, R. S., Hall, C. J., Crosier, P. S., Crosier, K. E. and Flores, M. V. (2009). Zebrafish runx1 promoter-EGFP transgenics mark discrete sites of definitive blood progenitors. *Blood* **113**, 1241–1249.

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

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, ??zge, Wilson, N. K., M??r??y, T., Bonifer, C., et al. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood* **120**, 314–322.

Le Blanc, K., Frassoni, F., Ball, L., Locatelli, F., Roelofs, H., Lewis, I., Lanino, E., Sundberg, B., Bernardo, M. E., Remberger, M., et al. (2008). Mesenchymal stem cells for treatment of steroid-resistant, severe, acute graft-versus-host disease: a phase II study. *Lancet* **371**, 1579–1586.

Ledbetter, J. A. and Herzenberg, L. A. (1979). Xenogeneic Monoclonal Antibodies to Mouse Lymphoid Differentiation Antigens. *Immunol. Rev.* **47**, 63–90.

Lee, Y., Manegold, J. E., Kim, A. D., Pouget, C., Stachura, D. L., Clements, W. K. and Traver, D. (2014). FGF signalling specifies haematopoietic stem cells through its regulation of somitic Notch signalling. *Nat. Commun.* **5**, 5583.

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

Levanon, D., Bernstein, Y., Negreanu, V., Ghozi, M. C., Bar-Am, I., Aloya, R., Goldenberg, D., Lotem, J. and Groner, Y. (1996). A large variety of alternatively spliced and differentially expressed mRNAs are encoded by the human acute myeloid leukemia gene AML1. *DNA Cell Biol.* **15**, 175–185.

Levanon, D., Goldstein, R. E., Bernstein, Y., Tang, H., Goldenberg, D., Stifani, S., Paroush, Z. and Groner, Y. (1998). Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11590–11595.

Levanon, D., Glusman, G., Bangsow, T., Ben-Asher, E., Male, D. A., Avidan, N., Bangsow, C., Hattori, M., Taylor, T. D., Taudien, S., et al. (2001). Architecture and anatomy of the genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. *Gene* **262**, 23–33.

Levine, M. and Davidson, E. (2005). Gene regulatory networks for development. *Pnas* **102**, 4936–4942.

Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., et al. (2010). DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433.

Li, Z., Lan, Y., He, W., Chen, D., Wang, J., Zhou, F., Wang, Y., Sun, H., Chen, X., Xu, C., et al. (2012). Mouse embryonic head as a site for hematopoietic stem cell development. *Cell Stem Cell* **11**, 663–675.

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

Li, P., Lahvic, J. L., Binder, V., Pugach, E. K., Riley, E. B., Tamplin, O. J., Panigrahy, D., Bowman, T. V., Barrett, F. G., Heffner, G. C., et al. (2015). Epoxyeicosatrienoic acids enhance embryonic haematopoiesis and adult marrow engraftment. *Nature* **523**, 468–471.

Li, Y., Wang, H., Wang, X., Jin, W., Tan, Y., Fang, H., Chen, S., Chen, Z. and Wang, K. (2016). Genome-wide studies identify a novel interplay between AML1 and AML1/ETO in t(8;21) acute myeloid leukemia. *Blood* **127**, 233–242.

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

Liao, E. C., Paw, B. H., Oates, A. C., Pratt, S. J., Postlethwait, J. H. and Zon, L. I. (1998). SCL/Tal-1 transcription factor acts downstream of cloche to specify hematopoietic and vascular progenitors in zebrafish. *Genes Dev.* **12**, 621–626.

Libermann, T. A., Pan, Z., Akbarali, Y., Hetherington, C. J., Boltax, J., Yergeau, D. A. and Zhang, D. E. (1999). AML1 (CBFα2) cooperates with B cell-specific activating protein (BSAP/PAX5) in activation of the B cell-specific BLK gene promoter. *J. Biol. Chem.* **274**, 24671–24676.

Lichtinger, M., Ingram, R., Hannah, R., Müller, D., Clarke, D., Assi, S. A., Lie-A-Ling, M., Noailles, L., Vijayabaskar, M. S., Wu, M., et al. (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J.* **31**, 4318–4333.

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

Lieschke, G. J., Oates, A. C., Paw, B. H., Thompson, M. A., Hall, N. E., Ward, A. C., Ho, R. K., Zon, L. I. and Layton, J. E. (2002). Zebrafish SPI-1 (PU.1) Marks a Site of Myeloid Development Independent of Primitive Erythropoiesis: Implications for Axial Patterning. *Dev. Biol.* **246**, 274–295.

Lin, H. F., Traver, D., Zhu, H., Dooley, K., Paw, B. H., Zon, L. I. and Handin, R. I. (2005). Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. *Blood* **106**, 3803–3810.

Lin, Y., Yoder, M. C. and Yoshimoto, M. (2014). Lymphoid progenitor emergence in the murine embryo and yolk sac precedes stem cell detection. *Stem Cells Dev.* **23**, 1168–77.

Link, K. A., Chou, F. S. and Mulloy, J. C. (2010). Core binding factor at the crossroads: Determining the fate of the HSC. *J. Cell. Physiol.* **222**, 50–56.

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

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H. and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat. Commun.* **4**, 1563.

Liu, X., Jia, X., Yuan, H., Ma, K., Chen, Y., Jin, Y., Deng, M., Pan, W., Chen, S., Chen, Z., et al. (2015). DNA methyltransferase 1 functions through C/ebpa to maintain hematopoietic stem and progenitor cells in zebrafish. *J. Hematol. Oncol.* **8**, 15.

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

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

Lorsbach, R. B., Moore, J., Ang, S. O., Sun, W., Lenny, N. and Downing, J. R. (2004). Role of RUNX1 in adult hematopoiesis: Analysis of RUNX1-IRES-GFP knock-in mice reveals differential lineage expression. *Blood* **103**, 2522–2529.

Lowenberg, B., Downing, J. R. and Burnett, A. (1999). Acute Myeloid Leukemia. *N. Engl. J. Med.* **341**, 1051–1062.

Lu, X., Li, X., He, Q., Gao, J., Gao, Y., Liu, B. and Liu, F. (2013). miR-142-3p regulates the formation and differentiation of hematopoietic stem cells in vertebrates. *Cell Res.* **23**, 1356–1368.

Luc, S., Luis, T. C., Boukarabila, H., Macaulay, I. C., Buza-Vidas, N., Bouriez-Jones, T., Lutteropp, M., Woll, P. S., Loughran, S. J., Mead, A. J., et al. (2012). The earliest thymic T cell progenitors sustain B cell and myeloid lineage potential. *Nat. Immunol.* **13**, 412–419.

Lucas, C. L., Chandra, A., Nejentsev, S., Condliffe, A. M. and Okkenhaug, K. (2016). PI3Kδ and primary immunodeficiencies. *Nat. Rev. Immunol.* **16**, 702–714.

Lutterbach, B., Westendorf, J. J., Linggi, B., Isaac, S., Seto, E. and Hiebert, S. W. (2000). A mechanism of repression by acute myeloid leukemia-1, the target of multiple chromosomal translocations in acute leukemia. *J. Biol. Chem.* **275**, 651–656.

Lyons, S. E., Shue, B. C., Lei, L., Oates, A. C., Zon, L. I. and Liu, P. P. (2001). Molecular cloning, genetic mapping, and expression analysis of four zebrafish c/ebp genes. *Gene* **281**, 43–51.

Lyons, S. E., Lawson, N. D., Lei, L., Bennett, P. E., Weinstein, B. M. and Liu, P. P. (2002). A nonsense mutation in zebrafish gata1 causes the bloodless phenotype in vlad tepes. *Proc Natl Acad Sci U S A* **99**, 5454–9.

MacRae, C. A. and Peterson, R. T. (2015). Zebrafish as tools for drug discovery. *Nat. Rev. Drug Discov.* **14**, 721–731.

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

Marsman, J., O'Neill, A. C., Kao, B. R. Y., Rhodes, J. M., Meier, M., Antony, J., Mönnich, M. and Horsfield, J. A. (2014). Cohesin and CTCF differentially regulate spatiotemporal runx1 expression during zebrafish development. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 50–61.

Martelli, a M., Nyåkern, M., Tabellini, G., Bortul, R., Tazzari, P. L., Evangelisti, C. and Cocco, L. (2006). Phosphoinositide 3-kinase/Akt signaling pathway and its therapeutical implications for human acute myeloid leukemia. *Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, U.K* **20**, 911–928.

Matsuyama, T., Kimura, T., Kitagawa, M., Pfeffer, K., Kawakami, T., Watanabe, N., Kündig, T. M., Amakawa, R., Kishihara, K., Wakeham, A., et al. (1993). Targeted disruption of IRF-1 or IRF-2 results in abnormal type I IFN gene induction and aberrant lymphocyte development. *Cell* **75**, 83–97.

McLarren, K. W., Lo, R., Grbavec, D., Thirunavukkarasu, K., Karsenty, G. and Stifani, S. (2000). The mammalian basic helix loop helix protein HES-1 binds to and modulates the transactivating function of the runt-related factor Cbfa1. *J. Biol. Chem.* **275**, 530–538.

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.

Melnikova, I. N., Crute, B. E., Wang, S. and Speck, N. A. (1993). Sequence specificity of the core-binding factor. *J. Virol.* **67**, 2408–2411.

Michaud, J., Simpson, K. M., Escher, R., Buchet-Poyau, K., Beissbarth, T., Carmichael, C., Ritchie, M. E., Schütz, F., Cannon, P., Liu, M., et al. (2008). Integrative analysis of RUNX1 downstream pathways and target genes. *BMC Genomics* **9**, 363.

Micklem, H. S., Ford, C. E., Evans, E. P., Ogden, D. A. and Papworth, D. S. (1972). Competitive in vivo proliferation of foetal and adult haematopoietic cells in lethally irradiated mice. *J. Cell. Physiol.* **79**, 293–298.

Mikkola, H. K. A., 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.

Mitjavila-Garcia, M. T., Cailleret, M., Godin, I., Nogueira, M. M., Cohen-Solal, K., Schiavon, V., Lecluse, Y., Le Pesteur, F., Lagrue, A. H. and Vainchenker, W. (2002). Expression of CD41 on hematopoietic progenitors derived from embryonic hematopoietic cells. *Development* **129**, 2003–2013.

Monteiro, R., Pouget, C. and Patient, R. (2011). The gata1/pu.1 lineage fate paradigm varies between blood populations and is modulated by tif1γ. *EMBO J.* **30**, 1093–103.

Monteiro, R., Pinheiro, P., Joseph, N., Peterkin, T., Koth, J., Repapi, E., Bonkhofer, F., Kirmizitas, A. and Patient, R. (2016). Transforming Growth Factor β Drives Hemogenic Endothelium Programming and the Transition to Hematopoietic Stem Cells. *Dev. Cell* **38**, 358–370.

Morrison, S. J. and Weissman, I. L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673.

Motoda, L., Osato, M., Yamashita, N., Jacob, B., Chen, L. Q., Yanagida, M., Ida, H., Wee, H. J., Sun, A. X., Taniuchi, I., et al. (2007). Runx1 protects hematopoietic stem/progenitor cells from oncogenic insult. *Stem Cells* **25**, 2976–2986.

Mukouyama, Y., Chiba, N., Hara, T., Okada, H., Ito, Y., Kanamaru, R., Miyajima, a, Satake, M. and Watanabe, T. (2000). The AML1 transcription factor functions to develop and maintain hematogenic precursor cells in the embryonic aorta-gonad-mesonephros region. *Dev. Biol.* **220**, 27–36.

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

Murayama, E., Kissa, K., Zapata, A., Mordelet, E., Briolat, V., Lin, H. F., Handin, R. I. and Herbomel, P. (2006). Tracing Hematopoietic Precursor Migration to Successive Hematopoietic Organs during Zebrafish Development. *Immunity* **25**, 963–975.

Murry, C. E. and Keller, G. (2008). Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. *Cell* **132**, 661–680.

Nakajima-Takagi, Y., Osawa, M., Oshima, M., Takagi, H., Miyagi, S., Endoh, M., Endo, T. A., Takayama, N., Eto, K., Toyoda, T., et al. (2013). Role of SOX17 in hematopoietic development from human embryonic stem cells. *Blood* **121**, 447–458.

Nakano, H., Liu, X., Arshi, A., Nakashima, Y., van Handel, B., Sasidharan, R., Harmon, A. W., Shin, J.-H., Schwartz, R. J., Conway, S. J., et al. (2013). Haemogenic endocardium contributes to transient definitive haematopoiesis. *Nat. Commun.* **4**, 1564.

Namba, K., Abe, M., Saito, S., Satake, M., Ohmoto, T., Watanabe, T. and Sato, Y. (2000). Indispensable role of the transcription factor PEBP2/CBF in angiogenic activity of a murine endothelial cell MSS31. *Oncogene* **19**, 106–14.

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

Nishikawa, S. I., Nishikawa, S., Hirashima, M., Matsuyoshi, N. and Kodama, H. (1998b). Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. *Development* **125**, 1747–1757.

Nishimura, M., Fukushima-nakase, Y., Fujita, Y., Nakao, M., Toda, S., Kitamura, N., Abe, T., Okuda, T. and Dc, W. (2004). VWRPY motif-dependent and -independent roles of AML1/Runx1 transcription factor in murine hematopoietic development. *Blood* **103**, 562–570.

North, T. E. and de Bruijn, M. F. T. R. (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**, 158–68.

North, T., Gu, T. L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marín-Padilla, M. and Speck, N. A. (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development* **126**, 2563–75.

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

North, T. E., Goessling, W., Walkley, C. R., Lengerke, C., Kopani, K. R., Lord, A. M., Weber, G. J., Bowman, T. V., Jang, I.-H., Grosser, T., et al. (2007). Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature* **447**, 1007–1011.

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.

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-San, 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. (1993). Molecular Cloning and Characterization of PEBP2β, the Heterodimeric Partner of a Novel Drosophila runt-Related DNA Binding Protein PEBP2α. *Virology* **194**, 314–331.

Ogawa, M., Kizumoto, M., Nishikawa, S., Fujimoto, T., Kodama, H. and Nishikawa, S. I. (1999). Expression of alpha4-integrin defines the earliest precursor of hematopoietic cell lineage diverged from endothelial cells. *Blood* **93**, 1168–1177.

Ogawa, S., Satake, M. and Ikuta, K. (2008). Physical and functional interactions between STAT5 and Runx transcription factors. *J. Biochem.* **143**, 695–709.

Okada, S., Nakauchi, H., Nagayoshi, K., Nishikawa, S., Miura, Y. and Suda, T. (1992). In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. *Blood* **80**, 3044–3050.

Okada, H., Watanabe, T., Niki, M., Takano, H., Chiba, N., Yanai, N., Tani, K., Hibino, H., Asano, S., Mucenski, M. L., et al. (1998). AML1(−/−) embryos do not express certain hematopoiesis-related gene transcripts including those of the PU.1 gene. *Oncogene* **17**, 2287–2293.

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.

Okuda, T., Takeda, K., Fujita, Y., Nishimura, M., Yagyu, S., Yoshida, M., Akira, S., Downing, J. R. and Abe, T. (2000). Biological characteristics of the leukemia-associated transcriptional factor AML1 disclosed by hematopoietic rescue of AML1-deficient embryonic stem cells by using a knock-in strategy. *Mol. Cell. Biol.* **20**, 319–28.

Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn Diagrams. *BioinfoGP of CNB-CSIC* http://bioinfogp.cnnb.csic.es/tools/venny/index.ht.

Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. *Nat Rev Genet* **1**, 57–64.

Orkin, S. H. and Zon, L. I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* **132**, 631–644.

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

Otto, F., Lübbert, M. and Stock, M. (2003). Upstream and downstream targets of RUNX proteins. *J. Cell. Biochem.* **89**, 9–18.

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.

Palis, J., Chan, R. J., Koniski, A., Patel, R., Starr, M. and Yoder, M. C. (2001). Spatial and temporal emergence of high proliferative potential hematopoietic precursors during murine embryogenesis. *Proc. Natl. Acad. Sci.* **98**, 4528–4533.

Papayannopoulou, T. and Craddock, C. (1997). Homing and trafficking of hemopoietic progenitor cells. *Acta Haematol.* **97**, 97–104.

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.

Patterson, L. J., Gering, M., Eckfeldt, C. E., Green, A. R., Verfaillie, C. M., Ekker, S. C. and Patient, R. (2007). The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. *Blood* **109**,

2389–2398.

Patton, E. E. and Zon, L. I. (2001). The art and design of genetic screens: zebrafish. *Nat. Rev. Genet.* **2**, 956–66.

Penninger, J. M., Sirard, C., Mittrücker, H. W., Chidgey, A., Kozieradzki, I., Nghiem, M., Hakem, A., Kimura, T., Timms, E., Boyd, R., et al. (1997). The interferon regulatory transcription factor IRF-1 controls positive and negative selection of CD8+ thymocytes. *Immunity* **7**, 243–254.

Pereira, C. F., Chang, B., Qiu, J., Niu, X., Papatsenko, D., Hendry, C. E., Clark, N. R., Nomura-Kitabayashi, A., Kovacic, J. C., Ma'Ayan, A., et al. (2013). Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell* **13**, 205–218.

Perié, L. and Duffy, K. R. (2016). Retracing the in vivo haematopoietic tree using single-cell methods. *FEBS Lett.* **590**, 4068–4083.

Petell, C. J., Alabdi, L., He, M., San Miguel, P., Rose, R. and Gowher, H. (2016). An epigenetic switch regulates de novo DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation. *Nucleic Acids Res.* **44**, 7605–7617.

Petersdorf, E. W. (2013). The major histocompatibility complex: A model for understanding graft-versus-host disease. *Blood* **122**, 1863–1872.

Petit-Cocault, L., Volle-Challier, C., Fleury, M., Péault, B. and Souyri, M. (2007). Dual role of Mpl receptor during the establishment of definitive hematopoiesis. *Development* **134**, 3031–3040.

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. (2007a). 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.

Pimanda, J. E., Donaldson, I. J., de Bruijn, M. F. T. R., Kinston, S., Knezevic, K., Huckle, L., Piltz, S., Landry, J.-R., Green, A. R., Tannahill, D., et al. (2007b). The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 840–5.

Pinto Do Ó, P., Kolterud, Å.˚ 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.

Porcher, C., Liao, E. C., Fujiwara, Y., Zon, L. I. and Orkin, S. H. (1999). Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. *Development* **126**, 4603–4615.

Pouget, C., Peterkin, T., Simões, F. C., Lee, Y., Traver, D. and Patient, R. (2014). FGF signalling restricts haematopoietic stem cell specification via modulation of the BMP pathway. *Nat. Commun.* **5**, 5588.

Pozner, A., Goldenberg, D., Negreanu, V., Le, S. Y., Elroy-Stein, O., Levanon, D. and Groner, Y. (2000). Transcription-coupled translation control of AML1/RUNX1 is mediated by cap- and internal ribosome entry site-dependent mechanisms. *Mol. Cell. Biol.* **20**, 2297–2307.

Pozner, A., Lotem, J., Xiao, C., Goldenberg, D., Brenner, O., Negreanu, V., Levanon, D. and Groner, Y. (2007). Developmentally regulated promoter-switch transcriptionally controls Runx1 function during embryonic hematopoiesis. *BMC Dev. Biol.* **7**, 84.

Ptasinska, A., Assi, S. A., Mannari, D., James, S. R., Williamson, D., Dunne, J., Hoogenkamp, M., Wu, M., Care, M., McNeill, H., et al. (2012). Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* **26**, 1829–1841.

Qian, F., Zhen, F., Xu, J., Huang, M., Li, W. and Wen, Z. (2007). Distinct functions for different scl isoforms in zebrafish primitive and definitive hematopoiesis. *PLoS Biol.* **5**, 1110–1119.

Ransom, D. G., Haffter, P., Odenthal, J., Brownlie, a, Vogelsang, E., Kelsh, R. N., Brand, M., van Eeden, F. J., Furutani-Seiki, M., Granato, M., et al. (1996). Characterization of zebrafish mutants with defects in embryonic hematopoiesis. *Development* **123**, 311–9.

Reed-Inderbitzin, E., Moreno-Miralles, I., Vanden-Eynden, S. K., Xie, J., Lutterbach, B., Durst-Goodwin, K. L., Luce, K. S., Irvin, B. J., Cleary, M. L., Brandt, S. J., et al. (2006). RUNX1 associates with histone deacetylases and SUV39H1 to repress transcription. *Oncogene* **25**, 5777–5786.

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. A. (2008). The Emergence of Hematopoietic Stem Cells Is Initiated in the Placental Vasculature in the Absence of Circulation. *Cell Stem Cell* **2**, 252–263.

Richard, C., Drevon, C., Canto, P. Y., Villain, G., Bollérot, K., Lempereur, A., Teillet, M. A., Vincent, C., Rosselló

Castillo, C., Torres, M., et al. (2013). Endothelio-Mesenchymal Interaction Controls runx1 Expression and Modulates the notch Pathway to Initiate Aortic Hematopoiesis. *Dev. Cell* **24**, 600–611.

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

Robertson, K. D. (2005). DNA methylation and human disease. *Nat Rev Genet* **6**, 597–610.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M. and Stainier, D. Y. R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230–233.

Rowley, J. D. (1998). The critical role of chromosome translocations in human leukemias. *Annu. Rev. Genet.* **32**, 495–519.

Russler-Germain, D. A., Spencer, D. H., Young, M. A., Lamprecht, T. L., Miller, C. A., Fulton, R., Meyer, M. R., Erdmann-Gilmore, P., Townsend, R. R., Wilson, R. K., et al. (2014). The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. *Cancer Cell* **25**, 442–454.

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S. and Medvinsky, A. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J. Exp. Med.* **208**, 1305–1315.

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O. and Medvinsky, A. (2014). Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43- embryonic precursor. *Stem Cell Reports* **3**, 489–501.

Sakai, E., Kitajima, K., Sato, A. and Nakano, T. (2009). Increase of hematopoietic progenitor and suppression of endothelial gene expression by Runx1 expression during in vitro ES differentiation. *Exp. Hematol.* **37**, 334–345.

Saleque, S., Kim, J., Rooke, H. M. and Orkin, S. H. (2007). Epigenetic Regulation of Hematopoietic Differentiation by Gfi-1 and Gfi-1b Is Mediated by the Cofactors CoREST and LSD1. *Mol. Cell* **27**, 562–572.

Sandler, V. M., Lis, R., Liu, Y., Kedem, A., James, D., Elemento, O., Butler, J. M., Scandura, J. M. and Rafii, S. (2014). Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. *Nature* **511**, 312–318.

Sasaki, K., Yagi, H., Bronson, R. T., Tominaga, K., Matsunashi, T., Deguchi, K., Tani, Y., Kishimoto, T. and Komori, T. (1996). Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor beta. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12359–63.

Sato, T., Goyama, S., Nitta, E., Takeshita, M., Yoshimi, M., Nakagawa, M., Kawazu, M., Ichikawa, M. and Kurokawa, M. (2008). Evi-1 promotes para-aortic splanchnopleural hematopoiesis through up-regulation of GATA-2 and repression of TGF-β signaling. *Cancer Sci.* **99**, 1407–1413.

Saunthararajah, Y., Boccuni, P. and Nucifora, G. (2006). Combinatorial action of RUNX1 and PU.1 in the regulation of hematopoiesis. *Crit. Rev. Eukaryot. Gene Expr.* **16**, 183–92.

Sawamiphak, S., Kontarakis, Z. and Stainier, D. Y. R. (2014). Interferon Gamma Signaling Positively Regulates Hematopoietic Stem Cell Emergence. *Dev. Cell* **31**, 640–653.

Sebzda, E., Hibbard, C., Sweeney, S., Abtahian, F., Bezman, N., Clemens, G., Maltzman, J. S., Cheng, L., Liu, F., Turner, M., et al. (2006). Syk and Slp-76 Mutant Mice Reveal a Cell-Autonomous Hematopoietic Cell Contribution to Vascular Development. *Dev. Cell* **11**, 349–361.

Seda, V. and Mraz, M. (2015). B-cell receptor signalling and its crosstalk with other pathways in normal and malignant cells. *Eur. J. Haematol.* **94**, 193–205.

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

Siekmann, A. F. and Lawson, N. D. (2007). Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries. *Nature* **445**, 781–784.

Sievers, F. and Higgins, D. G. (2014). Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079**, 105–116.

Simon, J. A. and Kingston, R. E. (2009). Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.*

Solar, G. P., Kerr, W. G., Zeigler, F. C., Hess, D., Donahue, C., de Sauvage, F. J. and Eaton, D. L. (1998). Role of c-mpl in early hematopoiesis. *Blood* **92**, 4–10.

Song, W. J., Sullivan, M. G., Legare, R. D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I. C., Haworth, C., Hock, R., et al. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23**, 166–175.

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

Soond, D. R., Bjørgo, E., Moltu, K., Dale, V. Q., Patton, D. T., Torgersen, K. M., Galleway, F., Twomey, B., Clark, J., Gaston, J. S. H., et al. (2010). PI3K p110delta regulates T-cell cytokine production during primary and secondary immune responses in mice and humans. *Blood* **115**, 2203–13.

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

Spangrude, G. J., Heimfeld, S. and Weissman, I. L. (1988a). Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62.

Spangrude, G. J., Aihara, Y., Weissman, I. L. and Klein, J. (1988b). The stem cell antigens Sca-1 and Sca-2 subdivide thymic and peripheral T lymphocytes into unique subsets. *J. Immunol.* **141**, 3697–707.

Speck, N. A. (1995). A new transcription factor family associated with human leukemias. *Crit. Rev. Eukaryot. Gene Expr.* **5**, 337–364.

Speck, N. A., Stacy, T., Wang, Q., North, T., Gu, T. L., Miller, J., Binder, M. and Mar??n-Padilla, M. (1999). Core-binding factor: A central player in hematopoiesis and leukemia. In *Cancer Research*, .

Sroczynska, P., Lancrin, C., Kouskoff, V. and Lacaud, G. (2009). The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood* **114**, 5279–5289.

Staal, F. J. T., Baum, C., Cowan, C., Dzierzak, E., Hacein-Bey-Abina, S., Karlsson, S., Lapidot, T., Lemischka, I., Mendez-Ferrer, S., Mikkers, H., et al. (2011). Stem cell self-renewal: lessons from bone marrow, gut and iPS toward clinical applications. *Leukemia* **25**, 1095–102.

Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F. and Soler, E. (2012). Transcription regulation by distal enhancers: who's in the loop? *Transcription* **3**, 181–186.

Stadtfeld, M. and Hochedlinger, K. (2010). Induced pluripotency: History, mechanisms, and applications. *Genes Dev.* **24**, 2239–2263.

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. and Mateo, J. L. (2015). CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One* **10**,.

Strom, D. K., John, N., Westendorf, J. J., Linggi, B., Lutterbach, B., Downing, J. R., Lenny, N. and Hiebert, S. W. (2000). Expression of the AML-1 oncogene shortens the G1 phase of the cell cycle. *J. Biol. Chem.* **275**, 3438–3445.

Sturgeon, C. M., Ditadi, A., Awong, G., Kennedy, M. and Keller, G. (2014). Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 554–561.

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

Sujobert, P., Bardet, V., Cornillet-Lefebvre, P., Hayflick, J. S., Prie, N., Verdier, F., Vanhaesebroeck, B., Muller, O., Pesce, F., Ifrah, N., et al. (2005). Essential role for the p110delta isoform in phosphoinositide 3-kinase activation and cell proliferation in acute myeloid leukemia. *Blood* **106**, 1063–1066.

Summerton, J. and Weller, D. (1997). Morpholino Antisense Oligomers: Design, Preparation, and Properties. *Antisense Nucleic Acid Drug Dev.* **7**, 187–195.

Suster, M. L., Kikuta, H., Urasaki, A., Asakawa, K. and Kawakami, K. (2009). Transgenesis in Zebrafish with the Tol2 transposon system. *Methods Mol. Biol.* **561**, 41–63.

Suster, M. L., Abe, G., Schouw, A. and Kawakami, K. (2011). Transposon-mediated BAC transgenesis in zebrafish. *Nat. Protoc.* **6**, 1998–2021.

Swiers, G., Rode, C., Azzoni, E. and De Bruijn, M. F. T. R. (2013a). A short history of hemogenic endothelium. *Blood Cells, Mol. Dis.* **51**, 206–212.

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K. D., et al. (2013b). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat. Commun.* **4**,.

Swinburne, I. A. and Silver, P. A. (2008). Intron delays and transcriptional timing during development. *Dev. Cell* **14**, 324–330.

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

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 CBFb. *Cell* **105**, 291.

Takahashi, K. and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676.

Takakura, N., Watanabe, T., Suenobu, S., Yamada, Y., Noda, T., Ito, Y., Satake, M. and Suda, T. (2000). A Role for Hematopoietic Stem Cells in Promoting Angiogenesis. *Cell* **102**, 199–209.

Takeuchi, M., Sekiguchi, T., Hara, T., Kinoshita, T. and Miyajima, A. (2002). Cultivation of aorta-gonad-mesonephros-derived hematopoietic stem cells in the fetal liver microenvironment amplifies long-term repopulating activity and enhances engraftment to the bone marrow. *Blood* **99**, 1190–1196.

Takeuchi, M., Fuse, Y., Watanabe, M., Andrea, C.-S., Takeuchi, M., Nakajima, H., Ohashi, K., Kaneko, H., Kobayashi-Osaki, M., Yamamoto, M., et al. (2015). LSD1/KDM1A promotes hematopoietic commitment of hemangioblasts through downregulation of Etv2. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13922–7.

Tamplin, O. J., Durand, E. M., Carr, L. A., Childs, S. J., Hagedorn, E. J., Li, P., Yzaguirre, A. D., Speck, N. A. and Zon, L. I. (2015). Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche. *Cell* **160**, 241–252.

Tamura, T., Yanai, H., Savitsky, D. and Taniguchi, T. (2008). The IRF Family Transcription Factors in Immunity and Oncogenesis. *Annu. Rev. Immunol.* **26**, 535–584.

Tanaka, T., Kurokawa, M., Ueki, K., Tanaka, K., Imai, Y., Mitani, K., Okazaki, K., Sagata, N., Yazaki, Y., Shibata, Y., et al. (1996). The extracellular signal-regulated kinase pathway phosphorylates AML1, an acute myeloid leukemia gene product, and potentially regulates its transactivation ability, Mol. *Cell. Biol.* **16**, 3967–3979.

Tanaka, Y., Joshi, A., Wilson, N. K., Kinston, S., Nishikawa, S. and Göttgens, B. (2012). The transcriptional programme controlled by Runx1 during early embryonic blood development. *Dev. Biol.* **366**, 404–419.

Tang, Y., Bai, H., Urs, S., Wang, Z. and Liaw, L. (2013). Notch1 activation in embryonic VE-cadherin populations selectively blocks hematopoietic stem cell generation and fetal liver hematopoiesis. *Transgenic Res.* **22**, 403–410.

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. (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.

Taoudi, S. and Medvinsky, A. (2007). Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9399–403.

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.

Telfer, J. C. and Rothenberg, E. V (2001). Expression and function of a stem cell promoter for the murine CBFalpha2 gene: distinct roles and regulation in natural killer and T cell development. *Dev. Biol.* **229**, 363–82.

Testa, U., Stellacci, E., Pelosi, E., Sestili, P., Venditti, M., Orsatti, R., Fragale, a, Petrucci, E., Pasquini, L., Belardelli, F., et al. (2004). Impaired myelopoiesis in mice devoid of interferon regulatory factor 1. *Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, U.K* **18**, 1864–1871.

Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., Lancrin, C., Clapes, T., Möröy, T., et al. (2015). GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat. Cell Biol.* **18**, 21–32.

Thambyrajah, R., Ucanok, D., Jalali, M., Hough, Y., Wilkinson, R. N., McMahon, K., Moore, C. and Gering, M. (2016). A gene trap transposon eliminates haematopoietic expression of zebrafish Gfi1aa, but does not interfere with haematopoiesis. *Dev. Biol.* **417**, 25–39.

Theriault, F. M. (2005). Role for Runx1 in the Proliferation and Neuronal Differentiation of Selected Progenitor Cells in the Mammalian Nervous System. *J. Neurosci.* **25**, 2050–2061.

Thol, F., Damm, F., Lüdeking, A., Winschel, C., Wagner, K., Morgan, M., Yun, H., Göhring, G., Schlegelberger, B., Hoelzer, D., et al. (2011). Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *J. Clin. Oncol.* **29**, 2889–2896.

Thompson, M. a, Ransom, D. G., Pratt, S. J., MacLennan, H., Kieran, M. W., Detrich 3rd, H. W., Vail, B., Huber, T. L., Paw, B., Brownlie, A. J., et al. (1998). The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis. *Dev. Biol.* **197**, 248–269.

Till, J. E. and McCulloch, E. A. (1961). A Direct Measurement of the Radiation Sensitivity of Normal Mouse Bone Marrow Cells. *Radiat. Res.* **14**, 213.

Tober, J., Koniski, A., McGrath, K. E., Vemishetti, R., Emerson, R., De Mesy-Bentley, K. K. L., 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.

Tober, J., Yzaguirre, A. D., Piwarzyk, E. and Speck, N. a (2013). Distinct temporal requirements for Runx1 in hematopoietic progenitors and stem cells. *Development* **140**, 3765–76.

Traver, D., Paw, B. H., Poss, K. D., Penberthy, W. T., Lin, S. and Zon, L. I. (2003). Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat. Immunol.* **4**, 1238–1246.

Trombly, D. J., Whitfield, T. W., Padmanabhan, S., Gordon, J. A. R., Lian, J. B., van Wijnen, A. J., Zaidi, S. K., Stein, J. L. and Stein, G. S. (2015). Genome-wide co-occupancy of AML1-ETO and N-CoR defines the t(8;21) AML signature in leukemic cells. *BMC Genomics* **16**, 309.

Truett, G. E., Heeger, P., Mynatt, R. L., Truett, A. A., Walker, J. A. and Warman, M. L. (2000). Preparation of PCR-quality mouse genomic dna with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* **29**, 52–54.

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.

Turner, M., Schweighoffer, E., Colucci, F., Di Santo, J. P. and Tybulewicz, V. L. (2000). Tyrosine kinase SYK: Essential functions for immunoreceptor signalling. *Immunol. Today* **21**, 148–154.

Van der Flier, A. and Sonnenberg, A. (2001). Function and interactions of integrins. *Cell Tissue Res.* **305**, 285–298.

Van Hoof, D., Munoz, J., Braam, S. R., Pinkse, M. W., Linding, R., Heck, A. J., Mummery, C. L. and Krijgsveld, J. (2009). Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* **5**, 214–226.

van Keuren, M. L., Gavrilina, G. B., Filipiak, W. E., Zeidler, M. G. and Saunders, T. L. (2009). Generating transgenic mice from bacterial artificial chromosomes: Transgenesis efficiency, integration and expression outcomes. *Transgenic Res.* **18**, 769–785.

Vargel, Ö., Zhang, Y., Kosim, K., Ganter, K., Foehr, S., Mardenborough, Y., Shvartsman, M., Enright, A. J., Krijgsveld, J. and Lancrin, C. (2016). Activation of the TGFβ pathway impairs endothelial to haematopoietic transition. *Sci. Rep.* **6**, 21518.

Varshney, G. K., Pei, W., Lafave, M. C., Idol, J., Xu, L., Gallardo, V., Carrington, B., Bishop, K., Jones, M., Li, M., et al. (2015). High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.* **25**, 1030–1042.

Voigt, P. and Reinberg, D. (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia The Cancer Genome Atlas Research Network. *N. Engl. J. Med.* **368**, 2059–74.

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**, 3324–39.

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, a H. and Speck, N. a (1996a). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 3444–9.

Wang, Q., Stacy, T., Miller, J. D., Lewis, A. F., Gu, T. L., Huang, X., Bushweller, J. H., Bories, J. C., Alt, F. W., Ryan, G., et al. (1996b). The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. *Cell* **87**, 697–708.

Wang, X., Blagden, C., Fan, J., Nowak, S. J., Taniuchi, I., Littman, D. R. and Burden, S. J. (2005). Runx1 prevents wasting, myofibrillar disorganization, and autophagy of skeletal muscle. *Genes Dev.* **19**, 1715–1722.

Wang, L., Huang, G., Zhao, X., Hatlen, M. A., Vu, L., Liu, F. and Nimer, S. D. (2009a). Post-translational modifications

of Runx1 regulate its activity in the cell. *Blood Cells, Mol. Dis.* **43**, 30–34.

Wang, Z., Li, G., Tse, W. and Bunting, K. D. (2009b). Conditional deletion of STAT5 in adult mouse hematopoietic stem cells causes loss of quiescence and permits efficient nonablative stem cell replacement. *Blood* **113**, 4856–4865.

Wang, L., Zhang, P., Wei, Y., Gao, Y., Patient, R. and Liu, F. (2011). Ablood flow-dependent klf2a-NO signaling cascade is required for stabilization of hematopoietic stem cell programming in zebrafish embryos. *Blood* **118**, 4102–4110.

Wang, C. Q., Krishnan, V., Tay, L. S., Chin, D. W. L., Koh, C. P., Chooi, J. Y., Nah, G. S. S., Du, L., Jacob, B., Yamashita, N., et al. (2014). Disruption of Runx1 and Runx3 leads to bone marrow failure and leukemia predisposition due to transcriptional and DNA repair defects. *Cell Rep.* **8**, 767–782.

Wang, S., He, Q., Ma, D., Xue, Y. and Liu, F. (2015). Irf4 Regulates the Choice between T Lymphoid-Primed Progenitor and Myeloid Lineage Fates during Embryogenesis. *Dev. Cell* **34**, 621–631.

Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. and Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, 1–12.

Watanabe, D., Suetake, I., Tajima, S. and Hanaoka, K. (2004). Expression of Dnmt3b in mouse hematopoietic progenitor cells and spermatogonia at specific stages. *Gene Expr. Patterns* **5**, 43–49.

Webber, B. R., Iacovino, M., Choi, S. H., Tolar, J., Kyba, M. and Blazar, B. R. (2013). DNA methylation of Runx1 regulatory regions correlates with transition from primitive to definitive hematopoietic potential in vitro and in vivo. *Blood* **122**, 2978–2986.

Wei, W., Wen, L., Huang, P., Zhang, Z., Chen, Y., Xiao, A., Huang, H., Zhu, Z., Zhang, B. and Lin, S. (2008). Gfi1.1 regulates hematopoietic lineage differentiation during zebrafish embryogenesis. *Cell Res.* **18**, 677–85.

Wheeler, J. C., Shigesada, K., Gergen, J. P. and Ito, Y. (2000). Mechanisms of transcriptional regulation by Runt domain proteins. *Semin. Cell Dev. Biol.* **11**, 369–375.

Wilkerson, M. D. and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573.

Wilkinson, R. N., Pouget, C., Gering, M., Russell, A. J., Davies, S. G., Kimelman, D. and Patient, R. (2009). Hedgehog and Bmp Polarize Hematopoietic Stem Cell Emergence in the Zebrafish Dorsal Aorta. *Dev. Cell* **16**, 909–916.

Wilson, A., Murphy, M. J., Oskarsson, T., Kaloulis, K., Bettess, M. D., Oser, G. M., Pasche, A.-C., Knabenhans, C., Macdonald, H. R. and Trumpp, A. (2004). c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–63.

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 program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. *Blood* **113**, 5456–65.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., Ouwehand, W. H., Dzierzak, E., et al. (2010a). 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., Timms, R. T., Kinston, S. J., Cheng, Y.-H., Oram, S. H., Landry, J.-R., Mullender, J., Ottersbach, K. and Gottgens, B. (2010b). Gfi1 Expression Is Controlled by Five Distinct Regulatory Regions Spread over 100 Kilobases, with Scl/Tal1, Gata2, PU.1, Erg, Meis1, and Runx1 Acting as Upstream Regulators in Early Hematopoietic Cells. *Mol. Cell. Biol.* **30**, 3853–3863.

Witzenbichler, B., Maisonpierre, P. C., Jones, P., Yancopoulos, G. D. and Isner, J. M. (1998). Chemotactic properties of angiopoietin-1 and -2, ligands for the endothelial-specific receptor tyrosine kinase Tie2. *J. Biol. Chem.* **273**, 18514–18521.

Wong, P. M., Chung, S. W., Reicheld, S. M. and Chui, D. H. (1986). Hemoglobin switching during murine embryonic development: evidence for two populations of embryonic erythropoietic progenitor cells. *Blood* **67**, 716–21.

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**, 7731–6.

Wotton, D., Ghysdael, J., Wang, S., Speck, N. A. and Owen, M. J. (1994). Cooperative binding of Ets-1 and core binding factor to DNA. *Mol. Cell. Biol.* **14**, 840–850.

Wotton, S., Terry, A., Kilbey, A., Jenkins, A., Herzyk, P., Cameron, E. and Neil, J. C. (2008). Gene array analysis reveals a common Runx transcriptional programme controlling cell adhesion and survival. *Oncogene* **27**, 5856–5866.

Xu, J., Zhu, L., He, S., Wu, Y., Jin, W., Yu, T., Qu, J. Y. and Wen, Z. (2015). Temporal-Spatial Resolution Fate Mapping Reveals Distinct Origins for Embryonic and Adult Microglia in Zebrafish. *Dev. Cell* **34**, 632–641.

Yagi, R., Chen, L. F., Shigesada, K., Murakami, Y. and Ito, Y. (1999). A WW domain-containing Yes-associated protein (YAP) is a novel transcriptional co-activator. *EMBO J.* **18**, 2551–2562.

Yamaguchi, Y., Kurokawa, M., Imai, Y., Izutsu, K., Asai, T., Ichikawa, M., Yamamoto, G., Nitta, E., Yamagata, T., Sasaki, K., et al. (2004). AML1 Is Functionally Regulated through p300-mediated Acetylation on Specific Lysine Residues. *J. Biol. Chem.* **279**, 15630–15638.

Yan, X.-J., Xu, J., Gu, Z.-H., Pan, C.-M., Lu, G., Shen, Y., Shi, J.-Y., Zhu, Y.-M., Tang, L., Zhang, X.-W., et al. (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat. Genet.* **43**, 309–315.

Yoder, M. C. (2014). Inducing definitive hematopoiesis in a dish. *Nat. Biotechnol.* **32**, 539–541.

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

Yoder, M. C., Hiatt, K. and Mukherjee, P. (1997b). In vivo repopulating hematopoietic stem cells are present in the murine yolk sac at day 9.0 postcoitus. *Proc. Natl. Acad. Sci.* **94**, 6776–6780.

Yokomizo, T. and Dzierzak, E. (2010). Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. *Development* **137**, 3651–3661.

Yokomizo, T., Ogawa, M., Osato, M., Kanno, T., Yoshida, H., Fujimoto, T., Fraser, S., Nishikawa, S., Okada, H., Satake, M., et al. (2001). Requirement of Runx1/AML1/PEBP2alphaB for the generation of haematopoietic cells from endothelial cells. *Genes Cells* **6**, 13–23.

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.

Yokota, T., Huang, J., Tavian, M., Nagai, Y., Hirose, J., Zúñiga-Pflücker, J.-C., Péault, B. and Kincade, P. W. (2006). Tracing the first waves of lymphopoiesis in mice. *Development* **133**, 2041–51.

Yoshihara, H., Arai, F., Hosokawa, K., Hagiwara, T., Takubo, K., Nakamura, Y., Gomei, Y., Iwasaki, H., Matsuoka, S., Miyamoto, K., et al. (2007). Thrombopoietin/MPL Signaling Regulates Hematopoietic Stem Cell Quiescence and Interaction with the Osteoblastic Niche. *Cell Stem Cell* **1**, 685–697.

Yoshimoto, M., Montecino-Rodriguez, E., Ferkowicz, M. J., Porayette, P., Shelley, W. C., Conway, S. J., Dorshkind, K. and Yoder, M. C. (2011). Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1468–1473.

Yoshimoto, M., Porayette, P., Glosson, N. L., Conway, S. J., Carlesso, N., Cardoso, A. A., Kaplan, M. H. and Yoder, M. C. (2012). Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence. *Blood* **119**, 5706–5714.

Yu, M., Mazor, T., Huang, H., Huang, H.-T., Kathrein, K., Woo, A., Chouinard, C., Labadorf, A., Akie, T., Moran, T., et al. (2012). Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Mol Cell* **45**, 330–343.

Zhang, P. and Liu, F. (2011). In vivo imaging of hematopoietic stem cell development in the zebrafish. *Front. Med.* **5**, 239–247.

Zhang, X. Y. and Rodaway, A. R. F. (2007). SCL-GFP transgenic zebrafish: In vivo imaging of blood and endothelial development and identification of the initial site of definitive hematopoiesis. *Dev. Biol.* **307**, 179–194.

Zhang, Y., Biggs, J. R. and Kraft, A. S. (2004). Phorbol ester treatment of K562 cells regulates the transcriptional activity of AML1c through phosphorylation. *J. Biol. Chem.* **279**, 53116–25.

Zhang, Y., Jin, H., Li, L., Qin, F. X. F. and Wen, Z. (2011). cMyb regulates hematopoietic stem/progenitor cell mobilization during zebrafish hematopoiesis. *Blood* **118**, 4093–4101.

Zhang, C., Patient, R. and Liu, F. (2013). Hematopoietic stem cell development and regulatory signaling in zebrafish. *Biochim. Biophys. Acta - Gen. Subj.* **1830**, 2370–2374.

Zhang, P., He, Q., Chen, D., Liu, W., Wang, L., Zhang, C., Ma, D., Li, W., Liu, B. and Liu, F. (2015). G protein-coupled receptor 183 facilitates endothelial-to-hematopoietic transition via Notch1 inhibition. *Cell Res.* **25**, 1093–1107.

Zhao, X., Jankovic, V., Gural, A., Huang, G., Pardanani, A., Menendez, S., Zhang, J., Dunne, R., Xiao, A., Erdjument-Bromage, H., et al. (2008). Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity. *Genes Dev.* **22**, 640–653.

Zhen, F., Lan, Y., Yan, B., Zhang, W. and Wen, Z. (2013). Hemogenic endothelium specification and hematopoietic stem cell maintenance employ distinct Scl isoforms. *Development* **140**, 3977–85.

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

Zhou, X. and Vize, P. D. (2004). Proximo-distal specialization of epithelial transport processes within the Xenopus pronephric kidney tubules. *Dev. Biol.* **271**, 322–338.

Zhou, B. O., Ding, L. and Morrison, S. J. (2015). Hematopoietic stem and progenitor cells regulate the regeneration of their niche by secreting Angiopoietin-1. *Elife* **2015**,.

Zhu, X., Yeadon, J. E. and Burden, S. J. (1994). AML1 is expressed in skeletal muscle and is regulated by innervation. *Mol. Cell. Biol.* **14**, 8051–8057.

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.