



University of
Nottingham

UK | CHINA | MALAYSIA

**A Conserved Non-Coding Element Is
Sufficient But Not Essential For *c-myb*
Expression During Zebrafish
Haematopoiesis**

Jawaher Almulhim

**Thesis submitted to the University of Nottingham for the
degree of Doctor of Philosophy**

July 2021

Acknowledgements

First and foremost, I would like to thank "**Allah**" Almighty for giving me the strength, knowledge, ability and opportunity to undertake this research study and to persevere and complete it satisfactorily. Without blessings of Allah, this achievement would not have been possible.

I am incredibly grateful to my supervisor **Dr. Martin Gering** for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. Thank you, Sir, I could not have imagined having a better supervisor and mentor. Without your guidance, this project would not have been possible.

I would like to deeply thank my **parents**, who have always believed in my abilities, it is your love and prayers **mum** and **dad** (May Allah rest his soul in peace) that have brought me this far. A special and profound thanks to my **sisters** and **brothers** for their encouragement and prayers. Thank you all.

To my kids, **Abdulaziz, Fajer, Abdulrahman** and **Meznah** for all the happiness and smiles that they gave to me, to get me out of the work pressure and stress. My kids, I am sorry for every single day and the moment I spent it away from you. This journey would

not have been possible without you, and I dedicate this thesis to you all. I love you.

At most, I would like to thank my husband **Abdullah** for supporting and encouraging me for everything throughout this experience.

I would like to express my gratitude to my best friend, **Fatemah**. A massive thank you for all the fun we have had in the last four years. Brilliant friend Fatemah, you have made my PhD experience a cherishable one.

Exceptional gratitude goes out to all down at **Saudi Arabia** government for helping and providing the funding for this work.

My heartfelt gratitude to **Dr Joanna Richens** and **Yasmin Hough** that have always encouraged me and who have been there for me in good times and bad also it was fantastic to have the opportunity to work with you.

I would also like to thank all of the lab members for making my stay more enjoyable and for sharing their knowledge - **Peter, Mariam, Farah, Aisha, Amal** and **Jawaher**.

List of Abbreviations

AGM	Aorta-Gonad-Mesonephros
ALM	Anterior Lateral Mesoderm
BA	Branchial Arches
BM	Bone Marrow
Bp	Base Pair
Cas9	CRISPR Associated Protein 9
ChIP-seq	Chromatin Immunoprecipitation (ChIP) Followed by Sequencing
CHT	Caudal Haematopoietic Tissue
CNE	Conserved Non-Coding Element
CpGIs	CpG (Cytosine Followed by Guanine and Linked Together By Phosphate) Islands
CREs	<i>Cis</i> -Regulatory Elements
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTCF	CCCTC-Binding Factor
dHSCs	Definitive Haematopoietic Stem Cells
dHSPCs	Definitive Haematopoietic Stem and Progenitor Cells
DNA	Deoxyribonucleic Acid
DNaseI HS	DNaseI Hypersensitive
DSB	Double-Strand Break
ECs	Endothelial Cells
EGFP	Enhanced Green Fluorescent Protein
EHT	Endothelial to Haematopoietic Transition

EMPs	Erythromyeloid Progenitors
ePreN	Early Neutrophil Progenitors
F0	Founding Generation
F1	First Generation Offspring
F2	Second Generation Offspring
FACS	Fluorescence-Activated Cell Sorting
FL	Foetal Liver
FSC	Forward Scatter
gRNA	Guide RNA
H3K4me1	Mono-Methylation Of Lysine 4 On Histone 3
H3K4me3	Tri-Methylation Of Lysine 4 On Histone 3
hCNE1	Human Conserved Non-Coding Element One
HECs	Haemogenic Endothelium Cells
hpf	Hours Post-Fertilisation
hRE1	Human Regulatory Element One
HSCs	Haematopoietic Stem Cells
HSPCs	Haematopoietic Stem and Progenitor Cells
ICM	Intermediate Cell Mass
Kb	kilobase
KM	Kidney Marrow
LT-HSC	Long-Term Haematopoietic Stem Cell
MHB	Mid/Hindbrain Boundary
mRE1	Mouse Regulatory Element One
mRNA	Messenger RNA
Neuts	Neutrophils
NHEJ	Non-Homologous End Joining

OP	Olfactory Placode
P/Ms	Promyelocytes and Myelocytes
PAM	Protospacer Adjacent Motif
PBI	Posterior Blood Island
PCR	Polymerase Chain Reaction
PCV	Posterior Cardinal Vein
PEs	Potential Enhancer Elements
pf	Days Post Fertilisation
PF5	Promoter Fragment Five
PND	Pronephric Duct
prRBCs	Primitive Red Blood Cells
RBI	Rostral Blood Island
Re	Retina
RNA	Ribonucleic Acid
RT	Room Temperature
scRNA-Seq	Single-Cell RNA Sequencing
SNPs	Single-Nucleotide Polymorphisms
SNs	Segmented Neutrophils
SSC	Side Scatter
TADs	Topologically Associating Domains
TAE	Tris Acetate EDTA Buffer
TE	Tris EDTA Buffer
TF	Transcription Factor
TS	Target Site
t-SNE	t-Stochastic Neighbour Embedding
UMI	Unique Molecular Identifier

vDA	Ventral Wall of the Dorsal Aorta
WISH	Whole-mount In Situ Hybridisation
WKM	Whole Kidney Marrow
WT	Wild-Type
YS	Yolk Sac
zRE1	Zebrafish Regulatory Element One

Table of Contents

Acknowledgements	I
List of Abbreviations	III
Table of Contents	VII
List of Figures	XII
List of Tables	XV
List of Appendix Figures	XVI
List of Appendix Tables	XVII
Abstract	XVIII

1 Introduction	1
1.1 Haematopoiesis	5
1.1.1 Haematopoiesis in mouse	5
1.1.1.1 Primitive haematopoiesis in mouse.....	7
1.1.1.2 The intermediate wave of mouse haematopoiesis gives rise to erythromyeloid progenitors in the yolk sac.....	9
1.1.1.3 Definitive haematopoiesis in mouse.....	9
1.1.2 The zebrafish as a model for haematopoiesis	12
1.1.3 Haematopoiesis in zebrafish	13
1.1.3.1 Primitive haematopoiesis in zebrafish	13
1.1.3.2 Erythromyeloid progenitors arise in the posterior blood island (PBI) during the intermediate wave.....	17
1.1.3.3 Definitive haematopoiesis in zebrafish	20
1.2 Non-coding <i>cis</i>-regulatory elements	25
1.2.1 Promoters	25
1.2.2 Enhancers	27
1.2.3 Insulator	30
1.3 Conserved non-coding elements (CNEs)	33
1.4 c-Myb and its importance during haematopoiesis	34

1.5	Identification and testing of <i>c-myb</i> promoter-proximal regulatory sequences	43
1.6	Testing of potential <i>c-myb</i> distal regulatory elements in zebrafish transgenic lines identified CNE1 as a conserved haematopoietic enhancer	46
1.7	Identification of H3K4me1 and H3K4me3 sites in the zebrafish genome	55
1.8	CRISPR/Cas9 system as a tool for gene editing	58
1.9	Background	61
1.10	Aims and objectives	63
2	Materials and Methods	65
2.1	Materials	65
2.1.1	Buffers and solutions.....	65
2.2	Methods.....	68
2.2.1	Single-cell gene expression profile.....	68
2.2.1.1	The 10x Genomics Chromium system	68
2.2.1.1.1	Preparation, constructing and sequencing of single-cell gene expression of zebrafish kidney marrow.....	68
2.2.1.2	Cell Ranger analysis	70
2.2.1.3	Re-analysis of Dinh and colleague' and Grassi colleagues' published data.....	71
2.2.2	Targeting interest regions of downstream of <i>c-myb</i> gene using CRISPR/Cas9 system	72
2.2.2.1	Single guide RNA (sgRNA) design.....	72
2.2.2.2	The generation of the complementary sgRNA	76
2.2.3	Establishment of stable human regulatory element one (hRE1) zebrafish transgenic line	81
2.2.3.1	Generation of a GFP reporter construct containing human C-MYB CNE1	81

2.2.4	Molecular biology methods.....	88
2.2.4.1	Fin clipping of adult zebrafish	88
2.2.4.2	Genomic DNA isolation from embryos and fin clip biopsies ..	88
2.2.4.3	gDNA extraction from embryos after whole-mount RNA in situ hybridisation	89
2.2.4.4	Primer design	90
2.2.4.5	Polymerase chain reaction (PCR).....	92
2.2.4.6	Agarose gel electrophoresis	92
2.2.4.7	DNA extraction from agarose gel.....	93
2.2.4.8	Ligation	93
2.2.4.9	Transformation and Mini/Midi-preparation of plasmid DNA...94	
2.2.4.10	Restriction enzyme digest	95
2.2.4.11	Measuring DNA and RNA concentration	96
2.2.4.12	DNA Sequencing	96
2.2.4.13	Cloning of the target site linker in pDR274	96
2.2.4.14	sgRNAs synthesis for Microinjection	98
2.2.5	Zebrafish methods.....	100
2.2.5.1	Maintenance of zebrafish and husbandry.....	100
2.2.5.2	Mutant and transgenic zebrafish lines used in this study ...	101
2.2.5.3	Whole-mount <i>in situ</i> hybridization (WISH).....	101
2.2.5.4	Adult Zebrafish Kidney FACS Analysis.....	104
2.2.5.4.1	Whole kidney marrow (WKM) cells isolation.....	104
2.2.5.4.2	Fluorescence activated cell sorting (FACS).....	105
2.2.5.5	Microinjection needle preparation.....	108
2.2.5.6	Microinjection	108
2.2.6	Imaging.....	109
3	Results.....	110
3.1	Single-cell RNA Sequencing (scRNA-Seq) data suggests that all GFP+ cells isolated from whole kidney marrow (WKM) of adult Tg(<i>c-myb-zRE1:PF5:egfp</i>)^{qmc85} zebrafish are neutrophilic granulocytes.....	110
3.1.1	Isolation of single GFP ⁺ cells from the adult WKM of four Tg(<i>c-myb-zRE1:PF5:egfp</i>) ^{qmc85} fish	110
3.1.2	All single GFP ⁺ kidney marrow cells of Tg(<i>c-myb-zRE1:PF5:egfp</i>) ^{qmc85} fish express neutrophil genes	116
3.1.3	The single-cell RNA Sequencing (scRNA-Seq) analysis of <i>qmc85</i> :GFP ⁺ kidney marrow cells reveals distinct neutrophil maturation stages	126

3.2	Verification of the DNA sequence of the region downstream of the <i>c-myb</i> gene	139
3.3	Deletion of zRE1 from the wild-type zebrafish genome using CRISPR/Cas9	142
3.4	Attempted CRISPR/Cas9-mediated deletion of three potential enhancers (PEs) downstream of zebrafish <i>c-myb</i> gene	156
3.5	Successful CRISPR/Cas9-mediated deletion of four potential enhancers downstream of <i>c-myb</i> from the wild-type zebrafish genome	160
3.6	<i>Lyz</i> and <i>c-myb</i> genes expression levels were unaffected in the compound heterozygous <i>c-myb</i>^{t25127/qmc193} or <i>c-myb</i>^{t25127/qmc194} mutant zebrafish embryos	175
3.6.1	<i>Lyz</i> and <i>c-myb</i> genes expression levels were normal in the compound heterozygous <i>c-myb</i> ^{t25127/qmc193} mutant zebrafish embryos	181
3.6.2	The expression level of <i>lyz</i> and <i>c-myb</i> genes were unchanged in the compound heterozygous <i>c-myb</i> ^{t25127/qmc194} mutant zebrafish embryos	185
3.7	Human regulatory element one (hRE1) is able to drive GFP expression in the zebrafish adult kidney	189
3.7.1	Establishment of stable human regulatory element one (hRE1) zebrafish transgenic line	192
3.7.2	Zebrafish embryos that carry the Tg(<i>c-myb</i> -hRE1-PF5: <i>egfp</i>) <i>qmc195</i> transgene display GFP reporter gene expression in cells of the caudal haematopoietic tissue.....	194
3.7.3	The <i>c-myb</i> -hRE1-PF5: <i>egfp</i> expression is retained in adult haematopoietic cells in the <i>qmc195</i> kidney	201
4	Discussion	206
4.1	Summary	206

4.2	Single-cell RNA Sequencing (scRNA-Seq) of whole kidney marrow (WKM) of adult zebrafish Tg(<i>c-myb</i> -zRE1:PF5: <i>egfp</i>) ^{qmc85} line	207
4.3	Deletion of zebrafish conserved non-coding element zCNE1 and potential enhancers (PEs) by CRISPR/Cas9 system	216
4.4	Incomplete functional conservation of human regulatory one (hRE1) in zebrafish transgenic line.....	222
5	Conclusion	235
6	Appendices	237
6.1	Appendix figures	237
6.2	Appendix Tables	265
7	References.....	294

List of Figures

Figure 1.1 Adult haematopoietic hierarchy.....	4
Figure 1.2 Embryonic haematopoietic sites in mouse embryos	6
Figure 1.3 Timeline of zebrafish haematopoiesis.....	19
Figure 1.4 Schematic illustration of human neutrophils.....	24
Figure 1.5 Schematic representation of the effects of <i>cis</i> -regulatory elements: promoters, enhancers and insulators	32
Figure 1.6 Pathway of C-MYB transcription factor network.....	41
Figure 1.7 The expression pattern of <i>c-myb</i> during zebrafish embryogenesis.	42
Figure 1.8 PF1-PF6 promoter-proximal elements are able to direct the EGFP expression of <i>c-myb</i> gene in non- haematopoietic tissues	45
Figure 1.9 Genomic sequence comparisons between vertebrate species identified a number of conserved non-coding elements (CNEs) in the <i>c-myb</i> locus	48
Figure 1.10 A map of the putative regulatory elements' fragments used to make the constructs to generate transgenic zebrafish lines	50
Figure 1.11 zRE1 and mRE1 are able to drive reporter genes expression in the different promoters to haematopoietic cells.....	53
Figure 1.12 Identification of potential enhancers (PEs)	56
Figure 1.13 CRISPR-Cas adaptive immunity in bacteria	60
Figure 2.1 Identification of suitable Cas9 target sites (TSs).....	74
Figure 2.2 sgRNA design and synthesis.....	78
Figure 2.3 Successful cloning of the TS7, 16 and 17 linkers.....	80
Figure 2.4 Amplification of hRE1 and cloning in pGEM-Teasy	83
Figure 2.5 Generation of the hRE1-zPF5: <i>egfp</i> reporter transgene – Part 1	86
Figure 2.6 Generation of the hRE1-PF5: <i>egfp</i> reporter transgene – Part 2	87
Figure 2.7 Gating strategy for the flow cytometric analysis of adult zebrafish whole kidney marrow (WKM) cells	107
Figure 3.1 Isolation of single GFP ⁺ cells from the adult kidneys of four Tg(<i>c-myb</i> -zRE1:PF5: <i>egfp</i>) ^{qmc85} fish	114
Figure 3.2 Calling cell barcodes.	118
Figure 3.3 Single-cell RNA Sequencing of zebrafish whole kidney marrow neutrophils of Tg(<i>c-myb</i> -zRE1:PF5: <i>egfp</i>) ^{qmc85} identified nine neutrophil cell clusters.	119

Figure 3.4 t-SNE visualization of single GFP+ kidney marrow cells of Tg(<i>c-myb</i> -zRE1:PF5: <i>egfp</i>) ^{qmc85} coloured by the number of UMI counts per cell	120
Figure 3.5 Single-cell RNA Sequencing reveals that all GFP+ cells that were isolated from the kidney marrow of Tg(<i>c-myb</i> -zRE1:PF5: <i>egfp</i>) ^{qmc85} fish express neutrophil genes.....	121
Figure 3.6 The correlation of selection of neutrophil markers	125
Figure 3.7 MA plots comparing human early neutrophil progenitors (ePreN) and neutrophils (Neuts) (Dinh et al., 2020).	129
Figure 3.8 Comparison of Dinh and colleagues' data with our human orthologues of zebrafish clusters one and two genes.	132
Figure 3.9 Up-regulation of the gene expression levels of human orthologues of zebrafish genes of cluster one in comparison of promyelocytes and myelocytes (P/Ms) relative to the segmented neutrophils (SNs)	137
Figure 3.10 Up-regulation of the gene expression levels of human orthologues of zebrafish genes of cluster two in comparison of the segmented neutrophils (SNs) relative to the promyelocytes and myelocytes (P/Ms).....	138
Figure 3.11 Verification of the DNA sequence downstream of <i>c-myb</i> ...	141
Figure 3.12 <i>c-myb</i> 2&7 sgRNAs successfully produced deletion in the target sites 2 and 7	144
Figure 3.13 Germline transmission of the <i>c-myb</i> TS2 and TS7 zRE1 deletion allele.....	148
Figure 3.14 Identify homozygous mutants in F2 embryos	150
Figure 3.15 Identification of F2 homozygous <i>qmc193</i> mutants' adult fish	152
Figure 3.16 <i>Lyz</i> and <i>c-myb</i> expression were maintained in 34-36 hpf and 4 dpf <i>qmc193/qmc193</i> embryos	155
Figure 3.17 <i>c-myb</i> 16&2 sgRNAs successfully produced deletion in the target sites 16 and 2	158
Figure 3.18 Successful CRISPR/Cas9-mediated deletion of three putative enhancer elements as well as zRE1 downstream of the <i>c-myb</i> gene....	162
Figure 3.19 Establishment of a zebrafish line that carries the deletion allele <i>qmc194</i> that lacks three potential enhancer elements as well as zRE1 downstream of <i>c-myb</i>	166
Figure 3.20 Identification of homozygous <i>qmc194</i> mutant embryos in the F2 progeny of two heterozygous <i>qmc194</i> F1 carriers.....	168
Figure 3.21 Identification of adult homozygous <i>qmc194</i> carriers in the progeny of two <i>qmc194</i> heterozygous fish.....	170
Figure 3.22 <i>Lyz</i> and <i>c-myb</i> expression are unaffected in 34-36 hpf and 4 dpf <i>qmc194/qmc194</i> embryos, respectively.....	174

Figure 3.23 The expression of <i>lyz</i> and <i>c-myb</i> are missing in the <i>c-myb</i> ^{t25127/t25127} embryos	178
Figure 3.24 Reduced /or loss of the expression of some genes in the <i>c-myb</i> ^{t25127/t25127}	180
Figure 3.25 <i>Lyz</i> expression was maintained in 34-36 hpf in <i>c-myb</i> ^{t25127/qmc193} embryos	183
Figure 3.26 <i>c-myb</i> expression was maintained in 3 dpf <i>c-myb</i> ^{t25127/qmc193} embryos	184
Figure 3.27 The expression level of <i>lyz</i> in 34-36 hpf was not changed in the compound heterozygote of <i>c-myb</i> ^{t25127/qmc194} embryos	187
Figure 3.28 The expression level of <i>c-myb</i> at 4 dpf was not altered in the compound heterozygote of <i>c-myb</i> ^{t25127/qmc194} embryos	188
Figure 3.29 Localisation of the conserved non-coding element one downstream of human <i>C-MYB</i> gene.....	191
Figure 3.30 Establishment of the stable human regulatory element one (hRE1) zebrafish transgenic line by injecting the Tol2 pJHE1- <i>egfp</i>	193
Figure 3.31 Expression of EGFP reporter gene in in the caudal haematopoietic tissue (CHT) of 2-day old embryos of Tg(<i>c-myb</i> -hRE1-PF5: <i>egfp</i>) ^{qmc195} transgene	197
Figure 3.32 Expression of <i>GFP</i> mRNA in CHT of Tg(<i>c-myb</i> -hRE1-PF5: <i>egfp</i>) ^{qmc195} transgenic zebrafish embryos. (A-D)	200
Figure 3.33 Flow cytometric analysis show that Tg(<i>c-myb</i> -hRE1:PF5: <i>egfp</i>) ^{qmc195} transgenic fish carry GFP+ haematopoietic progenitor and myelomonocytic cells in kidney marrow	204
Figure 4.1 Single-cell RNA Sequencing of Tg(<i>c-myb</i> - zRE1:PF5: <i>egfp</i>) ^{qmc85} of zebrafish kidney marrow reveals that cluster one cells express stem or early progenitor genes	215
Figure 4.2 Putative binding sites of transcription factors on conserved non-coding element one (CNE1)	232
Figure 4.3 Putative binding sites of transcription factors on individual sequences of zebrafish, mouse and human conserved non-coding element one (CNE1)	233
Figure 4.4 Localisation of the sites of DNase1 HS within the conserved non-coding element one downstream of human <i>c- MYB</i> gene.....	234

List of Tables

Table 1.1 A list of potential <i>c-myb</i> downstream enhancers (PEs) identified as H3K4me1 binding sites in chromatin isolated from 24 hpf zebrafish embryos in a CHIP-Seq experiment published by Aday et al., 2011	57
Table 2.1 List of buffers and solutions.....	65
Table 2.2 Sample QC.....	70
Table 2.3 Library Preparation QC	70
Table 2.4 List of the sequences and oligonucleotides used for targets site	75
Table 2.5 List of primers used for genotyping of <i>C-MYB</i> hRE1	81
Table 2.6 List of primers used for genotyping and sequencing of <i>c-myb</i> mutants.....	90
Table 2.7 PCR primers of <i>c-myb</i> map and locations.....	91
Table 2.8 PCR thermocycling conditions	92
Table 2.9 PCR reaction volumes.....	92
Table 2.10 The general ligation reaction.....	94
Table 2.11 The general restriction digest reaction.....	95
Table 2.12 The ligation mix to insert the linker into the pDR274 vector	97
Table 2.13 The Mixture reaction to make <i>in vitro</i> transcription	99
Table 2.14 Incubation times in Proteinase K of embryos varied with the age of embryos	102
Table 2.15 List of WISH probes used in this project	103
Table 3.1 Total transcript (UMI) counts in the entire cohort of 2246 GFP ⁺ kidney marrow cells isolated from adult Tg(<i>c-myb</i> -zRE1:PF5: <i>egfp</i>) ^{qmc85} fish	124
Table 3.2 Identified Tg(<i>c-myb</i> -hRE1-PF5: <i>egfp</i>) transgenic founders....	195

List of Appendix Figures

Appendix figure 6. 1 A sequence of the human <i>C-MYB</i> locus. Orange bars represent the PCR product. Grey boxes present the polymorphisms....	237
Appendix figure 6.2 Single-cell RNA Sequencing of Tg(<i>c-myb-zRE1:PF5:egfp</i>) ^{qmc85} of zebrafish kidney marrow express neutrophil genes	238
Appendix figure 6.3 A sequence of the zebrafish <i>c-myb</i> locus. Orange bars represent the PCR products. Grey boxes present the polymorphisms. .	239
Appendix figure 6. 4 <i>lyz</i> expression was maintained in 34-36 hpf <i>qmc193</i> embryos	247
Appendix figure 6.5 <i>c-myb</i> expression was maintained in anterior primitive myeloid and posterior primitive erythroid progenitors of all 15 ss <i>qmc193</i> embryos	248
Appendix figure 6. 6 <i>c-myb</i> expression was maintained in 30 hpf <i>qmc193</i> embryos	249
Appendix figure 6. 7 <i>c-myb</i> expression was maintained in 2 dpf <i>qmc193</i> embryos	250
Appendix figure 6.8 <i>c-myb</i> expression was not altered in 4 dpf <i>qmc193</i> embryos	251
Appendix figure 6.9 <i>lyz</i> expression is unaffected in 34-36 hpf <i>qmc194</i> embryos	252
Appendix figure 6.10 <i>c-myb</i> expression was similar in all 15 ss embryos derived from an incross of <i>qmc194</i> heterozygous carriers	253
Appendix figure 6.11 The level of <i>c-myb</i> expression was comparable in all 30 hpf embryos derived from an incross of <i>qmc194</i> heterozygous carriers	254
Appendix figure 6.12 The level of <i>c-myb</i> mRNA was similar in all 2 dpf embryos derived from an incross of <i>qmc194</i> heterozygous carriers	255
Appendix figure 6.13 <i>c-myb</i> expression is unaffected in 4 dpf <i>qmc194</i> embryos	256
Appendix figure 6.14 <i>c-myb</i> expression was unaffected in 15 ss <i>t25127</i> embryos	257
Appendix figure 6.15 <i>c-myb</i> expression was maintained in 30 hpf <i>t25127</i> embryos	258
Appendix figure 6.16 <i>c-myb</i> expression was maintained in 2 dpf <i>t25127</i> embryos	259
Appendix figure 6. 17 <i>Lyz</i> expression was maintained in 34-36 hpf in <i>c-myb</i> ^{t25127/qmc193} embryos	260
Appendix figure 6.18 <i>c-myb</i> expression was maintained in 3 dpf in <i>c-myb</i> ^{t25127/qmc193} embryos	261
Appendix figure 6.19 The expression level of <i>lyz</i> in 34-36 hpf was not changed in the compound heterozygote of <i>c-myb</i> ^{t25127/qmc194} embryos .	262

Appendix figure 6.20 The expression level of *c-myb* in 4 dpf was not altered in the compound heterozygote of *c-myb*^{t25127/qmc194} embryos 263

Appendix figure 6.21 Identification of Tg(*c-myb*-hRE1-PF5:*egfp*) transgenic founders 264

List of Appendix Tables

Appendix table 6.1 List of the differently expressed ($p < 0.05$) genes that are more highly expressed in cluster one than cluster two. 265

Appendix table 6.2 List of the differently expressed ($p < 0.05$) genes that are more highly expressed in cluster two than cluster one. 267

Appendix table 6.3 The list of human orthologues of cluster one-specific zebrafish genes..... 275

Appendix table 6.4 The list of human orthologues of cluster two-specific zebrafish genes..... 277

Appendix table 6.5 The list of the human orthologues of cluster one-specific zebrafish genes that were found within the Dinh et al dataset in SeqMonk. 283

Appendix table 6.6 List of human orthologues of cluster one-specific zebrafish genes that were significantly upregulated in human ePreNs relative to Neuts in the Dinh et al. dataset. 285

Appendix table 6.7 The list of the human orthologues of cluster two-specific zebrafish genes that were found within the Dinh et al dataset in SeqMonk. 286

Appendix table 6.8 List of human orthologues of cluster two-specific zebrafish genes that were significantly upregulated in human Neuts relative to ePreNs in the Dinh et al. dataset. 291

Appendix table 6.9 List of transcription factors (TFs) harbouring conserved binding sites in conserved non-coding element one (CNE1). 293

Abstract

Haematopoiesis is the process by which all blood cell types form in a vertebrate organism. *c-myb* is an essential transcription factor that is widely expressed in haematopoietic stem and progenitor cells and is known to induce leukaemia when overexpressed. During normal haematopoiesis, it plays important roles during stem cell formation, self-renewal, maintenance, proliferation and differentiation of HSCs/progenitors. To gain a deeper insight into the transcriptional regulation of this key transcription factor, it is important to learn more about the *cis*-regulatory DNA elements that control its expression. *Cis*-regulatory DNA elements, like promoters and enhancers, are generally less conserved than the corresponding coding sequences. Nevertheless, long-range genomic DNA sequence alignments have identified a conserved non-coding DNA element (CNE1) downstream of the *c-myb* gene that is remarkably conserved from fish to man. Testing the zebrafish regulatory element (zRE1) containing zebrafish conserved non-coding element (zCNE1) and its mouse equivalent mRE1 in stable transgenic zebrafish lines, named *qmc85* and *qmc156*, has revealed that both are sufficient to direct zebrafish *c-myb* promoter activity to haematopoietic cells. Both lines display GFP expression in blood cells of the larval caudal haematopoietic tissue and the adult kidney marrow (KM). In this study, single cell RNA-sequencing (scRNA-Seq) was used to determine the gene expression profile and, thereby, the nature of

the GFP+ KM cells in the *qmc85* line. The scRNA-Seq data revealed that all of the cells expressed lysozyme, a marker for neutrophilic granulocytes. Gene transcripts that are usually associated with other blood cell lineages were not found at significant levels. Instead, the cells expressed numerous additional neutrophil genes at levels that correlated with those of the lysozyme mRNA. Further analysis shows that the neutrophils were at different maturation stages. To investigate whether zRE1 containing zCNE1 and three potential *cis*-regulatory elements (PEs) are necessary for endogenous *c-myb* expression, CRISPR/Cas9 technology was used to delete the elements from the zebrafish *c-myb* locus. Successful deletion of zRE1 alone and in combination with all three PEs yielded the mutant alleles *qmc193* and *qmc194*, respectively. Embryos homozygous for the two mutant alleles did not display an obvious reduction in *c-myb* expression in blood cells of the CHT, and neither did embryos that were compound heterozygous for either *qmc193* or *qmc194* and the previously published *c-myb* null allele *t25127*. Larvae that are homozygous for *qmc193* and *qmc194* were viable and could be raised to adulthood. These data suggest that zCNE1, though sufficient to direct *c-myb* promoter activity to haematopoietic cells (neutrophil lineage) but is not essential for *c-myb* expression in those cells. Testing the human hRE1 element containing hCNE1 in the stable transgenic line *qmc195* revealed that hCNE1 is able to

direct transgene expression in embryonic haematopoietic progenitors and in the adult kidney marrow.

1 Introduction

Haematopoiesis is the process of blood cell generation that is maintained by haematopoietic stem cells (HSCs). HSCs are immature cells that can self-renew and are multipotent. HSCs give rise to all lineage-committed haematopoietic progenitors with enormous proliferative potential (Drissen et al., 2016) (Figure 1.1). In mammals, HSCs reside in the bone marrow of adults, maintaining the adult haematopoietic system throughout life. Because of these multipotency and self-renewal characteristics of HSCs, they can be used to treat blood diseases. They can be transplanted into the patient after chemotherapy (used to eliminate abnormal cells) to restore blood formation after the treatment (Birbrair and Frenette, 2016). Clinically, HSCs are important in transplantation procedures used for the treatment of blood disorders or cancers such as leukaemia and lymphoma (Laurenti and Göttgens, 2018). Currently, HSCs used for transplantation are obtained from bone marrow, peripheral blood after mobilisation, and the umbilical cord. In an allogeneic stem cell transplant, the donor stem cells produce immune cells, which helps destroy any leukaemia cells in the recipient after chemotherapy treatment. This is called the graft-versus-cancer effect. Despite the importance of HSC in medicine, graft versus host disease, a condition resulting from an allogeneic transplant, has remained a significant problem complication of haematopoietic stem cells transplantation (reviewed in (Jeevani,

2011)). Some of these complications could be resolved by reprogramming somatic cells to induced pluripotent stem cells (iPSCs) through modulation of the expression of some TFs. The derived iPSCs could be transfused back to the patient for the treatment process (Alsayegh et al., 2019).

In vertebrates, the formation of HSCs requires the coordinated action of many transcription factors (TFs) (Jagannathan-Bogdan and Zon, 2013). *c-Myb* is an essential transcription factor involved in the formation of HSCs during development and is expressed in HSCs and progenitors (Bengtson et al., 2015, Lorenzo et al., 2011). Expression of the *c-myb* gene is tightly regulated by the interaction between TFs with *c-myb*'s *cis*-regulatory elements (CREs) (Zhang et al., 2016). In general, CREs include promoter and enhancer elements that are typically located in non-coding genomic sequences (Andersson and Sandelin, 2019, Pennacchio et al., 2013). Identifying the CREs and their function is fundamental and will help us to understand the regulation of *c-myb* in HSCs during embryonic development. The formation of HSCs was studied by analysing the regulation of the *c-myb* gene in zebrafish embryos during haematopoiesis. In vertebrates, numerous CRE sequences have been conserved during evolution and are often referred to as conserved non-coding elements (CNEs) (Nelson and Wardle, 2013). Recently, we have studied the role of the distal conserved non-coding element one (CNE1) downstream of *c-myb* that was previously identified in our

lab. This CNE1 has been tested in combination with the *c-myb* promoter and an enhanced green fluorescent protein (EGFP) reporter gene in transgenic zebrafish lines. In zebrafish, CNE1 was able to drive reporter gene expression in sites of definitive haematopoiesis, such as the caudal haematopoietic tissue (CHT), kidney and thymus (Hsu, 2010).

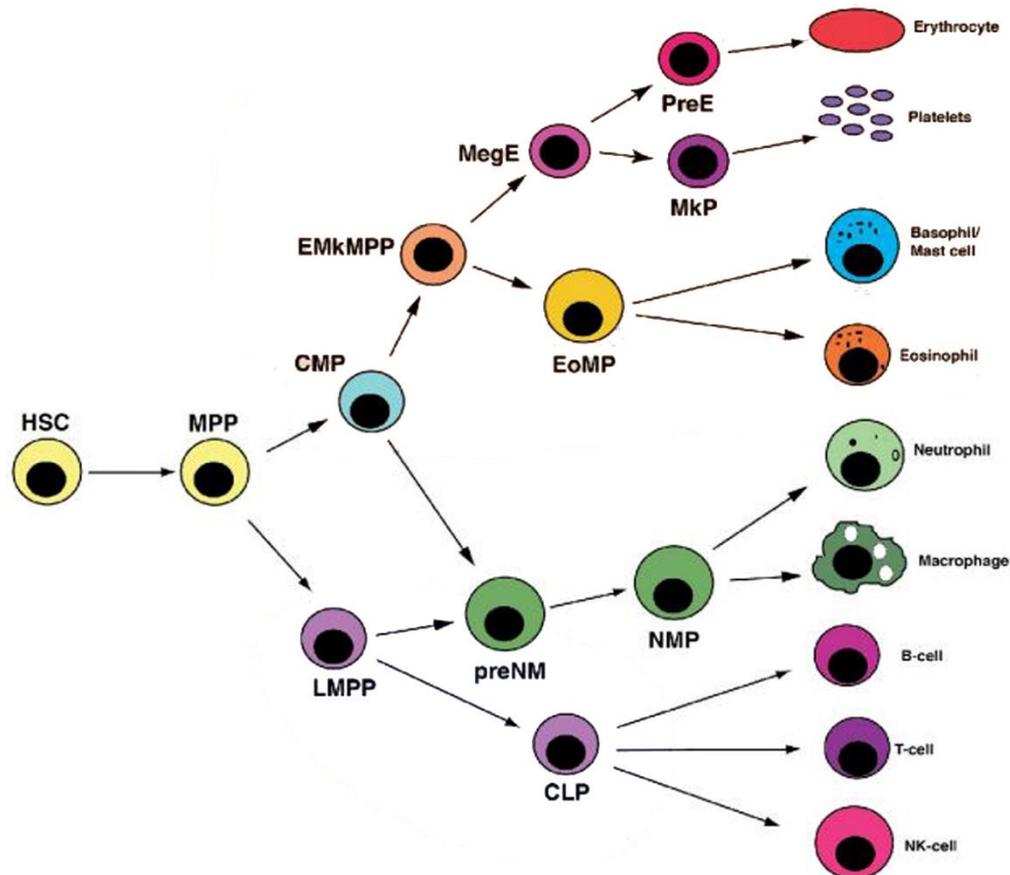


Figure 1.1 Adult haematopoietic hierarchy

In this model, HSCs give rise to multipotent progenitors (MPPs), which differentiate into either common myeloid progenitors (CMPs) or lymphoid-myeloid primed progenitors (LMPPs). CMPs placed upstream of erythroid-megakaryocyte-primed multipotent progenitors (EMkMPPs), and pre-neutrophil-monocytes (preNMs). EMkMPP can give rise to megakaryocyte-erythrocyte lineage (MegE) that can differentiate into erythrocytes and platelets or eosinophil-mast cell progenitors (EoMPs) that can differentiate into basophil/mast cells and eosinophils. LMPPs placed upstream of common lymphoid progenitors (CLPs) and preNMs. The CLPs can give rise to B-cells, T-cells and natural killer (NK) cells. The preNMs can differentiate into neutrophil-monocyte progenitors (NMPs) which can give rise to neutrophils and macrophages (Image from (Drissen et al., 2016)).

1.1 Haematopoiesis

Haematopoiesis is the process that forms all types of blood cells. In vertebrates, it is an evolutionarily conserved process through which blood cells are created and maintained throughout life. It begins during embryogenesis and continues throughout the lifetime of an organism. This section will focus on the different waves of haematopoiesis in the mouse and the zebrafish.

1.1.1 Haematopoiesis in mouse

Haematopoiesis takes place in various anatomical sites that change spatially and temporally in vertebrates. Haematopoiesis is conserved across all vertebrates with only minor differences. In the embryo, the mesodermal germ layer is responsible for producing the HSCs that give rise to the adult haematopoietic system. In mammalian embryos, haematopoiesis occurs by sequential events that originates extra-embryonically. Subsequently, the aorta-gonad-mesonephros (AGM) region is the site where intra-embryonic haematopoiesis occurs. Here, the earliest HSCs emerge. HSCs subsequently migrate to the foetal liver (FL) where they remain until birth. Around the time of birth, HSCs seed the bone marrow (BM), which is the main niche for adult HSCs throughout life (Figure 1.2) (Dzierzak and Medvinsky, 1995, Medvinsky and Dzierzak, 1996, de Bruijn et al., 2000a, Frame et al., 2013). In mammalian embryogenesis, haematopoiesis is divided into three waves: the primitive wave, the intermediate wave and the definitive wave.

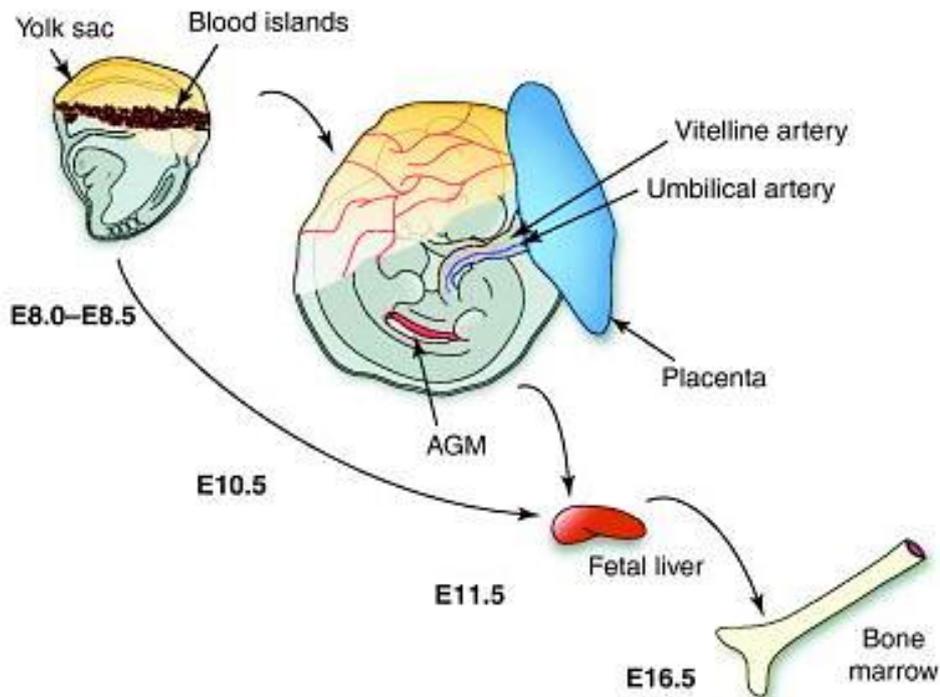


Figure 1.2 Embryonic haematopoietic sites in mouse embryos

The primitive haematopoietic cells are found extra-embryonically before E8.0 in the yolk sac blood islands. After E9.0, pre-definitive erythromyeloid progenitors arise and seed the foetal liver from E9.75 onwards. Cells from the HSC lineage are generated in the AGM region and also colonise the foetal liver where all the progenitors develop and become mature. Finally, at E16.5, the bone marrow is colonised by HSCs and becomes the main haematopoietic niche throughout adult life (Image from (Costa et al., 2012)).

1.1.1.1 Primitive haematopoiesis in mouse

The primitive wave is the initial and transitory wave of haematopoiesis; it is rapidly replaced by adult-type haematopoiesis that is called definitive. In the mouse, it begins at embryonic day 7 (E7.0) and gives rise to primitive erythrocytes and some of the myeloid cells. Primitive erythrocytes contain haemoglobin and are required for the immediate supply of oxygen, which is needed for rapid development (Orkin and Zon, 2008a).

During mouse development, the first primitive haematopoietic cells arise from the extra-embryonic YS (Yamane, 2018). This consists of two layers of cells, the visceral endoderm and the extra-embryonic mesoderm. The latter gives rise to the endothelium and the blood cells of the YS (Silver and Palis, 1997). In sites of primitive haematopoiesis and based on the emergence of the close relationship between haematopoietic and endothelial cells in the mouse foetus, it is believed that the blood and endothelial lineages derive from a common bipotential progenitor known as the haemangioblast. During gastrulation, haemangioblasts were first detected at the mid-streak stage then peaked in number during the neural plate stage (Huber et al., 2004). YS-derived mesoderm cells that are haemangioblasts and have the ability to give rise to transient haematopoietic cells and endothelial cells (Fraser and Baron, 2009, Gritz and Hirschi, 2016). These common progenitors may also produce smooth muscle cells (Ema et al., 2003). Molecular

analysis of *in vitro* blast colony-forming cells (BL-CFCs) derived from E7.0–7.5 mouse embryo indicates that haemangioblasts express genes related to haematopoiesis such as Scl, Gata1 and β H1 globin and vascular development genes including Flk-1 and VE-cadherin (Huber et al., 2004). At E7.5, haemangioblasts begin to differentiate into primitive haematopoietic cells. These cells can give rise to erythrocytes, macrophages and megakaryocytes. In mice, the levels of Gata1 and Pu.1 (now known as Spi1 in mouse) transcription factors are important for controlling the primitive erythroid or myeloid cell fates (Cantor and Orkin, 2002). Between E8.0 and E9.0, the blood island in the YS contains clusters of erythroid cells, which are known as primitive erythrocytes (primitive red blood cells (prRBCs)) and are surrounded by endothelial cells. These prRBCs are naturally large in size, are nucleated and contain the embryonic type of haemoglobin and establish an oxygen supply in the developing foetus (Kingsley et al., 2004). At the same time, blood vessels form and become connected, allowing the haematopoietic progenitor cells to enter the primitive circulation. At E9.5, and in addition to the erythroid cells, megakaryocytes and macrophages have been observed in the YS. Macrophages can form the microglial precursors, which then move to the developing central nervous system. None of these progenitors is pluripotent and loses their self-renewal (Tober et al., 2006, Palis et al., 1999).

1.1.1.2 The intermediate wave of mouse haematopoiesis gives rise to erythromyeloid progenitors in the yolk sac

The intermediate wave is a transient wave of haematopoiesis that is less characterised but clearly distinct and precedes the start of the definitive wave. It occurs in the YS and leads to the generation of erythromyeloid progenitors (EMPs) (Ferkowicz et al., 2003, Palis et al., 1999). In the yolk sac, EMPs are developed, giving rise to foetal erythrocytes, tissue-resident macrophages, megakaryocytes and granulocytes and can stay around for longer than E16.5 (Palis et al., 1999, Perdiguero et al., 2015, Soares-da-Silva et al., 2021). At E9, a population of cells isolated from the yolk sacs of mice lacking circulation were able to give rise to a subset of lymphoid B-cells upon transplantation (Yoshimoto et al., 2011). At around the time of E9.5, EMPs seed the FL, meaning that they can be present in the embryo before the definitive haematopoietic stem cells (dHSCs) and progenitors emerge. Unlike dHSCs, the EMPs are not capable of self-renewal and long-term repopulation (Bertrand et al., 2007, Chen et al., 2011, McGrath et al., 2011).

1.1.1.3 Definitive haematopoiesis in mouse

Definitive haematopoiesis is the third wave of haematopoiesis in the mouse embryo. It is the wave that forms long-term haematopoietic stem cells (LT-HSC). In this wave, the first HSCs are linked with the presence of cell clusters in the aorta, umbilical artery, and vitelline artery in mouse embryos (de Bruijn et al., 2000b). The definitive

HSCs produced in AGM at E10.0 were examined using the long-term repopulating activity (LTRA) assay (Medvinsky and Dzierzak, 1996). The definitive HSCs were also produced from both the dorsal aorta and its surrounding mesenchyme as well at the urogenital ridge from dissected AGM at E10 and E11, respectively, by organ culture and transplantation experiments (de Bruijn et al., 2000b). At E10.0 to E10.5, HSC progenitors appear in the ventral wall of the dorsal aorta (vDA) in the AGM region (Taoudi and Medvinsky, 2007). The vDA is the site of development and emergence of HSCs in all other vertebrates studied so far (Medvinsky et al., 2011, Sugiyama et al., 2011). HSCs are derived from the endothelial cells that line the vDA via a process known as endothelial to haematopoietic transition (EHT) (Ottersbach, 2019). These endothelial cells are called haemogenic endothelium cells (HECs) (Boisset et al., 2010, Jaffredo et al., 1998). HECs have the ability to give rise to HSCs during a process that is highly conserved among vertebrates (Carroll and North, 2014, Sturgeon et al., 2013). Therefore, during the definitive wave, these cells are responsible for the formation of all blood cell types.

After the HSC progenitors are generated in the AGM region, they leave the vDA and enter the dorsal aorta before joining the circulation. They subsequently migrate to colonise the FL. Several studies suggest that the FL does not produce haematopoietic cells itself but is colonised by EMPs and HSCs that come from both YS and

AGM between E9.0 and E11.0 (Cumano and Godin, 2007, Sugiyama et al., 2011). At around the time of E13.5, the number of HSC increases dramatically in the FL, and the HSCs mature and differentiate into various haematopoietic cell types. Pu.1 is a factor belonging to the ETS family of transcription factors that regulates the maintenance and development of HSCs in the FL (Rybtsov et al., 2016). After E16.5 and before birth, HSCs progressively migrate to the BM, the main adult haematopoietic organ, where they seed (Mazo et al., 2011). *Runx1* knockout mice fail to generate the definitive erythroid, myeloid and lymphoid cells demonstrating the importance of *Runx1* in definitive haematopoiesis (Okuda et al., 1996, Wang et al., 1996). Runx1 transcription factor (also known as AML1, Cbfa2, or Pebp2 α) is absolutely essential for definitive haematopoiesis in the mouse (Yzaguirre et al., 2017). Cells derived from AGM of the *Runx1*-null embryos failed to generate definitive HSCs in irradiated mice after transplantation, suggesting the importance of Runx1 for functional HSCs (Cai et al., 2000). Runx1 is an essential master regulator of endothelial-to-hematopoietic transition (ETH) (Chen et al., 2009). The expression of *Runx1* was observed in the endothelial cells at all sites of haematopoietic cluster formation, confirming that it is required for the ETH process (North et al., 1999, Yzaguirre et al., 2017). It has been found that the mouse embryos with a homozygous mutation in *Runx1* had normal primitive erythrocytes and endothelial cells, which were derived from

the blood islands of the yolk sac but were deficient in foetal liver haematopoiesis and died at around E12.5 (Okuda et al., 1996). Notch1 signaling plays an essential role in the development of definitive hematopoietic cells. In mice, Notch1-null cells fail to contribute to long-term definitive haematopoiesis (Hadland et al., 2004). In addition, the c-Myb transcription factor is also required for definitive haematopoiesis. *c-Myb*-null mouse embryos suffer from severe anaemia and die due to impaired definitive haematopoiesis, however primitive haematopoiesis is normal in these mouse embryos (Mucenski et al., 1991).

1.1.2 The zebrafish as a model for haematopoiesis

During the past 20 years, the zebrafish system has served as a powerful model for research into vertebrate haematopoiesis. Compared to other vertebrate models, the zebrafish model has several advantageous characteristics, it displays external development, has a short life cycle, and is amenable to genetic manipulation. In addition, large numbers of embryonic offspring can make phenotype-based, forward genetics studies possible (de Jong and Zon, 2005). Another advantage of zebrafish is that their embryos are transparent, making visualisation of the early stages of *in vivo* development possible. Furthermore, the external development of zebrafish facilitates the simple injection of enzymes, nucleotides and morpholinos into one-cell-stage embryos enabling its use in genetic experiments. Most importantly, they are similar to

other “higher-order” vertebrates in cellular structure, signalling and physiology. In addition, many of the gene functions are conserved in mammals, making the zebrafish an ideal model for biological research. Understanding the functionality of the hematopoietic system in zebrafish will also help to improve the understanding of the human hematopoietic system and the functions of particular genes in human blood diseases (Berghmans et al., 2005, Gore et al., 2018, Shafizadeh and Paw, 2004). New knowledge obtained from the zebrafish can provide insights that are also applicable to mice and humans. For example, the generation of HSC progenitors is a reasonably conserved process analogous in fish and mammals (Veldman and Lin, 2008).

1.1.3 Haematopoiesis in zebrafish

During zebrafish embryogenesis, as in other vertebrates, haematopoiesis occurs in three waves from the mesodermal germ layer with shifting sites, the primitive, intermediate and the definitive waves each of which will be explained in this section (Figure 1.3). However, unlike in the mouse, all the waves take place intra-embryonically in zebrafish.

1.1.3.1 Primitive haematopoiesis in zebrafish

In zebrafish, primitive haematopoiesis is the first wave of blood development and starts at 10 hours post-fertilisation (hpf) (Carroll and North, 2014). The generation of HSCs requires the processes

from the gastrulation stage when the three germ layers (ectoderm, mesoderm and endoderm) are formed at 5 hpf. In particular, during the segmentation period, the ventral lateral mesoderm produces bipotent haemangioblasts.

Within the ventral lateral mesoderm, the anterior lateral mesoderm (ALM) is the predominant site of primitive myelopoiesis, whereas the posterior lateral mesoderm (PLM) is the major site of the primitive erythroid cell (primitive red blood cells (prRBCs) development in the embryo (Paik and Zon, 2010). In addition, whereas the angioblasts of the ALM produce the head vasculature, those of the PLM form the trunk and tail vessels, including the dorsal aorta (DA) and the axial vein, which are also called the posterior cardinal vein (PCV) (Gering et al., 1998, Gore et al., 2012). Around the 2-somite stage (10 hpf), haemangioblasts are identified as cells that express several genes characteristic for haematopoietic cells (such as *scf*, *lmo2*, *gata2*) and vascular cells (for example, *fli1* and *flk1*) and can be detected in both the ALM and the PLM (Brown and Wittwer, 2000, Chen and Zon, 2009, Davidson and Zon, 2004, Patterson et al., 2007). The transcription factor *scf* is expressed in primitive haematopoietic cells and is required and sufficient to drive blood cell development by activating the expression of myeloid genes such as *pu.1* in the ALM as well as erythroid genes such as *gata1* in the PLM. The loss of both primitive and definitive hematopoietic cell lineages has been found in *scf* morphants in the zebrafish embryos (Dooley et al., 2005). It

has been shown that the over-expression of *scl* in zebrafish leads to an increase in the number of haematopoietic and endothelial cells (Gering et al., 1998). The expression of *scl* is reduced in the zebrafish *cloche* mutant, which defects both blood and endothelial cells (Liao et al., 1998, Stainier et al., 1995). A study in *cloche* indicates that transcription factor Npas4 regulates the early stage of endothelial and haematopoietic progenitor differentiation (Reischauer et al., 2016).

At the 4-somite stage (11 hpf), within the ALM, the formation of either myeloid cells or endothelial (vascular) cells have been identified by the expression of *pu.1* and *flk-1*, respectively (Bennett et al., 2001, Lieschke et al., 2002). At the 6-somite stage (12 hpf), the cells with myelopoietic ability can be found as the *pu.1* expressing cells emerge in the rostral lateral plate mesoderm (RLPM) of the ALM (Lieschke et al., 2002). At around 14-somite stage (16 hpf), the *pu.1*+ cells display medial migration around the yolk sac then gather at the midline between the eyes and above the developing heart in a region known as the rostral blood island (RBI) (Bennett et al., 2001, Paik and Zon, 2010). In the RBI, embryonic myelopoiesis in zebrafish occurs, and the myeloid progenitors then differentiate to give rise to either macrophages or neutrophils (Le Guyader et al., 2008). The level of *pu.1* can determine the choice of macrophage or neutrophil lineage. A high level of *pu.1* has been indicated to favour macrophage differentiation over neutrophil

differentiation. In contrast, the low level of *pu.1* is sufficient and essential for neutrophil development (Jin et al., 2012). During neutrophil development in zebrafish, *runx1* regulates embryonic myeloid fate decisions by promoting granulocytic over macrophage lineage. Runx1 binds to *pu.1* promoter and suppresses *pu.1* expression (Jin et al., 2012). At around 18 hpf, once at the midline, myeloid cell differentiation begins. The myeloid progenitor cells start to express *I-plastin* in macrophages and *mpx* (myeloperoxidase) in granulocytes (neutrophil committed cells) (Galloway et al., 2005, Lieschke et al., 2001). Jin and colleagues demonstrated that the *ceb1* (CCAAT/enhancer-binding protein 1) gene can be detected at 18 hpf and is the earliest neutrophil marker in the RBI. It was also found that the expression of *ceb1* significantly preceded that of *lyz* (lysozyme C encodes a primary and secondary granule protein in granulocytes) and *mpx*, are considered to be neutrophil specific markers in zebrafish (Jin et al., 2012). The transgenic reporter lines for the *mpx* (*mpx:GFP*) (Renshaw et al., 2006) and *lyz* (*Lyz:dsRed*) (Hall et al., 2007) marker can be used to study the neutrophils and their behaviour during inflammation. Upon initiation of embryonic circulation, the primitive myeloid cells enter the circulation and later are distributed throughout the zebrafish embryonic tissues (Hall et al., 2007).

In the PLM, the haemangioblasts that give rise to endothelial and erythroid cells can be identified by the expression of *scl*. In the PLM,

at the 6-somite stage (12 hpf), the emergence of the erythrocyte is marked by the expression of *gata1* (Galloway *et al.*, 2005). At the 10-somite stage (14 hpf), the erythroid progenitors and endothelial progenitors of the PLM begin to migrate medially to the midline between the notochord, the endoderm and the somites to form the intermediate cell mass (ICM) in the trunk and the posterior blood island (PBI) in the tail region of the zebrafish embryo (Paik and Zon, 2010). Moreover, between 10 and 20 somites stage, *pu.1* can be detected in the ICM blood cells which then express *mpx* and *l-plastin* as granulocyte and macrophage markers, respectively (Bennett *et al.*, 2001). At 22 hpf, the onset of heart contractions occurs. By 24 hpf, the ICM gives rise to the erythroid cells, the dorsal aorta and the posterior cardinal vein. While in circulation, *gata1*⁺ cells continue to mature, expressing genes involved in erythrocyte development, such as β -globin (Brownlie *et al.*, 2003). The primitive erythrocytes enter into the circulation between 24 and 26 hpf, after the cardinal vein is formed.

1.1.3.2 Erythromyeloid progenitors arise in the posterior blood island (PBI) during the intermediate wave

During the second wave in the zebrafish embryo and similar to the mouse embryo, haematopoiesis shifts to a transient wave that gives rise to a population of erythromyeloid progenitors (EMPs) after the initiation of circulation. This wave occurs in the PBI, a posterior

extension of the ICM (Bertrand et al., 2007). In the PBI, the EMPs have erythroid and myeloid differentiation potential but lack lymphoid potential. The EMPs arise before HSC emerge in the AGM and, unlike the HSCs, lack the capacity for self-renewal (Ciau-Uitz et al., 2014, Jin et al., 2007). The EMPs fate is controlled by cross-antagonism between transcription factors: Gata1, which is essential for erythroid differentiation and Pu.1, which is required for the differentiation of myeloid cells (Rhodes et al., 2005). However, while the transient population of EMPs might give rise to a fraction of embryonic myeloid cells, the EMPs do not colonise the thymus and populate the kidney marrow (KM) (Carroll and North, 2014).

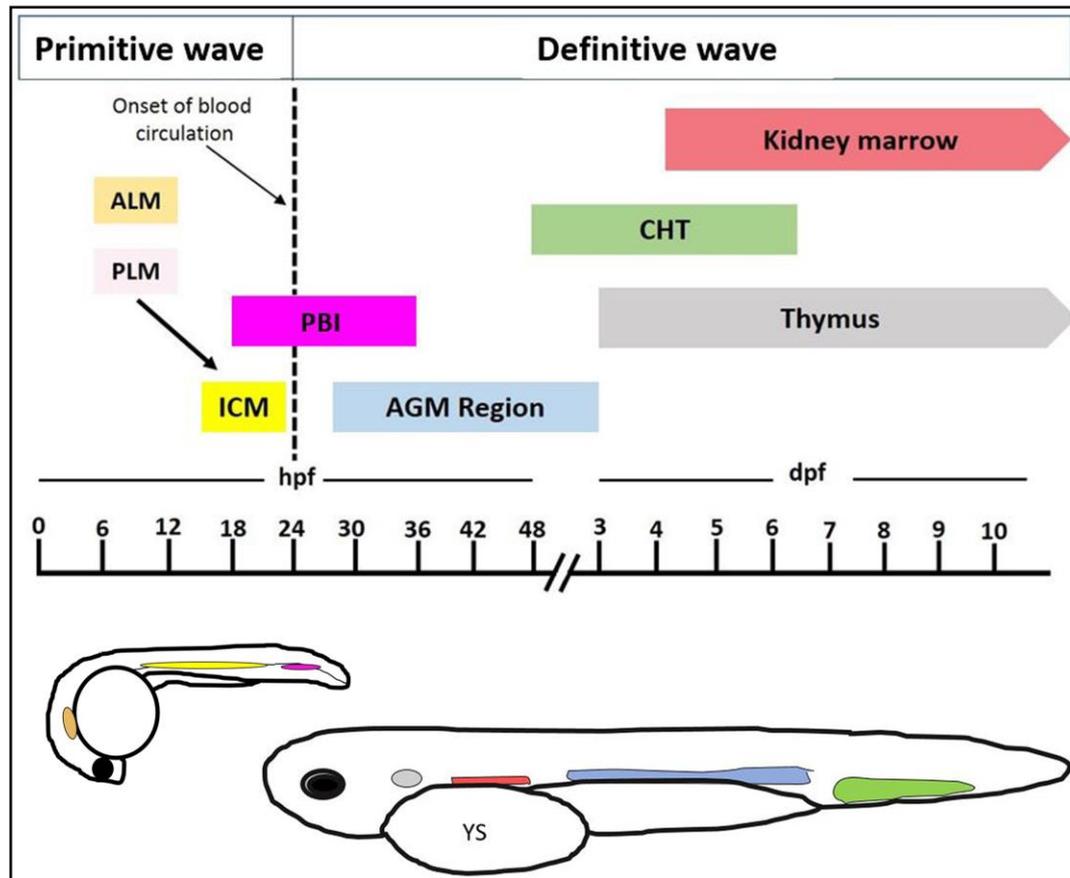


Figure 1.3 Timeline of zebrafish haematopoiesis

The development of blood cells of the primitive wave starts at 10 hpf when prRBCs emerge in the posterior lateral mesoderm (PLM), and the macrophage and neutrophil progenitors emerge in the anterior lateral mesoderm (ALM) (orange). By 14 hpf, all of the PLM cells migrate to the intermediate cell mass (ICM) (yellow). At 24 hpf, after circulation begins, haematopoiesis moves temporarily to the posterior blood island (PBI) (purple) where the erythroid-myeloid progenitors (EMPs) emerge. The definitive wave begins in the aorta-gonad-mesonephros (AGM) (light blue) at 30 hpf, where haemogenic endothelial cell (HEC) transition occurs. By 3 days post fertilisation (dpf), haematopoietic stem cells (HSCs) migrate to the caudal haematopoietic tissue (CHT) (green) and thymus (grey) to differentiate. By 4 dpf, cells migrate to the kidney (red).

1.1.3.3 Definitive haematopoiesis in zebrafish

In the zebrafish, definitive haematopoiesis begins at approximately 30 hpf, which includes the generation of definitive haematopoietic stem cells (dHSCs) that maintain the blood system throughout life (Kissa et al., 2008). In the AGM region of the zebrafish embryo, this wave starts when haemogenic endothelial cells (HECs) in the ventral wall of the dorsal aorta (vDA) undergo a *runx1* transcription factor-dependent endothelial to mesenchymal transition as they differentiate into haematopoietic cells, through a process known as endothelial to haematopoietic transition (EHT) (Ciau-Uitz et al., 2014, Moore et al., 2018). During this transition, the endothelial cells change their shape from a spindle (long and flat) to a rounded form. These cells leave the lining of the vDA and enter the underlying mesenchyme (sub-aortic space) before migrating into the axial vein known as the posterior cardinal vein (PCV) (Kissa and Herbomel, 2010, Kissa et al., 2008). During this process, these cells are identified as HSCs and are able to differentiate into all the blood cell lineages, including the lymphoid lineage (Jagannathan-Bogdan and Zon, 2013). The process of dHSCs emerging from HECs (a common precursor of definitive haematopoietic cells) in the ventral wall of the dorsal aorta in AGM region is equivalent to the other vertebrates (Chen et al., 2011, Gordon-Keylock et al., 2013). Transcription factors *Runx1* and *c-myb* are associated with definitive haematopoiesis. The *runx1*⁺ cells sit in the vDA. It has been shown

that the *c-myb* expression is probably seen before the cells leave the vessel (Gering and Patient, 2005). In zebrafish, it has been demonstrated that in *runx1* morphant embryo, there is a complete absence of the definitive haematopoietic cells (Kalev-Zylinska et al., 2002). The *runx1* lies downstream of Notch signalling network. In zebrafish *mindbomb* (*mib*) mutant, a deficiency in the Notch signalling pathway leads to a defect in definitive haematopoiesis but do not affect primitive haematopoiesis (Itoh et al., 2003). In addition, in *runx1* morphant zebrafish embryo, it has been found that the loss of *runx1* expression causes a loss of *c-myb* expression in definitive progenitors (Gering and Patient, 2005). Studies have also demonstrated that the expression of both *runx1* and *c-myb* is an important regulator of HSC formation in the AGM region of other vertebrates as well (Bertrand et al., 2010, North et al., 2009). By 32-48 hpf, the first HSC progenitors join the circulation via the dorsal wall of the cardinal vein (CV) in the zebrafish. Through the circulation, they reach the tail mesenchyme, which then becomes the larval definitive haematopoietic site, the caudal haematopoietic tissue (CHT). The function of this tissue is likely to be similar to the FL in mammalian embryos (Murayama et al., 2006), it is the transient site for definitive HSC development, where HSCs divide and give rise to progenitors that then differentiate (Tamplin et al., 2015). At 3-4 days post fertilisation (dpf), the HSCs join the circulation via the tail veins from the CHT and begin to seed the thymus and kidney.

At 3 dpf, T cell progenitors first appear in the thymus, where they subsequently differentiate into T-lymphocytes (Jagannathan-Bogdan and Zon, 2013). The expression of *rag1* (recombination activating gene) can be detected in T lymphocytes within the developing thymus at the beginning of 4 dpf via whole-mount in situ hybridisation (Willett et al., 1997). By 4 dpf, the HSCs migrate to the KM (the marrow of the head kidney). In zebrafish, the kidney is responsible for maintaining adult HSCs and is analogous to the mammalian bone marrow (Kulkeaw and Sugiyama, 2012, Medvinsky et al., 2011). Flow cytometry analysis of whole kidney marrow (WKM) revealed the observation of several haematopoietic lineages; erythroid, lymphoid, precursors and myeloid, can be resolved by light-scatter characteristics (Traver et al., 2003b). Transplantation of adult *gata*^{GFP} zebrafish WKM into *gata*^{-/-} mutant resulted in the embryo that lacks the lethal phenotype, thus also generating both lymphoid and myeloid cell types (Traver et al., 2003b). These confirm the critical role of the zebrafish WKM as a main site of hematopoiesis in adult zebrafish. Light and electron microscopy have confirmed the kidney as the primary site of granulopoiesis in the adult zebrafish due to the presence of neutrophil granulocytes and their progenitors' cells (Lieschke et al., 2001). In mammals, neutrophils are short-lived in the blood, and so, they are continuously being replenished from progenitors in the bone marrow till they reach maturation at different stages of differentiation (Grassi

et al., 2018). In humans, neutrophils with a regular production rate of $\sim 10^{11}$ (70% of all leukocytes cells) per day are the most abundant type of leukocytes, and they play a significant role in the innate immune system (Figure 1.4) (Strydom and Rankin, 2013). In zebrafish, more than 90% of blood granulocytes are neutrophils (Traver et al., 2003a). The maturation of neutrophil precursors in the zebrafish kidney shares many of the characteristics of neutrophil development in human bone marrow (Bennett et al., 2001). The development of neutrophils and their lobulated nuclei in zebrafish is also conserved among vertebrates (Gaines et al., 2008, Hoffmann et al., 2007, Lieschke et al., 2001, Lian et al., 2018).

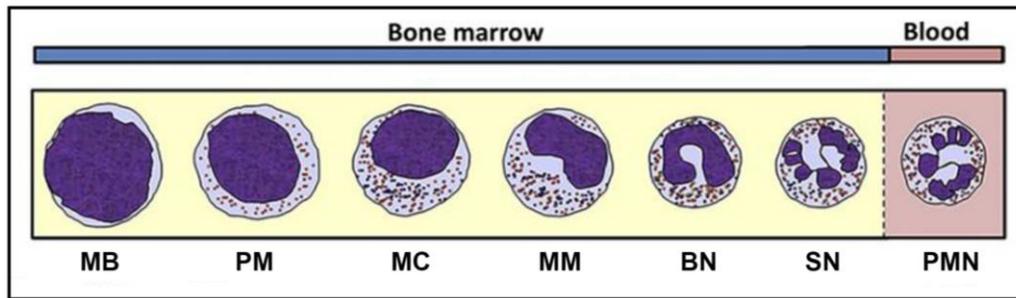


Figure 1.4 Schematic illustration of human neutrophils.

Neutrophils are generated in the bone marrow, and after several stages of differentiation, they can reach the bloodstream. Myeloblasts (MB) can proliferate, followed by maturation, then gradually moving through promyelocytes (PM), myelocytes (MC), metamyelocytes (MM), band neutrophils (BN), and finally segmented neutrophils (SN). Segmented neutrophils are released into the peripheral bloodstream as mature polymorphonuclear neutrophils (PMNs) (Image from (Cowland and Borregaard, 2016)).

1.2 Non-coding *cis*-regulatory elements

During embryonic development, the expression of a gene is controlled accurately, both spatially and temporally. This control occurs via the interaction of particular transcription factors (TFs) with *cis*-regulatory elements (CREs). The main classes of non-coding *cis*-regulatory elements in the genome include promoters, insulators, enhancers and silencers. (Nelson and Wardle, 2013, Raab and Kamakaka, 2010, Latchman, 2015).

1.2.1 Promoters

All eukaryotic genes have promoters that are approximately 100-1000 base pairs long, primarily located at 5' of a gene's coding sequence and crucial for gene regulation (Wright, 2019). One of the most important functions of the promoter is its interaction with the transcription initiation machinery. Promoters are usually located either upstream of the transcription start site (TSS) or overlapping with it. The TSS is the location where gene transcription is initiated. In order for transcription to be initiated, specific sequence elements are required to be bound by RNA polymerase II. In many eukaryotic genes, the TATA box is the most crucial element of polymerase II promoters and is located at -35 to -20 in front of the initiator (INR) that contains the TSS. The consensus sequence of the TATA box is TATAAA. The TATA box is recognised and bound by TBP (TATA-binding protein) and other transcription factors. Once they bind into the TATA box, the recruitment of RNA polymerase II and together

they orchestrate the initiation of gene transcription (Albert, 2015, Lenhard et al., 2012). TATA box and INR are the core promoter elements. Other elements such as the CAAT box and GC box, also referred to as promoter-proximal elements, are located within 200 bp upstream from the transcriptional start site. They are actively involved in the transcriptional regulation of various genes. CAAT box is a region with a consensus sequence of GGCCAATCT, which binds CAAT-binding protein (CBP). GC box is a region with a consensus sequence of GGGCGG recognised by Sp1s (a member of a large family of zinc-fingers TFs) which enhances the assembly of the transcriptional machinery (Dolfini et al., 2009, Godbey, 2014). CpG (cytosine followed by guanine and linked together by phosphate) islands (CpGIs) are well-known regions of genomic regulation inside and outside promoters. Majority of the promoters of the housekeeping genes which are constantly being expressed in the mammalian genome co-localise with CpGIs around TSS. CpGI regions remain unmethylated in active genes. The methylation (addition of a methyl group -CH₃ group to a nucleotide base) is a genomic modification that acts to inactivate gene transcription. CpGIs are regularly marked by tri-methylation of lysine 4 on histone 3 (H3K4me3), a signature of active promoters. The appearance of high levels of H3K4me3, acetylation of histone 3 at lysine 27 (H3K27ac), and the small numbers of mono-methylation of lysine 4

on histone 3 (H3K4me1) are also particular distinctive of promoters (Ernst et al., 2011).

1.2.2 Enhancers

An enhancer is a short region of DNA that is typically 50-1500 bp in length (Wright, 2019). An enhancer usually contains several binding sites for transcription factors (TFs) that range between of 6 to 12 bp in size. Enhancers can be bound by multiple transcription factors TFs aiming to enhance the transcription of a particular gene (Spitz and Furlong, 2012). When the TFs binds to the enhancer, general transcription factors (mediators), as well as chromatin remodelers, are recruited, facilitating enhancer-promoter interaction, which ultimately activates the gene transcription (Visel et al., 2009) (Figure 1.5). Enhancers are commonly located either upstream or downstream of a gene in the intergenic regions or in the introns of the genes. Enhancers are capable of regulating transcription in an orientation and distance independent manner (Ong and Corces, 2011). They can be located very far distance from the targeted gene. For example, a highly conserved limb-specific enhancer is located within the intron of the neighbouring gene *Lmbr1* and 1Mb away from its target Sonic Hedgehog (*Shh*) gene (Lettice et al., 2003). A typical gene can have multiple enhancers that often work independently and bind to particular TFs allowing the gene to be expressed in different cell types. For example, at least six enhancers are associated with the human apolipoprotein E (*APOE*) gene that

controls other tissues (i.e kidney, liver, skin and brain) of *APOE* expression (reviewed in (Visel et al., 2007)). Certain genes can also be regulated by several enhancers that drive similar or overlapping activity patterns named redundant enhancers. For instance, redundant enhancers regulate the *Gli3* gene expression in limb on mice (Osterwalder et al., 2018).

Over the years, progress has been made on enhancer study as a result of advances in sequencing technology and high-throughput approaches that have helped to find and validate candidate enhancers (Ryan and Farley, 2020). Genome-wide based studies are used to detect putative enhancers. Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq). ChIP-seq uses antibodies targeting a specific TF to ascertain the location of its binding sites within enhancers (Johnson et al., 2007, Robertson et al., 2007). ChIP-seq with particular antibodies against some histone modifications (i.e., H3K4me1 and H3K27ac) can be used to identify active enhancers (Bonn et al., 2012, Rada-Iglesias et al., 2011). Active enhancers can be found within open or accessible chromatin that the nucleosomal DNA is unwrapping. DNase I digestion can be used to detect the accessible chromatin followed by high-throughput sequencing (DNase-seq) (Boyle et al., 2008, Thurman et al., 2012). Open chromatin can also be identified by assay for transposase-accessible chromatin using sequencing (ATAC-seq), which inserts sequencing adapters into

accessible DNA regions using Tn5 transposase (Buenrostro et al., 2013).

Enhancers can act in far away distances when they are brought close to the promoter by the formation of a chromatin loop to regulate a particular target gene. The physical enhancer-promoter interactions have been shown in vivo using the chromatin conformation capture (3C) technique and its high throughput derivatives assays 4C (circularised 3C), 5C (carbon-copy 3C) and Hi-C (reviewed in (Ryan and Farley, 2020)). The human and mice genomic analysis of the genomes using Hi-C showed that it is widely organised in megabase-size sections named topologically associating domain (TAD) (Dixon et al., 2012). Within the same TAD, regulatory elements and genes especially interact, suggesting that the boundaries (CTCF sites) of TADs can work to restrict the influence of enhancers (Dixon et al., 2012). In mammals, TAD boundaries are largely conserved across different cell types and even between species. In mouse embryonic stem cells, approximately 2,200 TADs are organised within more than 90% of the genome with a median size of ~880 kb (Dixon et al., 2012). In the zebrafish genome, approximately 1,700 TADs have been identified with a median size of ~500 kb (Kaaij et al., 2018).

Deletion or change in the enhancer sequence may cause effects in gene regulation. For example, mutations in the enhancer regulate the expression of the sonic hedgehog (*SHH*) gene, resulting in

different limb skeletal defects in mice (Lettice et al., 2008). In zebrafish, *gata2a* expression is driven by the i4 (intron four) enhancer, which is important to program haematopoietic endothelium (HE) in embryos. Homozygous *gata2a*^{Δi4/Δi4} larvae at 5 dpf revealed normal numbers of haematopoietic stem and progenitor cells (HSPCs) (Dobrzycki et al., 2020). The reduction of *gata2b* and *runx1* expressions after the deletion of the enhancer was recovered by the activities of Notch signalling, which drives the expression of both genes in HE. However, the homozygous *gata2a*^{Δi4/Δi4} adults showed phenotypic abnormalities such as heart-swelling, skin infection and kidney marrow failure (Dobrzycki et al., 2020).

1.2.3 Insulator

In eukaryotes, chromatin is classified into accessible and inaccessible regions, called euchromatin (actively transcribed regions) and heterochromatin (silenced parts), respectively. Both accessible, as well as inaccessible regions are involved in gene regulation. In general, the insulator is a *cis*-regulatory element that targets one gene, preventing it from interfering with a neighbouring gene. They act to prevent enhancers or silencers from interacting with the promoter (Figure 1.5). The insulator is typically 300-2000 bp in length and has been categorised into two types; insulator barriers and enhancer blockers. Insulator barriers protect active genes from the spread of heterochromatic chromatin flanking them.

They, thus, keep the chromatin open. Enhancer-blocking insulators prevent the interaction between an enhancer and the promoters of neighbouring genes that lie beyond the insulator on the same chromosome. In vertebrates, the majority of insulators are bound by CTCF (CCCTC-binding factor), which is thought to be the primary part of the activity of insulators (Allison, 2007, Krivega and Dean, 2012, Raab and Kamakaka, 2010).

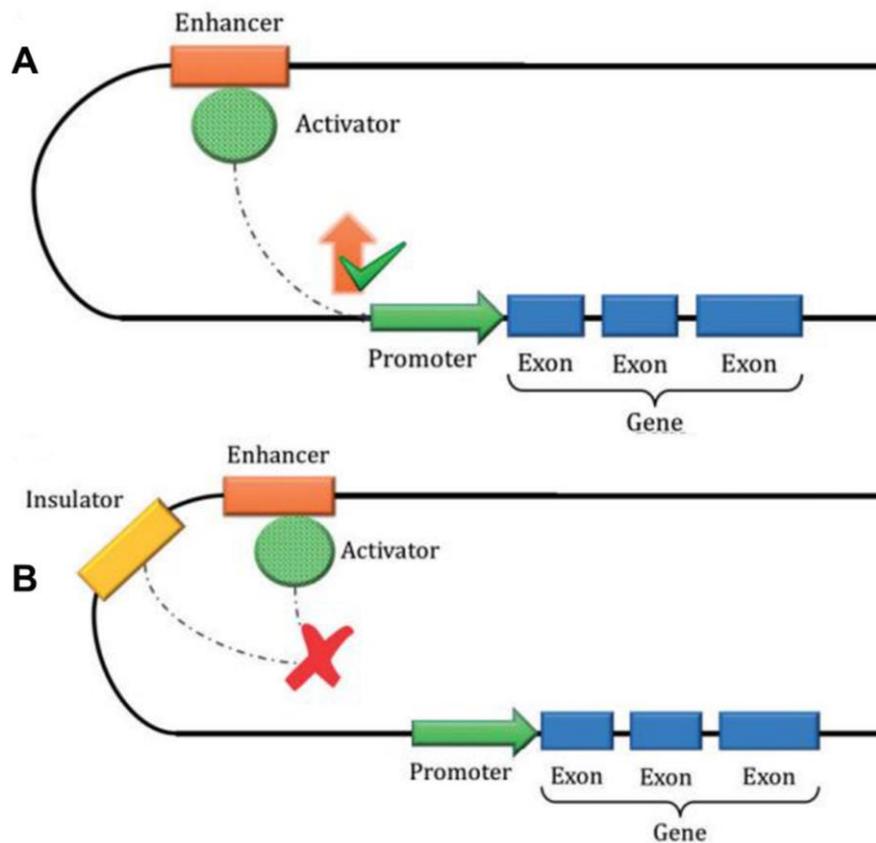


Figure 1.5 Schematic representation of the effects of *cis*-regulatory elements: promoters, enhancers and insulators

(A) The enhancer binds to an activator that joins to transcription factor (TF) in the promoter leading to upregulation of the target gene. **(B)** The insulator interacts with the activator of an enhancer to block its binding to the promoter, which in turn inhibits gene expression (Rojano et al., 2019).

1.3 Conserved non-coding elements (CNEs)

Genome sequencing of various species has revealed that several regions of non-coding DNA sequence have remained conserved over millions of years of evolution. Multi-species alignment of genomic sequences allows detection of these elements which have been called conserved non-coding elements (CNEs) (Nelson and Wardle, 2013). CNE can be classified based on the degree of conservation between two or more organisms, conserved non-coding elements (60-70% sequence identity) (Li et al., 2010), highly conserved non-coding elements (HCNE) (70-98%) (Woolfe et al., 2004) and ultra-conserved elements (UCNE) (95-100%) (Bejerano et al., 2004).

It has been reported that two *cis*-regulatory elements at the sequence level are conserved between (i.e., earthworms and insects) and deuterostomes (invertebrates and vertebrates) (Clarke et al., 2012). In addition, a series of TF binding sites can be identified in that conserved sequence (Clarke et al., 2012). CNEs can act as *cis*-regulatory CREs which control the expression of genes during the normal development of the embryo. Remarkably, CNEs are present in large numbers in the region around the genes which are responsible for encoding transcriptional regulators or factors that participate in embryonic development. CNEs also indicate that the conservation of these sequences is due to their consistent gene regulator function (Li et al., 2010). Many CNEs function as

developmental enhancers, and any alteration of this elements sequence may lead to phenotypic changes.

1.4 c-Myb and its importance during haematopoiesis

The c-Myb protein is a transcription factor (TF) that is encoded by the *c-Myb* gene. It regulates the expression of many genes and helps to coordinate the proliferation and differentiation of cells in different normal cell types (Ramsay, 2005). It is one of the most essential regulators of haematopoiesis and is highly conserved in vertebrates. c-Myb is expressed in stem and progenitor cells of all hematopoietic lineages, including erythroid, myeloid, and lymphoid cells (Lieu and Reddy, 2009, Sumner et al., 2000, Thomas et al., 2005, Vegiopoulos et al., 2006). c-Myb is a member of the Myb (myeloblastosis) TF family in vertebrates, first identified as proto-oncogenes involved in the development of avian leukaemia (Baluda and Reddy, 1994). c-Myb is a 75-kDa protein that consists of three functional domains; an amino-terminal DNA-binding domain (DBD), a central transactivation domain (TAD) and a negative regulatory domain (NRD). The DBD binds to the DNA $T/C AAC^G/TG$ sequence. The c-Myb interacts with proteins CBP and p300 to activate transcription. These proteins provide a link between c-Myb and the transcription machinery (Greig et al., 2008). There are three members of the Myb transcription factor family A-Myb, B-Myb and c-Myb. Knockout studies of genes that encode these TFs show that they play different functions during their expression in the development process. The

A-Myb (encoded by *MYBL1* gene) is required during spermatogenesis. B-Myb (encoded by *MYBL2* gene) plays an important role in all dividing cells (Mucenski et al., 1991, Tanaka et al., 1999, Toscani et al., 1997).

The *c-Myb* gene is highly expressed in immature proliferating haematopoietic cells and is down-regulated as soon as these cells differentiate, suggesting that *c-Myb* is essential during the transition between proliferation and differentiation (Gonda and Metcalf, 1984). Transcriptional regulation of the *c-Myb* gene typically begins at the transcriptional start site (TSS) of the promoter upstream of the first exon. The *c-MYB'* upstream promoter sequence has been found to be GC rich and lacks both TATA box and CAAT box. Many Sp1 (specificity protein 1) factor binding sites (GC box) were also found both upstream and downstream of the TSS. The *c-Myb* transcription is initiated by the Sp1 binding factor (Dvořák et al., 1989). The Sp1 factor activates the RNA polymerase II by binding to promoter sequences to initiate the transcription of the *c-MYB* gene. It has been found that the GC-rich regions (2 CpG islands) noticed in the promoter-proximal upstream sequence play an important role in *c-myb* gene transcriptional activation (Hsu, 2010). This *c-myb* promoter fragment can direct the expression of green fluorescent protein (GFP) in several tissues in the zebrafish transgenic line (Hsu, 2010).

The regulation of c-Myb can also involve other transcription factors. Within the TSS, there are two binding sites, E2 Factor (E2f) and Nuclear Factor kappa B (NF- κ B), which are activators of c-MYB expression in T-cells. Deleting conserved E2F and NF- κ B binding sites results in the elimination of MYB transcription in these cells (Lauder et al., 2001). The expression of c-MYB is also regulated by an attenuator site (stem-loop) within the intron one sequence that controls transcriptional elongation. Mutations in the MYB intron one attenuator region were found responsible for the overexpression of c-MYB protein in colorectal cancer cells (Hugo et al., 2006). The c-Myb gene is also regulated by distant enhancers. It has been found that the distal enhancer (-28kb) controls the regulation of the *c-Myb* gene expression during the differentiation of myeloid cells in mice (Zhang et al., 2016). During mouse erythroid differentiation, enhancers within the *Myb-Hbs1l* intergenic region interact with the *c-MYB* promoter and the first intron by forming an active chromatin hub (ACH) (Stadhouders et al., 2012). Upon erythroid differentiation, the ACH is lost, and the expression of the *c-Myb* is downregulated (Stadhouders et al., 2012). Genomic sequence comparisons between vertebrate species identified conserved non-coding element one (CNE1) in their *c-myb* loci that regulate the *c-myb* gene expression (Hsu, 2010). Several putative TF binding sites identified in CNE1, such as Ets, Gata, Runx1, and c-Myb, play an essential role during the development of haematopoietic cells (Hsu,

2010). During haematopoiesis, the expression level of *c-myb* may be influenced by specific microRNAs (miRNAs), which act in post-transcriptional regulation. Interactions of miR-126 with the 3'-UTR of *c-myb* act to down-regulate *c-myb* post-transcriptionally. A knockdown of miR26 in zebrafish embryos revealed that there was an upregulation in erythropoiesis and a downregulation in thrombopoiesis as a result of the high levels of c-Myb protein (Grabher et al., 2011).

c-MYB is part of a complex genetic network that consists of several proteins which are involved in the c-MYB transcriptional factor network (Figure 1.6). The start of haematopoietic activities is coordinated by major transcription factors that regulate the expression of the genes during haematopoietic activities (Jagannathan-Bogdan and Zon, 2013). In haematopoiesis, the c-Myb protein work with various transcription factors by binding to promoters or enhancers of a number of genes to regulate their expression (Ness, 1999). For instance, *mim-1* gene's promoter is activated in a combination of c-Myb and C/EBPB for the regulation in the differentiation of immature myeloid cells (Ness et al., 1993). The promoter region of *Rag-2* gene has binding sites for c-Myb and Pax-5 that regulate the activities of *RAG-2* in immature B cells (Kishi et al., 2002). The c-Myb TF can also act as either a transcriptional repressor or activator at the CD4 silencer and promoter, respectively, in T-cell development (Allen III et al., 2001). c-Myb

and HES-1 (a known transcriptional repressor) form a multifactor interaction complex during the transcription process to mediate the function of the CD4 silencer (Allen III et al., 2001). The TFs *c-myb* and *runx1* are essential for granule-related genes (*lyz*, *mpx*, *npsn* and *srgn*) expression in neutrophil maturation in zebrafish embryonic myelopoiesis. These TFs bind to these genes' promoters and interact to cooperatively regulate their expressions (Huang et al., 2021). During neutrophil differentiation, *c-myb* and *cebp1* work in parallel and synergetically by regulating granular protein gene transcription in zebrafish embryos (Jin et al., 2016).

In the mouse embryo, although *c-Myb* expression was found at a low level in primitive haematopoiesis in the YS at E7.5 (Palis et al., 1999), *c-Myb* plays an essential role in definitive haematopoiesis. It is expressed in the mouse foetal liver, a site of definitive wave (Sitzmann et al., 1995). Upon deletion of *c-Myb*, mouse embryos appear normal until E13.0, indicating that it is dispensable for primitive haematopoiesis. By E15.0, mice suffer from severe anaemia, and homozygous mice do not survive beyond day 21 (Mucenski et al., 1991).

During zebrafish embryogenesis, *c-myb* is expressed at a different time point in different haematopoietic and non-haematopoietic tissues. The expression is initiated in the ALM and the PLM at 4 to 5-somite stage. The expression of the *c-myb* gene is normally detected around 10 to 12-somite stage in the ALM and the PLM. At 12-somite

stage, *c-myb* expression is detected in the primitive *pu.1*⁺ myeloid progenitors in the ALM. At 18 hpf, the expression of *c-myb* is observed in erythrocytes of the ICM. Around 26 hpf, *c-myb* is expressed in *runx1*-expressing cells in the vDA. At about 2-3 dpf, cells expressing *c-myb* reaches the tail region to colonise the CHT. At 4-5 dpf, *c-myb*⁺ cells migrate to seed the thymus and the kidney, which is the adult site of haematopoiesis in zebrafish (Murayama et al., 2006). During zebrafish embryonic development, between 1-5 dpf, *c-myb* is also expressed in several non-haematopoietic tissues such as retina, midbrain, olfactory epithelium and pharyngeal arches (Hsu, 2010) (Figure 1.7). In zebrafish, homozygous *c-myb* mutants produce normal primitive erythrocytes suggesting that *c-myb* is not essential for early erythropoiesis in the zebrafish either. By approximately 20 dpf, the number of blood cells was reduced, and no erythrocytes were detected in the blood vessels. Although *c-myb* homozygous null zebrafish display severe anaemia, they are able to survive for 2-3 months. These mutant fish are capable of surviving for that long, presumably due to their ability to perform gas exchange by diffusion through the skin (Soza-Ried et al., 2010). It has been observed that the number of lymphocyte progenitors is reduced, and the thymus is not colonised in *c-myb*-null embryos (Hess et al., 2013). In *c-myb* mutant embryos, in the RBI, it has been found that whilst primitive myelopoiesis appeared to be independent of *c-myb* as shown by the normal expression of *pu.1*,

primitive neutrophils but not macrophages are dependent on its expression. Nevertheless, early neutrophil markers, for instance, *ceb1* are unaffected in *c-myb*^{-/-} embryos. By contrast, the expression of some neutrophil markers like *mpx*, *srgn* and *lyz* is reduced or lost in the zebrafish mutant embryos (Jin et al., 2016).

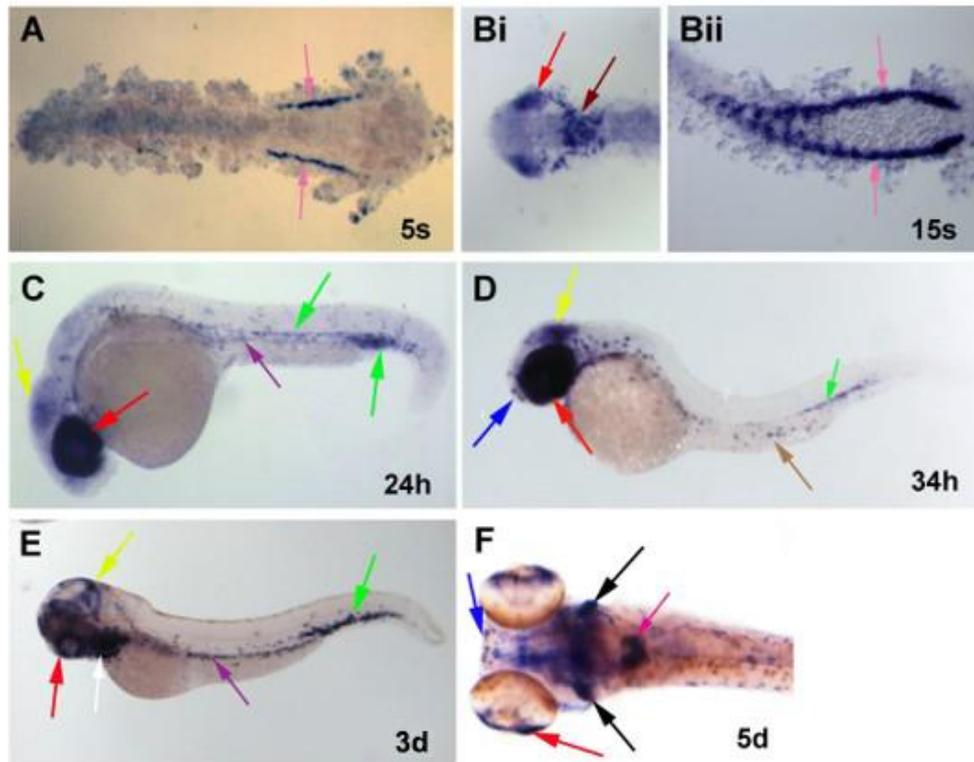


Figure 1.7 The expression pattern of *c-myb* during zebrafish embryogenesis.

(A) At the 5-somites stage, *c-myb* expression can be observed (pink arrow) in the posterior lateral mesoderm (PLM) during primitive haematopoiesis. **(Bi)** At 10-15 somites *c-myb* expression is detectable in the myeloid progenitors (maroon arrow), and in the retina (red arrow). **(Bii)** *c-myb* expression is still observed in the PLM cells as they migrate towards the midline at 15-somites stage. **(C)** *c-myb* is expressed in the posterior blood island (PBI) (right green arrow), vDA (left green arrow), pronephric duct (purple arrow), midbrain (yellow arrow) and retina at 24-26 hpf. **(D)** At 34 hpf, *c-myb* is also expressed in the olfactory system (blue arrow), retina, midbrain, surface skin cells (brown arrow), and the PBI (green arrow). **(E)** *c-myb* expression is seen in the pharyngeal arches (white arrow), CHT, retina, midbrain and pronephric duct at 3 dpf. **(F)** At 5 dpf, *c-myb* expression can be detected in the olfactory system, kidney (pink arrow), retina and thymus (black arrows) (Hsu, 2010).

1.5 Identification and testing of *c-myb* promoter-proximal regulatory sequences

In order to examine the transcriptional regulation of *c-myb* in the zebrafish embryo, a previous PhD student, Jui-Cheng Hsu tested both promoter-proximal and distal regulatory elements of the zebrafish *c-myb* gene in transgenic zebrafish lines by generating EGFP reporter transgenes (Hsu, 2010) (Figure 1.8A). The first promoter fragments (PFs) tested were called promoter fragments PF1 and PF2. They harboured upstream sequences of ~10 and ~5.3 kb, respectively. In the transgene constructs, the EGFP reporter gene was inserted into exon 1 of *c-myb* downstream the *c-myb* ATG start codon. In stable transgenic zebrafish generated with the transgene, low levels of EGFP expression were found in the retina, olfactory placode, pronephric duct and tubules in transgenic zebrafish lines and significant ectopic expression was observed in some tissues, for example, in the spinal cord (Figure 1.8B).

A further four promoter fragments (PF3-PF6) were tested. They differed from PF1 and PF2 in that they contained the complete exon 1 and intron 1 sequences and had the GFP reading frame inserted in exon 2. The transgenic reporter lines carrying the PF3, PF4, PF5 and PF6 promoter fragments showed variable levels of EGFP expression in the non-haematopoietic tissues like the retina, the olfactory placodes and the branchial arches (Figure 1.8C). The levels of

reporter expression were higher in the PF3, PF4 and PF5 lines and lower in the PF6 line. The expression of the reporter gene was more robust in these lines than in the lines generated with PF1 and PF2, suggesting that intron 1 of *c-myb* plays an important role in the transcriptional regulation of zebrafish *c-myb*. The PF3-PF6 lines also showed less ectopic expression and tighter control over the promoter activity. Disappointingly, none of the transgenic lines that carried different promoter fragments displayed any EGFP expression in haematopoietic cells. This suggested that the CREs directing *c-myb* expression to haematopoietic sites were neither located immediately upstream of *c-myb* nor in intron 1 of *c-myb*. Elements in other introns or downstream or distal enhancer elements had to be responsible for *c-myb* expression in the haematopoietic cells.

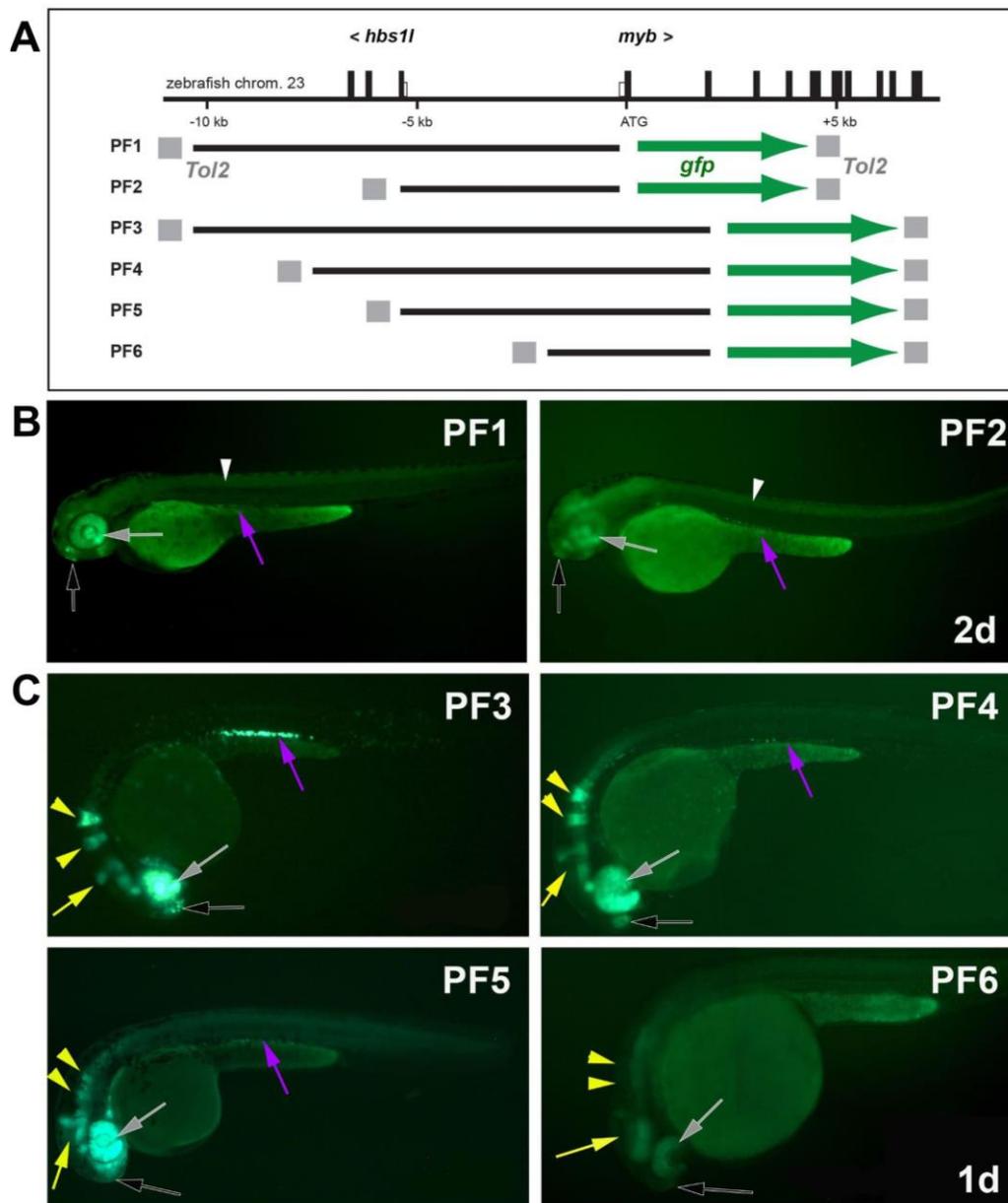


Figure 1.8 PF1-PF6 promoter-proximal elements are able to direct the EGFP expression of *c-myb* gene in non-haematopoietic tissues

(A) A schematic representation of all the *c-myb* promoter fragments (PF1-PF6) cloned to generate transgenic lines. Black vertical lines indicate *c-myb* exons. The green arrows show the green fluorescent protein GFP reporters. The grey boxes show the Tol2 sequence. Horizontal black lines represent the promoter sequences. **(B)** Images depict Tg(*c-myb*-PF1:*egfp*) and Tg(*c-myb*-PF2:*egfp*) embryos showing weak reporter gene expression in the retina (Re) (grey arrow), olfactory placode (OP) (black arrow), spinal cord (white arrowhead) and pronephric duct (PND) (purple arrow) at 2 dpf. **(C)** Images show Tg(*c-myb*-PF3:*egfp*)-Tg(*c-myb*-PF6:*egfp*) embryos, eGFP is detected in the Re, OP, PND, the brain (BR) (yellow arrow) and rhombomeres (yellow arrowheads) at 1 dpf. The figure has been kindly provided by Dr. Martin Gering.

1.6 Testing of potential *c-myb* distal regulatory elements in zebrafish transgenic lines identified CNE1 as a conserved haematopoietic enhancer

As the GFP expression was not detected in haematopoietic tissues in any of the previous transgenic zebrafish lines, it was assumed that the promoter itself and intron 1 were insufficient to drive gene expression in haematopoietic cells and that a distal regulatory element controlled haematopoietic *c-myb* expression. Analysis of the sequence surrounding the *c-myb* locus was performed to uncover any potential regulatory elements (REs). Conserved sequences were identified in the downstream *c-myb* gene among eight species, including species of fish to human, using the M-LAGAN software (Multi-Limited Area Global Alignment of Nucleotides). M-LAGAN is an applied method for creating alignments of many multiple genomic sequences at any evolutionary distance (Brudno et al., 2003). Firstly, two of the conserved non-coding elements (CNEs) were found downstream of the *c-myb* gene in all fish species (Figure 1.9A). Then the sequence alignment was extended to include vertebrates. Interestingly, sequence alignments that included sequences from all major vertebrate clades revealed that the zebrafish conserved non-coding element one (zCNE1) (formerly known as CNE8 (Hsu, 2010)) is conserved in all species from fish to man (Figure 1.9B). Subsequently, the DNA sequence of the *c-myb* locus was examined for the presence of CpG islands. For this

purpose, the identification of CpG islands was performed as a region of 200 bp windows with a GC content of over 49% and the calculated an 'observed/expected' ratio of more than 0.6 using the web- based CpG island Searcher.

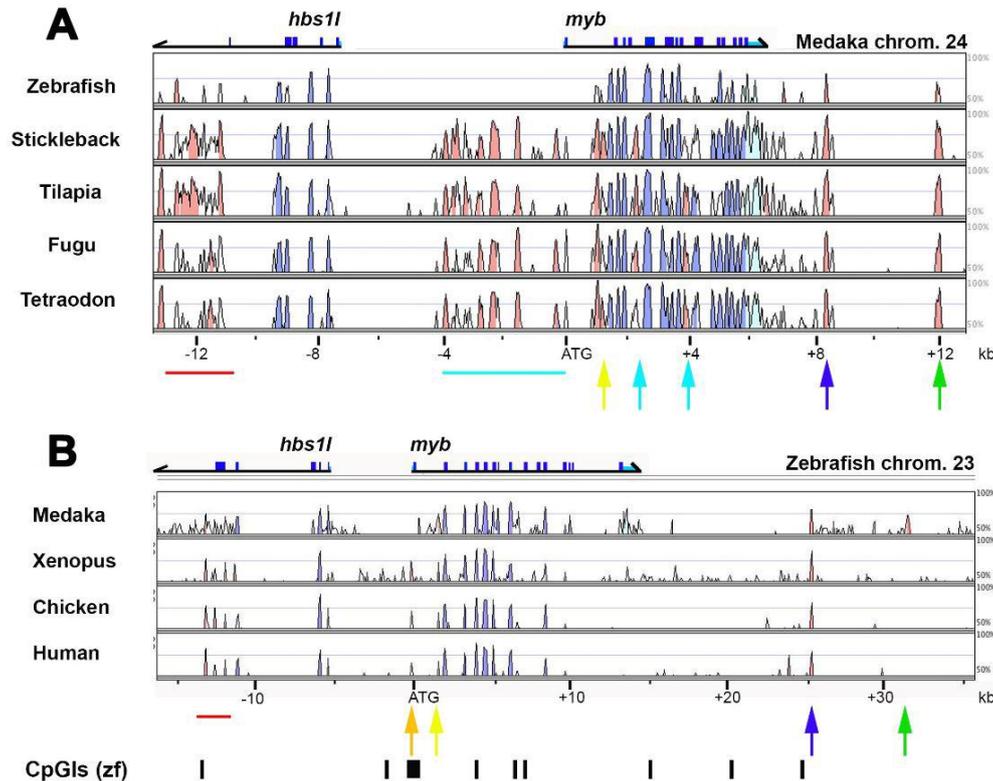


Figure 1.9 Genomic sequence comparisons between vertebrate species identified a number of conserved non-coding elements (CNEs) in the *c-myb* locus

In the above of **(A and B)** maps of the medaka and zebrafish *c-myb* locus, respectively. The horizontal lines show the *c-myb* locus on medaka and zebrafish chromosomes 24 and 23, respectively. The vertical blue lines represent exons of the *c-myb* and *hbs1* genes. **(A)** Vista blots highlighting regions of sequence identity that were found in pairwise sequence alignments of the medaka *c-myb* sequence with that of zebrafish, stickleback, tilapia, fugu and tetraodon. **(B)** Vista blots highlighting regions of sequence identity that were found in pairwise sequence alignments of the zebrafish *c-myb* sequence with that of medaka, xenopus, chick and human. The coloured peaks were higher than 70%. Blue peaks show conserved sequences in coding exons. Pink peaks show conserved non-coding elements. The numbering of nucleotides is done relative to the transcription start sites of medaka and zebrafish *c-myb*. The red line shows conserved non-coding elements of neighbouring gene *hbs1*. The light blue line shows conserved non-coding elements upstream of the *c-myb* gene promoter. The yellow arrow shows a non-coding element in intron 1. The light blue arrows show non-coding elements within introns. The dark blue arrow shows CNE1 that is conserved across species. The green arrow indicates CNE that is conserved in fish species only. The orange arrow shows the start codon ATG. Black boxes in the bottom are CpGIs in zebrafish. The figure has been kindly provided by Dr. Martin Gering.

Nine CpGIs were located in the region of the *c-myb* locus, with some being isolated and others occurring in clusters. One of the CpGIs, was located close to CNE1. Based on the finding of the sequence alignments and CpGIs analysis, DNA fragments that carry CNEs with or without CpGIs were defined as putative regulatory elements (REs) that were subsequently tested to see whether they possessed enhancer activity in haematopoietic cells. For this purpose, the putative REs were cloned upstream of the small PF6 *c-myb* promoter fragment in the EGFP reporter transgene construct. The constructs were used to generate stable zebrafish transgenic lines that were examined for EGFP expression (Figure 1.10). The zebrafish RE1 (zRE1) contained the zCNE1 and together with a neighbouring CpGI.

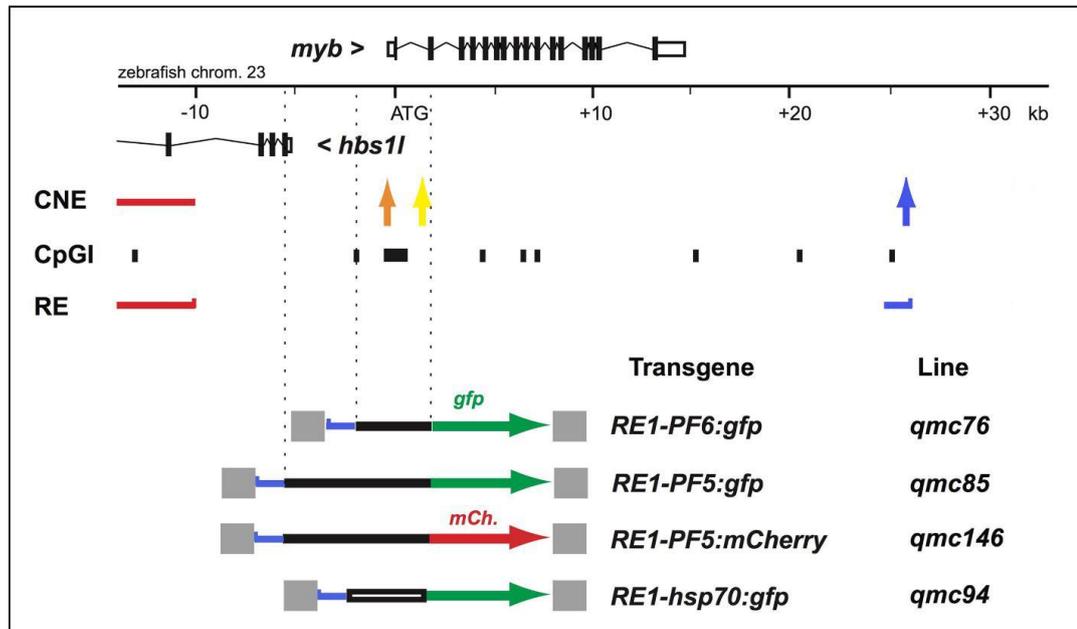


Figure 1.10 A map of the putative regulatory elements' fragments used to make the constructs to generate transgenic zebrafish lines

At the top, the vertical black lines represent exons of the *c-myb* locus. The *c-myb* promoter is indicated by the orange arrow, and the yellow arrow signifies intron 1. The blue arrow indicates zCNE1 that is conserved across species. The red line points to numerous conserved sequences that were tested in reporter constructs. Black boxes are CpGIs. The horizontal blue line indicates the putative regulatory element that was identified and tested for enhancer activity. At the bottom, the grey boxes show *To12* sequences. The blue lines represent the putative regulatory element. The promoter fragments (PF5, PF6 and hsp70) are shown as black lines. The figure has been kindly provided by Dr. Martin Gering.

The *Tg(c-myb-zRE1-PF6-egfp)^{qmc76}* transgenic line showed strong EGFP expression in the retina, olfactory placodes and pronephric duct. Interestingly, there was an expression in the definitive blood cells below the vDA and in the CHT, which was not found in the transgenic line that contained PF6 alone. The expression of EGFP in HSC progenitors in the vDA was observed at 30 hpf and peaked by 2 dpf. It declined by 3 dpf because of the migration of cells and seeding of the CHT (Hsu, 2010). EGFP expression in *qmc76* was also observed in the cells of the larval kidney and was maintained during adulthood. The cells analysed were haematopoietic cells (Figure 1.11Ai-Aiii) (Savage, 2012).

Because the PF6 gave weaker expression, a new construct was generated contained promoter fragment five (PF5), which would provide stronger reporter gene expression in blood cells than PF6 when combined with zRE1. This line was named *qmc85*. In the *Tg(c-myb-zRE1-PF5-egfp)^{qmc85}* transgenic line, zCNE1 was able to direct robust EGFP expression in blood cells in the CHT at 3 dpf (Figure 1.11Bi). In addition, FACS analysis on cells from the adult kidney revealed that approximately 22% of GFP+ cells were found in the kidney of *Tg(c-myb-zRE1-PF5-egfp)^{qmc85}*. Moreover, the EGFP expression was found in myeloid, lymphoid and progenitor cells (Figure 1.11Bii) (Savage, 2012).

To examine the ability of the zCNE1 to regulate different promoters and to verify what expression pattern could be observed, a transgenic line was generated that contained other promoters combined with zRE1. This line had the heterologous heat-shock 70 promoter (HSP70) which contained zRE1 and EGFP and was named *qmc94*. In the $Tg(c-myb-zRE1-hsp:egfp)^{qmc94}$ transgenic line, EGFP reporter gene expression was observed in the CHT definitive blood cells (Figure 1.11C). Some of the individual EGFP+ cells were observed in the circulation at 4 dpf seeded the kidney and the thymus via circulation (Tsikandelova, 2013). In addition, *qmc146* line was generated which contains the regulatory element zRE1 fused to a zebrafish PF5 and a *mCherry* reporter (Figure 1.11D).

To examine the evolutionary conservation of RE1 (that contains CNE1), the mouse orthologue (mRE1) (which contains mCNE1) was used in combination with zebrafish promoter PF5 to generate an EGFP transgenic line. It has been found that mRE1 is able to direct zebrafish *c-myb* promoter activity to definitive haematopoietic cells in the zebrafish transgenic line $Tg(c-myb-mRE1-PF5:egfp)^{qmc156}$ (data not shown) (Mohamed, 2015).

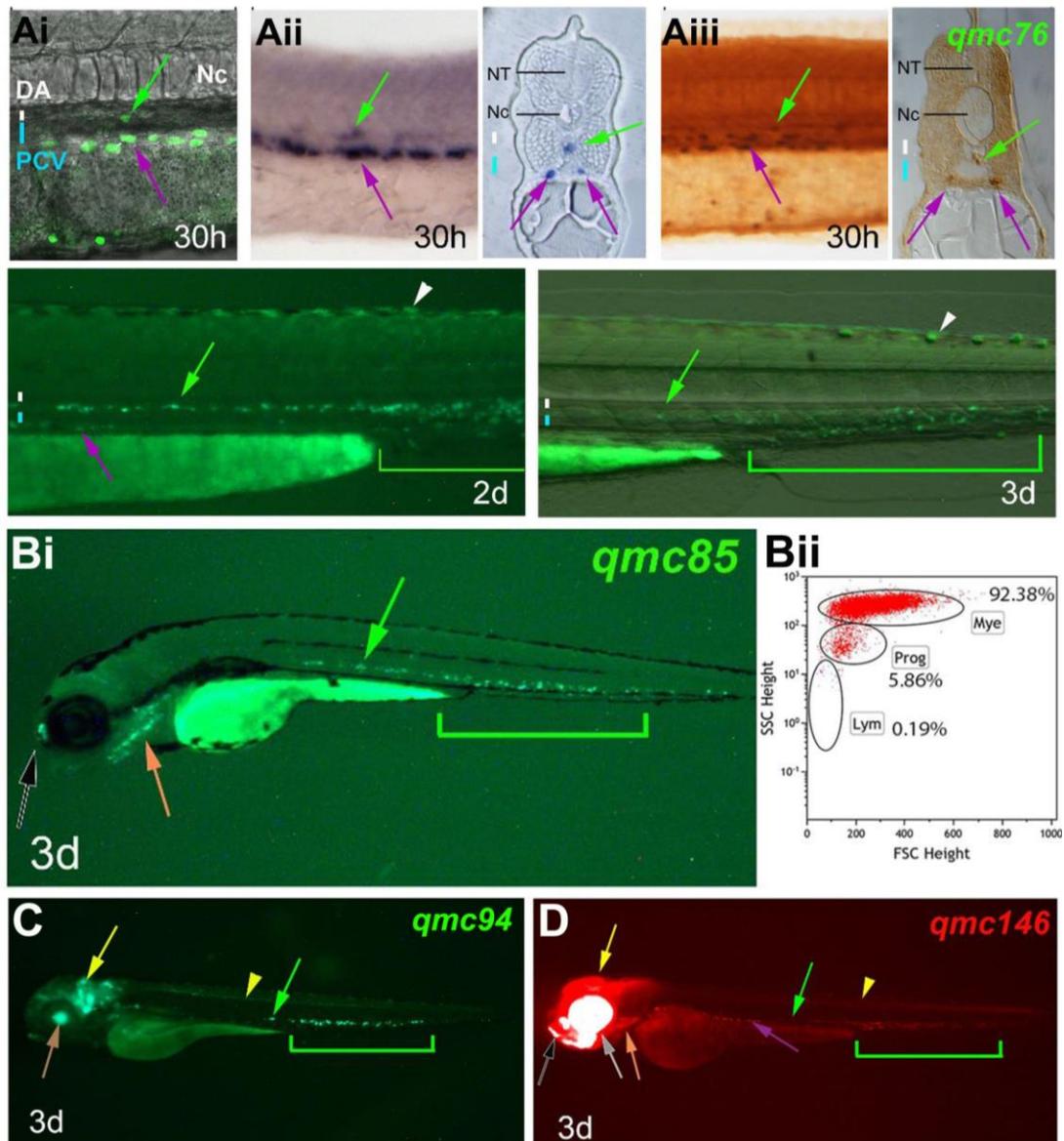


Figure 1.11 zRE1 and mRE1 are able to drive reporter genes expression in the different promoters to haematopoietic cells

(Ai) Confocal live image from a $Tg(c-myb-zRE1-PF6:egfp)^{qmc76}$ embryo at 30 hpf depicts *egfp* expression in the vDA (green arrow) and (pronephric duct) PND (purple arrow). **(Aii)** At 30 hpf, EGFP is expressed in the trunk of $Tg(c-myb-zRE1-PF6:egfp)^{qmc76}$ embryo as verified via whole-mount in situ (WISH). Transverse section from the WISH confirms that the expression of *gfp* can be detected in the under of vDA and the PND at 30 hpf. **(Aiii)** By immunohistochemistry, expression of *egfp* is observed vDA and the PND of the trunk of a $Tg(c-myb-zRE1-PF6:egfp)^{qmc76}$ embryo. A transverse section has also been confirmed. EGFP expression is obvious in vDA and the CHT (green bracket) between 2 dpf and 3 dpf. **(Bi)** The expression of EGFP can be observed in vDA, the CHT, olfactory placode (OP) (black arrow) and branchial arches (BA) (orange arrow) in a $Tg(c-myb-zRE1-PF5:egfp)^{qmc85}$ embryo at 3 dpf. **(Bii)** Flow cytometric analysis of whole marrow kidney cells from $Tg(c-myb-zRE1-PF5:egfp)^{qmc85}$ embryo. **(C)** EGFP expression can be seen in vDA, the CHT, lens (brown arrow)

and brain (BR) (yellow arrow) in a $Tg(c-myb-zRE1-hsp:egfp)^{qmc94}$ embryo at 3 dpf. **(D)** Expression of mCherry can be detected in PND, OP, vDA, the CHT, BA and retina (Re) (grey arrow) in a $Tg(c-myb-mRE1-PF5:mCherry)^{qmc146}$ embryo at 3 dpf. The images have been kindly provided by Dr. Martin Gering.

1.7 Identification of H3K4me1 and H3K4me3 sites in the zebrafish genome

In addition to zRE1, other enhancer sequences may also be present downstream of the zebrafish *c-myb* gene. A study published by Nathan Lawson's lab used a CHIP-Seq approach, an experiment in which chromatin immunoprecipitation was followed by deep sequencing, to determine the genome-wide binding profile of trimethylated (H3K4me3) and mono-methylated (H3K4me1) lysin 4 on histone 3 in the chromatin of whole zebrafish embryos at 24 hpf (Aday et al., 2011) (Figure 1.12). These marks had previously been shown to mark promoters and enhancers, respectively (Akkers et al., 2009, Barski et al., 2007). The Lawson data are available as a track on the Zv9 assembly of the zebrafish genome sequence on the UCSC genome browser (<https://genome-euro.ucsc.edu>). Close inspection of these data for information on regulatory elements present in the *c-myb* locus shows that there are several candidates for enhancer elements downstream of the *c-myb* gene. Table 1.1 A list of potential *c-myb* downstream enhancers (PEs) identified as H3K4me1 binding sites in chromatin isolated from 24 hpf zebrafish embryos in a CHIP-Seq experiment published by Aday et al., 2011 provides the exact coordinates of these potential enhancer elements (PES). Interestingly, one of the sequences enriched by H3K4me1 partially overlaps with zRE1.

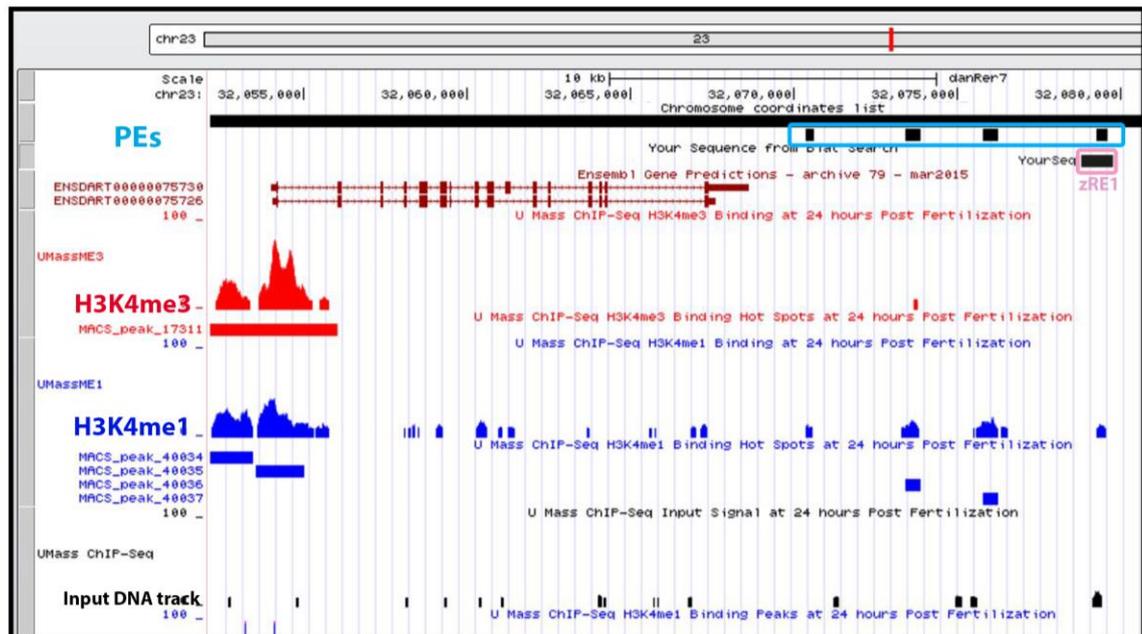


Figure 1.12 Identification of potential enhancers (PEs)

The panel shows an overview of the zebrafish *c-myb* gene along chromosome 23. The track on the UCSC browser displays data from a genome-wide analysis that investigated the presence of H3K4 mono and trimethylation throughout the zebrafish genome at 24 hpf (UCSC Genome Browser on Zebrafish Jul. 2010 (Zv9/danRer7)). The potential enhancers (PEs) of *c-myb* downstream are surrounded by a light blue frame. Black boxes in the top show the PEs, from left to right; MP40036-, MP40036, MP40037 and MP40037+, respectively. The zebrafish regulatory element one (zRE1) of *c-myb* downstream is surrounded by a light pink frame.

Table 1.1 A list of potential *c-myb* downstream enhancers (PEs) identified as H3K4me1 binding sites in chromatin isolated from 24 hpf zebrafish embryos in a CHIP-Seq experiment published by Aday et al., 2011

Two PEs, MP40036 and MP40037, were called by the authors' algorithm. Two others, MP40036- and MP40037+, can be deduced by interrogating the data track on the UCSC genome browser. The data track can be found on the Zv9/danRer7 version of the zebrafish genome assembly.

UCSC #elements	Genomic Location
MP40036-	chr23 32070369 32070599
MP40036	chr23 32073399 32073899
MP40037	chr23 32075776 32076247
MP40037+	chr23 32079264 32079585

1.8 CRISPR/Cas9 system as a tool for gene editing

CRISPR/Cas is a natural adaptive immune system used by prokaryotes to protect against invading nucleic acids, such as viruses and plasmids. Upon infection, the information of foreign invaders is saved in short DNA sequences called protospacers that are repeated and inserted within a CRISPR locus in the genome of bacteria. These short DNA sequences serve as a type of immune system memory. The protospacers are expressed as short guide RNAs that help Cas endonucleases to find a complementary sequence next to a protospacer adjacent motif (PAM). The invading DNA is cleaved by the Cas nuclease to prevent recurrent infections (Bhaya et al., 2011, Wiedenheft et al., 2012, Ran et al., 2013) (Figure 1.13).

The CRISPR/Cas system has been modified to allow genetic engineering in a variety of species. There are different types of CRISPR systems. Because of its efficiency, simplicity and multiplexing potentials, type II from *Streptococcus pyogenes* was used in this project as a genome-editing tool (Gupta and Musunuru, 2014, Shmakov et al., 2015). Two primary components are required for genome editing. They are CRISPR-associated Cas9 nuclease and a single guide RNA (gRNA). The gRNA contains two different segments of RNA: CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA contains a 18-20 bp sequence that is complementary to the DNA sequence of the target site for Cas9. The

tracrRNA acts as a scaffold for the crRNA-Cas9 interaction. It has been shown that crRNA and tracrRNA can be fused to a single guide RNA (sgRNA) that retains the activity of the two individual RNAs. Suitable target sites for Cas9 are 20 bp sequences that lie upstream of a protospacer adjacent motif (PAM) (NGG) (Gupta and Musunuru, 2014, Jinek et al., 2012, Cong et al., 2013). Guided to the target by the sgRNA, Cas9 introduces a double-strand break (DSB) 3 bps upstream of the PAM sequence. The DSB generated by Cas9 activates one of two repair mechanisms. One of them is non-homologous end joining (NHEJ), which generally produces insertions/deletions (indels) disrupting a target gene. The other one is homology-directed repair (HDR). This is a process whereby a DSB is repaired by homologous recombination using a DNA template (Sander and Joung, 2014, Chang et al., 2017). In zebrafish, several studies have successfully used the CRISPR/Cas9 technology to generate gene editing with high rates in several genomic loci (Chang et al., 2013, Irion et al., 2014, Jao et al., 2013).

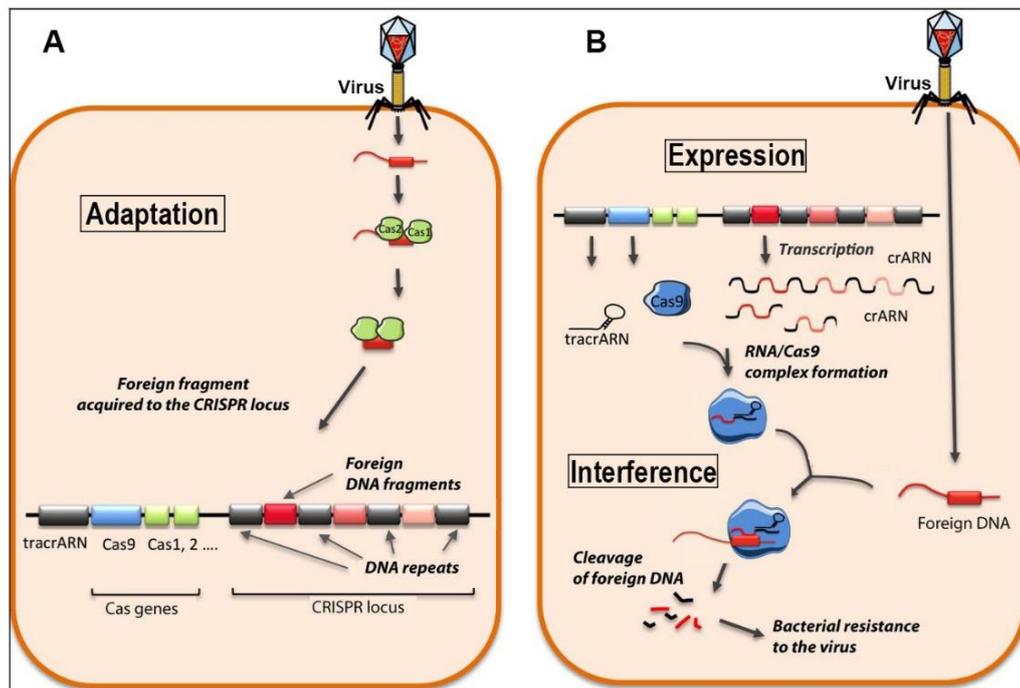


Figure 1.13 CRISPR-Cas adaptive immunity in bacteria

(A) Adaptation. During infection, the bacteria is invaded by foreign DNA. Numerous of Cas genes are expressed and bind proteins of the invading DNA (protospacers). Protospacers are incorporated into the spacer region of CRISPR locus and stored in the genome. **(B)** Expression and Interference. Upon subsequent infection, the CRISPR region is transcribed to produce individual spacer-repeat units of RNA, called CRISPR RNA (crRNA). Then, the Cas -nuclease binds the crRNA. Interference occurs when the Cas-crRNA complex recognises a protospacer sequence that is complementary to the spacer, and Cas-nuclease cleaves the invading DNA. The figure was obtained from (Duroux-Richard et al., 2017).

1.9 Background

It was previously reported that the majority of non-coding DNA is so-called 'junk' DNA. However, large-scale genome studies such as the ENCODE (Encyclopaedia of DNA Elements Project) has revealed the functional role of some of these non-coding DNA (Consortium, 2007, Doolittle, 2013). These previously believed non-functional DNA regions have important roles in transcription and gene regulation. Some of these non-coding DNA (elements) are evolutionally conserved among vertebrates of different species. For instance, 22% of these non-coding elements in human are also conserved in zebrafish (Hiller et al., 2013). Various conserved non-coding elements (CNEs) can act as *cis*-regulatory elements that are associated with gene regulation. The nature of different cell types in vertebrates is based on genomic regulatory information. Regulatory information is controlled by specific gene regulatory networks (GRNs) created by TFs that cooperate to activate or repress the expression of target genes (Nowick and Stubbs, 2010). The activation and gene expression of GRNs are supported by *cis*-regulatory elements such as promoters and enhancers, which both contain specific DNA motifs to which TFs binds. Enhancers containing specific DNA binding motifs form a link between TFs and the target gene within the GRNs (Spitz and Furlong, 2012). *c-Myb* is an important TF that is evolutionally conserved in vertebrates during haematopoiesis (Soza-Ried et al., 2010). *c-myb* is a proto-oncogene,

whose expression produces a transcription factor that is involved in the differentiation and proliferation of neutrophil cells during haematopoiesis. However, the overexpression of *c-myb* gene could be a major health issue that may lead to leukaemia and other cancers like breast and colon cancer (Mitra, 2018). The expression of the *c-myb* gene is controlled by regulatory elements including enhancers (Zhang et al., 2016). Since the c-Myb transcription factor is essential for regulating the haematopoietic process, it will be interesting to understand how the *cis*-regulatory elements associated to the *c-myb* gene expression carry out the regulation of haematopoiesis. This study attempts to elucidate some of the important functional roles of CNE in the expression of c-Myb transcriptional regulation in zebrafish model organism. Zebrafish share about 70% of their genetic structure with humans (Howe et al., 2013). Understanding the regulation of haematopoiesis in zebrafish will also help us to understand the regulation of haematopoiesis in humans. Findings from this study, once published, will provide molecular insight into the functions of the regulatory elements in enhancing the expression of the *c-myb* gene for the regulation of the haematopoietic process. Information and data to be generated from this study could be relevant to studies on *cis*-regulatory elements as well as to researchers, clinicians, and educators.

1.10 Aims and objectives

c-Myb is an important TF that is evolutionally conserved in vertebrates during haematopoiesis. It plays a vital role in maintaining haematopoietic progenitors and regulating their differentiation (Soza-Ried et al., 2010). The expression of the *c-myb* gene is controlled by regulatory elements, including enhancers (Zhang et al., 2016). Enhancers are *cis*-regulatory elements that are important in the gene expression of specific cell types. Enhancers are usually bound by specific TFs to increase the transcription of a gene. Distal enhancers (-28 kb) are bound by the TFs Hoxa9 and PU.1 and play an important role in the regulation of the expression of *c-Myb* gene during mouse myeloid differentiation (Zhang et al., 2016). During erythroid differentiation in mice, the transcription regulation of *c-Myb* gene is controlled by distal enhancers that are bound by specific TFs like KLF1 and the GATA1/TAL1/LDB1 complex (Stadhouders et al., 2012). Previous work has shown that the downstream *c-myb* enhancer is sufficient to enhance the *c-myb* promoter activity in definitive haematopoietic cells in zebrafish embryos and adults (Hsu, 2010). Since our enhancer downstream of the zebrafish *c-myb* gene is sufficient to direct *c-myb* promoter activity in the definitive haematopoietic cells, we wanted to know whether this element was essential for *c-myb* expression during zebrafish haematopoiesis and whether it was functionally conserved

among vertebrates. To achieve this aim, we carried out the following objectives;

1. We examined the exact nature of the haematopoietic cells that express our reporter transgene in the adult kidney using single-cell RNA sequencing (scRNA-Seq).
2. We deleted the zCNE1 from the *c-myb* locus using the CRISPR/Cas9 system to determine whether the zCNE1 is also necessary for *c-myb* expression in the definitive blood cells.
3. We also deleted nearby potential enhancers (PEs) downstream of zebrafish *c-myb* gene that have been identified through the identification of histone modification. We wanted to see whether the PEs contributed to *c-myb* expression in definitive blood cells.
4. We tested whether the conserved human enhancer is able to drive *c-myb* promoter activity to definitive haematopoietic cells in the embryo and the adult zebrafish.

2 Materials and Methods

2.1 Materials

2.1.1 Buffers and solutions

Table 2.1 List of buffers and solutions

All buffers and solutions used were listed including their ingredients and the temperature they were stored at.

Number	Buffers/ solutions	Ingredients	Store
1	0.03% PTU	0.3g U-Phenylthiourea (PTU) (Sigma, P7629) in 1 L sterile dH ₂ O.	4°C
2	0.2X SSCTw	5 ml 2xSSC, 45 ml dH ₂ O; 225 µl 20%Tween20	RT
3	10mg/ml Proteinase K	Proteinase K from Tritirachium album (Sigma, P2308 100 MG, 30 units/mg) in glycerol storage buffer.	-20°C
4	1X HBSS buffer	10XHBSS (GIBCO, 4065049); sterile dH ₂ O	RT
5	1X Phosphate buffered saline (PBS)	10X PBS (GIBCO, 14190250); sterile dH ₂ O	RT
6	1X Phosphate buffered saline Tween20 (PBSTw)	1xPBS; 0.1% Tween20	RT
7	20X Saline Sodium Citrate (SSC) pH7	3 M NaCl; 0.3 M Sodium Citrate	RT
8	2X SSC Tween20 (SSCTw)	5 ml 20xSSC; 45 ml dH ₂ O; 225 µl 20%Tween20	RT
9	4% paraformaldehyde (PFA)	2 g paraformaldehyde (Sigma, 15812-7); 1M NaOH; 1xPBS	4°C
10	50mM NaOH	0.1 ml of 1 M NaOH in 2 ml of dH ₂ O	RT
11	50X TAE buffer (stock)	Tris Base; glacial acetic acid; 0.5 M EDTA (pH8); dH ₂ O	RT
12	5X base buffer	0.5 g NaOH; 2 ml 0.5 M EDTA (pH8) bring up the volume to 100ml with sterile dH ₂ O	RT

Number	Buffers/ solutions	Ingredients	Store
13	5X Neutralisation Solution	3.15 g 1M Tris-HCl (pH8) in 100 ml of sterile dH ₂ O	RT
14	60x E3 Buffer	34.4 g NaCl; 1.52 g KCl; 5.8 g CaCl ₂ .2H ₂ O; 9.8 g MgSO ₄ .7H ₂ O made up to 2 L with dH ₂ O.	RT
15	Ampicillin- containing LB-Medium (100µg/ml)	300 µl of Ampicillin (100 mg/ml) in 300 ml of LB-Medium	4°C
16	BCL3 Buffer	0.1 mM Tris-HCl (pH9.5); 0.1 mM NaCl; 50 mM MgCl ₂ ; 0.1% Tween20	RT (up to 2 weeks only)
17	DNA loading dye	6X gel loading dye (NEB, B7024S)	RT
18	E3 Buffer (working solution)	15 µl of 60x E3 Buffer & 500 µl of methylene blue in 900 ml of dH ₂ O	RT
19	Embryo Lysis Buffer (ELB) stock solution	10 ml 1M Tris-HCl pH8.5; 0.4ml 0.5 M EDTA pH8.0; 10 ml 10% Tween 20 (0.5%) in 200 ml sterile dH ₂ O	RT
20	Embryo lysis buffer (ELB) working solution ProK (10mg/m)	ELB stock solution; 300 µg/ml Proteinase K to be added fresh only at the time of use	Make up fresh
21	Ethidium bromide	Ethidium bromide solution for fluorescence, ~1% in H ₂ O (Sigma cat. #E1510)	RT
22	FCS/PBS	4.5 ml of 10xPBS; 2.5 ml of FCS and 45 ml sterile dH ₂ O	-20°C
23	Hybe ^{-/-} Buffer (pH6.0)	50% Formamide (Sigma, F9037); 5X SSC (pH7); 9.2 mM Citric acid; 0.1% Tween20; sterile dH ₂ O	-20°C
24	Hybe ^{+/+} Buffer (pH6.0)	50% Formamide (Sigma, F9037); 5X SSC (pH7); tRNA (50 mg/ml) (torula yeast, Type VI; Sigma, R6625) 50 mg/ml Heparin; 9.2 mM Citric acid (1 M); 0.1% Tween20; sterile dH ₂ O	-20°C
25	Kanamycin - containing LB-Medium	150 µl of Kanamycin (50 µg) in 300 ml of LB-Medium	4°C
26	LB- Medium	10 g Bacto-Tryptone; 5 g Bacto-yeast extract; 10 g NaCl; dH ₂ O up to 1 liter	RT
27	MAB Block	2% Roche-Boehringer Blocking reagent TM in MAB	-20°C
28	MABTw	50 ml MAB Buffer; 250 µl 20% Tween20 (0.1%) added when needed	4°C

Number	Buffers/ solutions	Ingredients	Store
29	Maleic Acid Buffer (MAB) pH7.5	0.1 M Maleic acid; 0.15 M NaCl sterile dH ₂ O; adjust pH to 7.5 using NaOH pellets; autoclave	4°C
30	MS222 4g/L pH7-7.5	4 g Ethyl 3-aminobenzoate (Tricaine) methanesulfonate (MS222) (Sigma, A5040-100G) adjust pH with Tris-HCl pH9.5	4°C
31	PBSTw 20 mM EDTA	PBSTw; 0.5 M EDTA pH8	RT
32	Proteinase K working solution for WISH experiments	Use 0.5 µg/ml of PBSTw; i.e. 1 µl of stock of Proteinase K (10 mg/ml) in 1 ml of PBSTw.	Make up fresh
33	SOC Medium	Tryptone 2% (w/v); Yeast extract 0.5% (w/v); NaCl 10 mM; KCl 2.5 mM; MgCl ₂ 10 mM; Glucose 20 mM	RT
34	Tween20	10 ml Tween 20; 40 ml dH ₂ O	RT

2.2 Methods

2.2.1 Single-cell gene expression profile

2.2.1.1 The 10x Genomics Chromium system

2.2.1.1.1 Preparation, constructing and sequencing of single-cell gene expression of zebrafish kidney marrow

The kidney of transgenic zebrafish line *qmc85* was prepared as described in 2.2.5.4 and 3.1.1. Gated GFP-Positive cells were sorted using flow cytometry then sent to Deep Seq unit, University of Nottingham Next Generation Sequencing Facility. These steps were performed by Dr Nadine Holmes in the School of Life Sciences' Deep Seq unit. In summary, the procedure used by deep seq involved the preparation of a single cell 3' Whole Transcriptome Sequencing library from a dissociated cell suspension using the Chromium Single Cell 3' Library and Gel Beak Kit v2, the Chromium Single Cell A Chip Kit and the Chromium Multiplex Kit, 96 reactions (10X Genomics; PN-120267, PN-1000009 and PN-120262). Cell counts and viability estimates were obtained using the LUNA-II Automated Cell Counter (Logos Biosystems), Trypan Blue Stain (0.4%) and Luna Cell Counting Slides (Logos Biosystems; T13001 and L12001). Cells that are not stained with Trypan blue are live cells, while those stained with Trypan blue are dead cells. In LUNA-II Cell Counter, live cells were labelled with green circles, and dead cells were labelled with red circles, making it easy to confirm the accuracy of each count. Because the viability of 82% was only just below the 90%

recommended by the manufacturer

<https://kb.10xgenomics.com/hc/en-us/articles/115001800523->

[What-is-the-minimum-number-of-cells-that-can-be-profiled-](#), it was

decided to carry on the experiment (Table 2.2). The number of input

cells targeted was 5000 cells per sample, with the aim of generating

sequencing libraries from ~ 2,500 single cells. The 10x Genomics

Chromium system uses microfluidics to combine individual cells with

Single Cell 3' v2 Barcoded Gel beads (Gel Bead-In-EMulsions

(GEMs)) in tiny water droplets in oil. In these droplets, the cells were

lysed to release their RNA and the 3' ends of the polyadenylated

RNAs were reverse transcribed and marked with a cell-specific

barcode as well as a transcript-specific unique molecular identifier

(UMI). All the steps, including GEM Generation and Barcoding, Post

GEM-RT Clean-up and cDNA Amplification and Library Construction,

were performed according to the Chromium Single Cell 3' Reagents

Kits v2 User Guide. Variable steps of this protocol included using 12

cycles of cDNA amplification and 14 cycles of library amplification.

Amplified cDNA was quantified using the Qubit Fluorometer and the

Qubit dsDNA High Sensitivity (HS) Assay Kit (ThermoFisher

Scientific; Q32854), and fragment length profiles were assessed

using the Agilent 2100 Bioanalyzer and Agilent High Sensitivity DNA

Kit (Agilent; 5067-4626). Completed sequencing libraries were

quantified using the Qubit Fluorometer and the Qubit dsDNA HS

Assay Kit and fragment length distributions assessed using the

Agilent 4200 TapeStation and the HS D1000 ScreenTape Assay (Agilent; 5067-5584, 5067-5585) (Table 2.3). This library was pooled in equimolar amounts and the final library pool was quantified using the KAPA Library Quantification Kit for Illumina Platforms (Roche; KK4824). Libraries were sequenced on the Illumina NextSeq 500 using a NextSeq 500 High Output v2.5 150 cycle kit (Illumina; 20024907) to generate > 50,000 raw reads per cell for each sample, using custom sequencing run parameters described in the protocol. All sequences were then analyzed with cell ranger version 2.1.1.

Table 2.2 Sample QC

Sample	Cell conc. (cells/ μ l)	Viability %	Amount cells used (μ l)	Amount H2O
Martin_Zfish_Kidney	1200	> 82	5	28.8

Table 2.3 Library Preparation QC

Sample	No. cycles cDNA Amp	cDNA conc. - Qubit HS (ng/ μ l)	No. cycles Library Amp	Library conc. Qubit HS (ng/ μ l)	Dilution factor for HSD1000	Av frag size TS	nM
Martin_Zfish_Kidney	12	1.74	14	36.3	15	444	123.87

2.2.1.2 Cell Ranger analysis

Cell Ranger is a set of pipelines that are used to process output from Chromium single-cell data. Cell Ranger version 2.1.1 was used to align reads, generate feature-barcode matrices followed by clustering and gene expression analysis of our 10x Chromium single-cell RNA-seq output (2.2.1.1.1) The processes of the Cell Ranger constitute the conversion of Illumina raw's base call files (BCLs) into

FASTQ files using 10x Cell Ranger mkfastq. The FASTQ files were then imported into Cell Ranger Count to generate the gene cell matrix. Here, the sequences of barcode and UMI were extracted and aligned to the transcriptome. The non or duplicated barcodes and UMIs were then filtered. The clustering differential expression gene was run by the following steps; Principal Component Analysis (PCA) was run on filtered gene cell matrix to reduce the number of genes, graph-based (Louvain) and *K*-means clustering were then performed to identify clusters of cells, t-SNE (t-stochastic neighbour embedding) dimensionality was then run to create the two-dimensional projection, sSeq (a bioinformatic tool for differential gene expression) analysis was then used to find the genes that differentially characterize each cluster. These steps were performed by Dr Fei Sang, a bioinformatician based in Deep Seq. 10x Genomics' Loupe Cell software (version 3.1.0) was then used to visualize the files produced by Cell Ranger.

2.2.1.3 Re-analysis of Dinh and colleague' and Grassi colleagues' published data

In order to re-analyse Dinh and colleagues' data (Dinh et al., 2020), the Short Read Archive (SRA) files for their data were obtained from NCBI, which was deposited in the Gene Expression Omnibus (GEO) under the accession number GSE153263 and imported to CyVerse online analysis platform (<https://www.cyverse.org/>). In CyVerse, the SRA files were converted into FASTQ files using NCBI-SRA-Fastq-

2.8.1 application. The sequence reads were then mapped against the human reference genome and gene annotation (GRCh38.p13) to obtain Binary Alignment Map (BAM) files using HISAT2-index-align-2.1 application. The BAM files then were imported into the SeqMonk program (v1.47.1) using Homo sapiens GRCH38_v97 for visualization and analysis data. Regarding Grassi and colleagues' study, the comma separated values (CSV) tables were obtained from their data then imported to GraphPad PRISM (v9.0.0 (121)) software for more analysis (Grassi et al., 2018).

2.2.2 Targeting interest regions of downstream of *c-myb* gene using CRISPR/Cas9 system

2.2.2.1 Single guide RNA (sgRNA) design

Suitable target sites (TSs) within the region downstream of *c-myb* of zebrafish (upstream and downstream of zRE1) were identified using the CRISPRscan UCSC track tool (<https://www.crisprscan.org/>) (Moreno-Mateos et al., 2015) (Figure 2.1A). The CRISPRscan tool identifies target sites as 20-bp sequences that start with a GN/NG dinucleotide and are followed by an NGG PAM sequence. All suitable TSs must have sequences of either 5'-GN-(N₁₈)-PAM-3' or 5'-NG-(N₁₈)-PAM-3'. In the sgRNA, these sequences are modified to a 5'-GG-N18-3' sequence. The PAM sequence at the 3' end is required by Cas9. The GN/NG dinucleotide at the 5' end is needed to allow efficient *in vitro* transcription of the

single guide RNA gene by T7 RNA polymerase once the TS sequence was cloned into the sgRNA expression vector pDR274 that was created by Joung lab's (Hwang et al., 2015). Primers were designed and they contain 5' overhangs to allow the integration into the *BsaI* digested pDR274 vector fragment. All oligonucleotides were ordered from Sigma-Aldrich (Figure 2.1B). Table 2.4 shows all TSs sequences and primers designed.

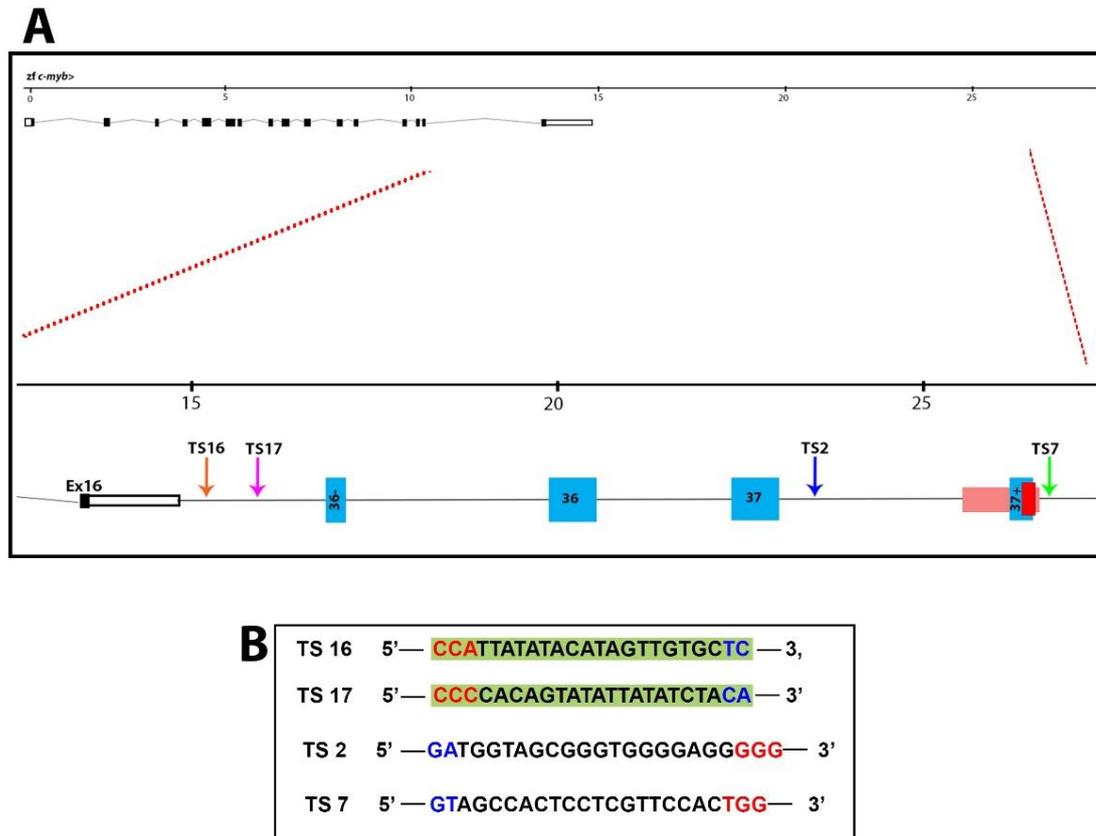


Figure 2.1 Identification of suitable Cas9 target sites (TSs)

(A) The location of candidate TSs 16, 17, 2 and 7 are indicated by orange, purple, blue and light green lines, respectively. The zebrafish conserved noncoding element number one (zCNE1) is shown as a red box. A pink box shows the zebrafish regulatory element one (zRE1). The light blue boxes show the potential enhancers MP40036-, MP40036, MP40037 and MP40037+. The latter PE MP40037+ overlaps with zRE1. **(B)** Sequences of suitable Cas9 target sites are shown as they appear on the plus strand of chromosome 23. Please note that TSs 16 and 17 are on the minus strand while TSs 2 and 7 are on the plus strand of chromosome 23. Blue letters indicate the dinucleotides at the 5'-ends of target sites. These dinucleotides are required to be GG, GN or NG sequences. In the guide RNA expression constructs, GN and NG sequences are modified to GG dinucleotides to allow the efficient use of the T7 promoter in the in vitro transcription reaction. Red letters represent the protospacer adjacent motif (PAM) sequences.

Table 2.4 List of the sequences and oligonucleotides used for targets site

TS	Strand	Sequence 5'→3' with PAM A = Point of SNP T>A	Score value	Off- target site	Gering Lab Data Base (DB) # Foreword Primers/Revers Primers	Oligos sequence (F)5'→3'/(R)3'→5' with overhangs
2	+	GATGGTAGCGGGTGGGG AGGGGG	92	0	DB978	taGgTGGTAGCGGGTGGGGAGG
					DB979	aaacCCTCCCCACCCGCTACCA
7	+	GTAGCCACTCCTCGTTCCACT TGG	44	0	DB1263	taGgAGCCACTCCTCGTTCCAC
					DB1264	aaacGTGGAACGAGGAGTGGCT
16	-	CCA TTATATACATAGTTGTGCTC	42	0	DB1363	taGgGCACAACCTATGTATATAA
					DB1364	aaacTTATATACATAGTTGTGC
17	-	CCC CACAGTATATTATATCTACA	51	0	DB1367	tagGTAGATATAATATACTGTG
					DB1368	aaacCACAGTATATTATATCTA

2.2.2.2 The generation of the complementary sgRNA

To clone the TS sequences into pDR274, two complementary TS sequence oligonucleotides were made that carried additional sequences at their 5' ends that generated cohesive ends that were compatible with the 5' overhangs that were generated by cutting the pDR274 vector with the *BsaI* restriction endonuclease as described in section 2.2.4.13 (Figure 2.2Ai-ii). The oligonucleotide pairs were annealed and then ligated with the *BsaI*-digested vector fragment (Hwang et al., 2013b) (Figure 2.2Aiii). Successful *BsaI* digestion of the vector had been confirmed by agarose gel electrophoresis (Figure 2.2Aiii). The ligation products were used to transform chemically competent *E.coli*. Kanamycin-resistant bacteria were identified on Kanamycin-containing agar plates. Plasmid DNA was prepared from these bacteria. The DNA was digested with *BsaI* and *DraI*. Plasmids that contained a TS sequence insert were expected to be digested with *DraI* but not with *BsaI* as the integration of the TS sequence should abolish the *BsaI* restriction sites in the vector (Figure 2.2B). The DNA sequences of the plasmids were subsequently confirmed by Sanger sequencing (Figure 2.2C). Plasmids that contained the correct TS sequence were used to generate sgRNA run-off transcripts. For this purpose, *DraI* was used to linearise the sgRNA vector as described in section 2.2.4.14 (Figure 2.2D-F). Next, T7 RNA polymerase was used to *in vitro* transcribe the plasmid and generate a sgRNA (Figure 2.2G). In this project, all

gRNAs were designed and produced using the same procedure (Figure 2.3A-J).

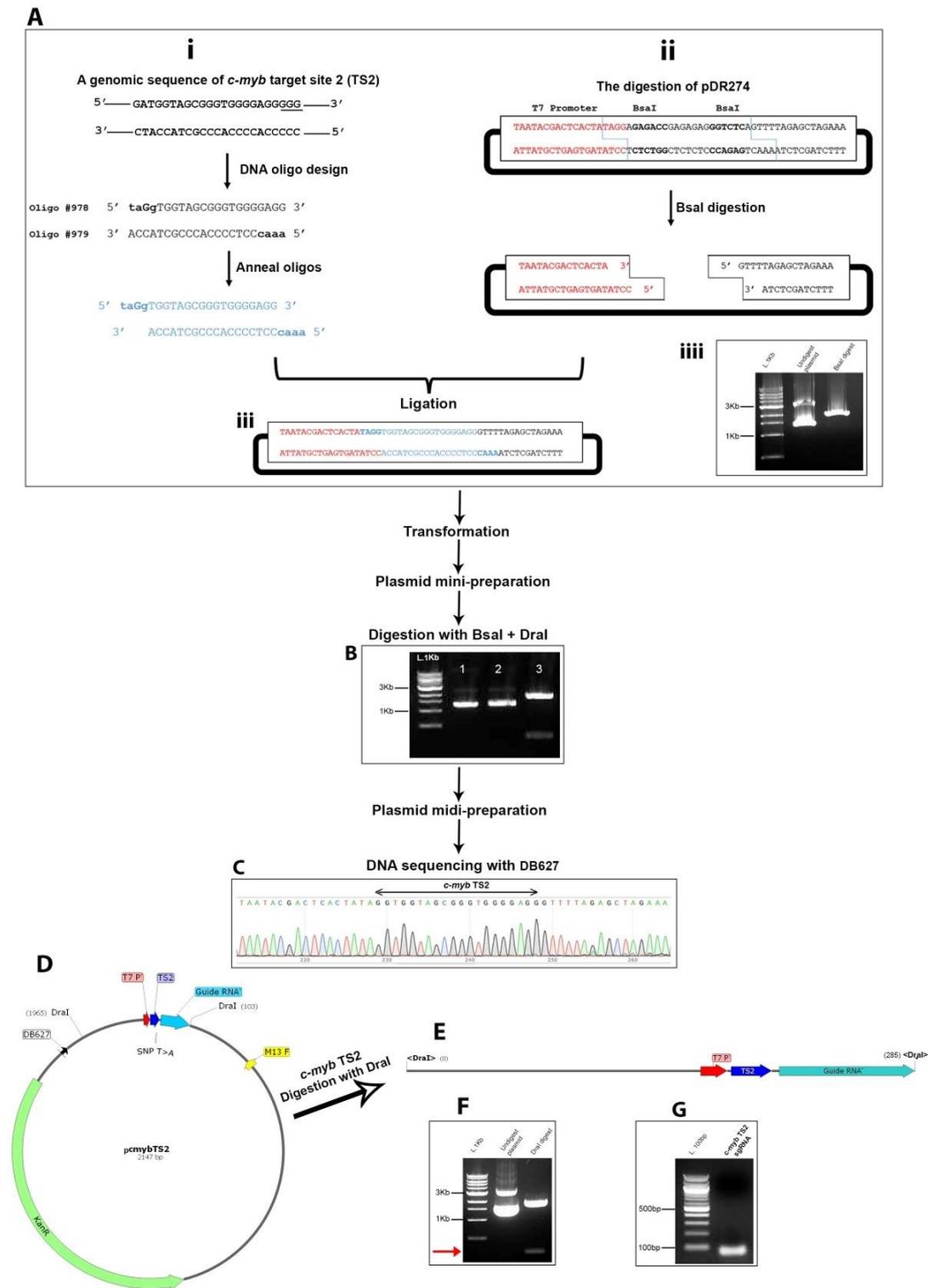


Figure 2.2 sgRNA design and synthesis

sgRNA of *c-myb* TS2 is used as an example. **(A)** the diagram shows the cloning of TS2 into the pDR274 vector. **(i)** Generation of the TS2 linker with pDR274-compatible sticky ends. **(ii)** Generation of the pDR274 vector fragment by BsaI digestion. **(iii)** A sequence of the recombinant pDR274 plasmid that carries the TS2 linkers after ligation (blue letters). **(iiii)** A representative agarose gel after electrophoresis of a restriction digest of

plasmid DR274 with *BsaI*. After transformation of *E.coli* with the ligation mix, plasmids were prepared and digested with the restriction enzymes *BsaI* and *DraI*. These digests were used to identify the colonies that carry the plasmid with the TS2 insert. Successful cloning of the TS sequence eliminates the *BsaI* restriction sites from the plasmid. Thus, plasmids with TS sequences can be digested with *DraI* but not with *BsaI*. **(B)** A representative gel of plasmid DNAs digested using *BsaI* and *DraI* enzymes. Lane 1 shows undigested plasmid. Lane 2 depicts *BsaI* digest. Lane 3 show plasmid digested with *DraI*. **(C)** Sanger sequencing with primer DB627 shows the presence of TS2 within the pDR274 plasmid. **(D)** Map of the pcmybTS2 plasmid containing *c-myb* TS2 insert. **(E)** Map of *c-myb* TS2 plasmid linearised with *DraI* which contains T7 promoter. **(F)** Shows representative gel of *c-myb* TS2 plasmid digested using *DraI* enzyme. A red arrow refers to the *DraI* fragment. **(G)** Shows representative agarose gel after electrophoresis of the *c-myb* TS2 sgRNA product generated in an *in vitro* transcription reaction with T7 RNA polymerase. Please note that a DNA ladder was used here. As single-stranded RNA migrates through the gel faster than double-stranded DNA of the same size, sizes are not directly comparable.

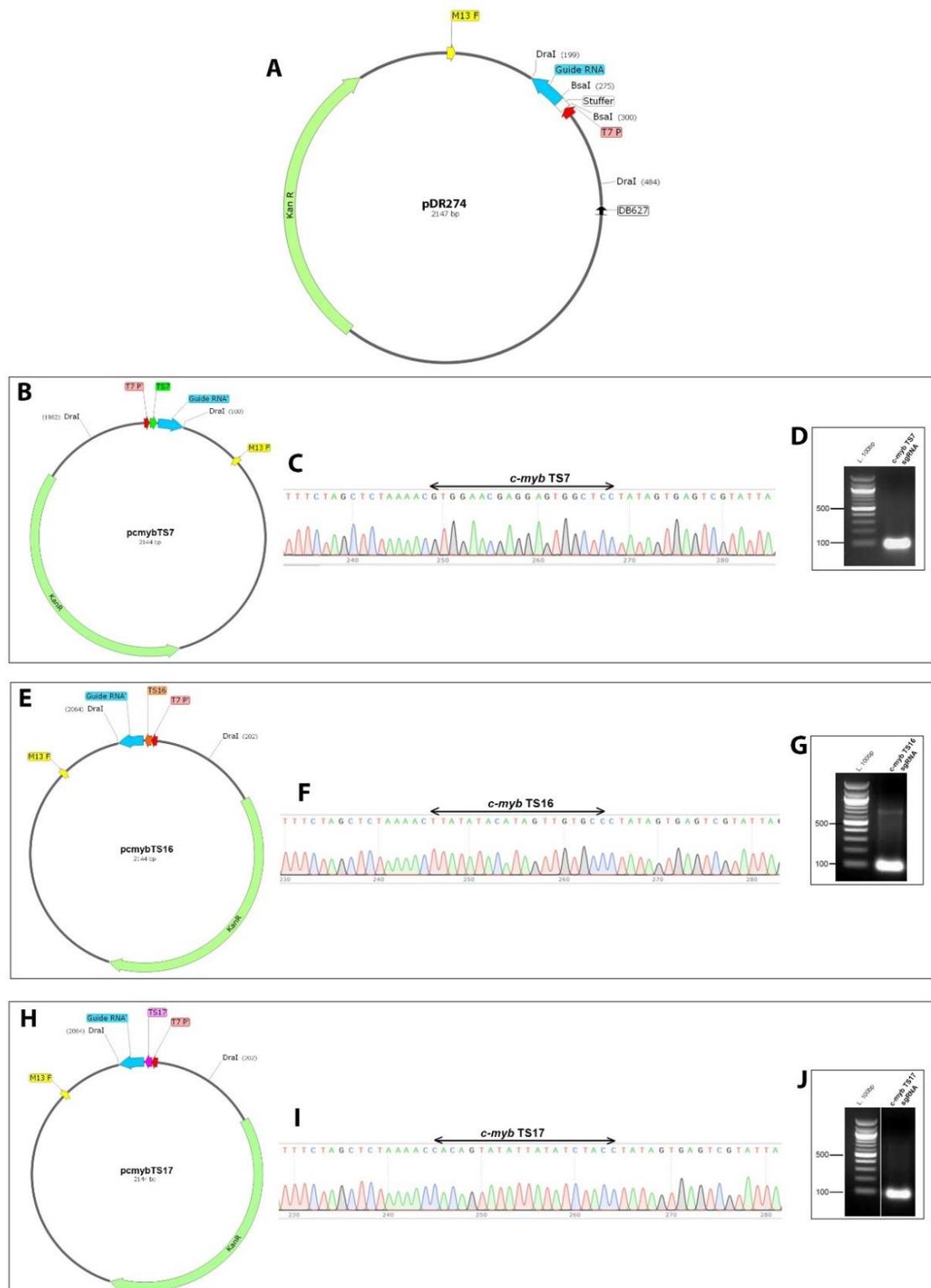


Figure 2.3 Successful cloning of the TS7, 16 and 17 linkers

(A) Map of plasmid pDR274. (B, E and H) Maps of the recombinant pDR274 plasmids that carry the TS7 (green arrow), TS16 (orange arrow) and TS17 (purple arrow), respectively. (C, F and I) Sanger sequencing with M13 forward primer confirms the presence of TSs 7, 16, 17 within the pDR274 plasmid. (D, G and J) Show representative agarose gel after

electrophoresis of the *c-myb* TS7, 16 and 17 sgRNAs products generated by *in vitro* transcription with T7 RNA polymerase.

2.2.3 Establishment of stable human regulatory element one (hRE1) zebrafish transgenic line

2.2.3.1 Generation of a GFP reporter construct containing human C-MYB CNE1

The hRE1 fragment was PCR amplified on the genomic DNA isolated from human embryonic stem cells using primers DB1216 and DB1217 (Table 2.5) with a genomic of human embryonic stem cells. The cells were obtained from cell line of male HUES7 hESCs which were maintained in Essential 8™ (E8) medium with 278 supplements (#A1517001) on Matrigel™-coated tissue culture flasks at 37 °C with 5 % CO₂. The genomic DNA of human embryonic stem cells was prepared and kindly provided by Abdulkadir Abakir from Dr Alexey Ruzov's lab.

Table 2.5 List of primers used for genotyping of C-MYB hRE1

Gering Lab Data Base (DB) # Forward Primer/Reverse Primer	Sequence 5'-3'
DB1216	GGGGTTGGAGCAGGAATAATT
DB1217	AAGTTTGGTAGCTTCAGTTTGG

The PCR amplification generated a 1 kb fragment. Its sequence was confirmed by Sanger sequencing. The sequence also revealed the presence of three SNPs relative to the reference sequence (Appendix figure 6. 1). Next, this 1 kb hRE1 fragment was cloned in a two-step process into the Tol2 *c-myb* reporter construct. First, it was inserted into pGEM-Teasy to introduce ApaI and SbfI restriction sites at its ends. Then, these sites were used to clone the fragment in the place of zRE1 in the Tol2 *c-myb* reporter construct on pJC117.

To perform the first cloning step, the 1 kb hRE1 fragment was ligated to the linear pGEM-T Easy plasmid using T4 DNA ligase (Figure 2.4A). Chemically competent *E. coli* were transformed with the ligation mix. Plasmid DNA was prepared from the transformants and was digested with the restriction enzyme *EcoRI* to confirm that the fragment had been cloned. The newly generated plasmid, called pJHB, was expected to carry 3 *EcoRI* restriction sites, one of which is present in the hRE1 fragment (Figure 2.4B). Agarose gel electrophoresis revealed three bands representing the expected ~3 kb vector fragment, as well as ~700 bp and ~300 bp fragments. The *EcoRI* digests verified the presence of three *EcoRI* restriction sites, suggesting that the new plasmid construct carried the hRE1 insert (Figure 2.4B). In addition, the DNA sequence of the plasmid was confirmed by Sanger sequencing (Figure 2.4C). The sequence also revealed the orientation of the hRE1 fragment in the pGEM-Teasy vector.

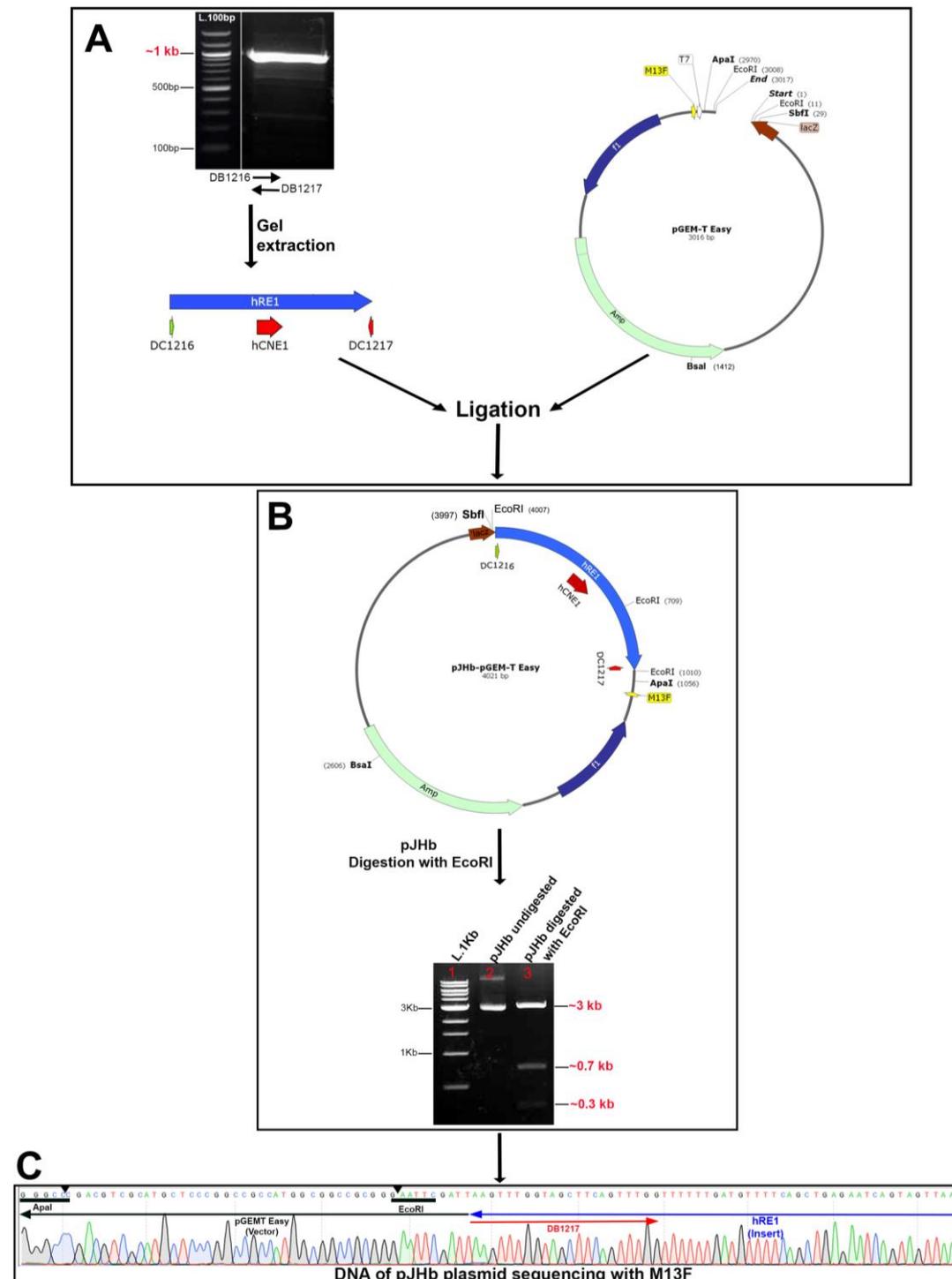


Figure 2.4 Amplification of hRE1 and cloning in pGEM-Teasy

(A) The 1 kb human regulatory element one (hRE1) was amplified by PCR using the primers DB1216 and DB1217 on human genomic DNA. PCR products were separated on a 1% agarose gel and visualised with ethidium bromide (EtBr). The fragment that had the expected size of ~1 kb was purified from the gel. The purified hRE1 fragment was ligated with the linear pGEM-Teasy vector using T4 DNA ligase. Chemically competent *E. coli* were transformed with the ligation mix and plated on Ampicillin-containing agar plates. **(B)** Plasmid DNA was extracted from Ampicillin-

resistant bacteria and digested with *EcoRI*. The products of the restriction digests were analysed by agarose gel electrophoresis. A representative agarose gel is shown here. Lane 1 shows a 1 kb ladder. Lane 2 shows the undigested plasmid. Lane 3 depicts the DNA of the *EcoRI* digest. **(C)** Sanger sequencing using M13F reveals the presence of hRE1 fragment and its orientation within the pGEM-Teasy plasmid. The plasmid was called pJHb. Its restriction map is shown.

In the second step, the *ApaI* and *SbfI* restriction sites that flank the hRE1 fragment were used to replace the zRE1 fragment in the transgene construct on plasmid pJC117. For this purpose, both plasmids were digested with *ApaI* and *SbfI* (Figure 2.5A). Following digestion, preparative gel electrophoresis was carried out to separate the digestion products and to isolate the hRE1 fragment of pJHb and the vector fragment of pJC117 (Figure 2.5B). The two DNA fragments were ligated, and the ligation mix was used to transform *E. coli*. Ampicillin-resistant colonies were identified (Figure 2.5C). Plasmid DNA isolated from resistant bacteria were then subjected to restriction digestion with *ApaI* and *SbfI* to determine whether the hRE1 insert had been ligated into the pJC117 vector. Gel electrophoresis was then performed. The gel revealed a large DNA fragment of approximately 14 kb that corresponded to the pJC117 vector fragment. The small ~1 kb fragment represented the hRE1 insert (Figure 2.6A). Because zRE1 and hRE1 have a similar size, *BsaI* restriction digests were used to distinguish between the two REs. The original pJC117 plasmid carries three *BsaI* restriction sites, one of which is present in zRE1. Unlike zRE1, hRE1 does not possess

a *BsaI* site. Thus, the new hRE1-containing plasmid was expected to have only 2 *BsaI* sites (Figure 2.6B). The *BsaI* digests confirmed the presence of only two *BsaI* restriction sites, suggesting that the new plasmid construct, called pJHE1, carried the hRE1 insert (Figure 2.6B). Sanger sequencing with primer DB307 confirmed the presence of the hRE1 insert (Figure 2.6C). It is worth noting that the orientation of the hRE1 sequence in pJHE1 was identical to that of zRE1 and mRE1 in their respective *Tol2* plasmids. In all three reporter constructs, the orientation of the CNE1 is flipped relative to the orientation of the EGFP reporter gene.

The co-injection of *Tol2* vectors and transposase mRNA were performed by injecting solution mix. This solution was prepared by adding 1 μl of the *Tol2* plasmid (with concentration 450 ng/ μl) and 0.5 μl of transposase mRNA (with concentration 467 ng/ μl), then dH₂O was added to a final volume to 10 μl . One-cell stage wt embryos were injected with 0.5 μl of this solution. The generation of stable hRE1 transgenic line was explained in detail in 3.7.1.

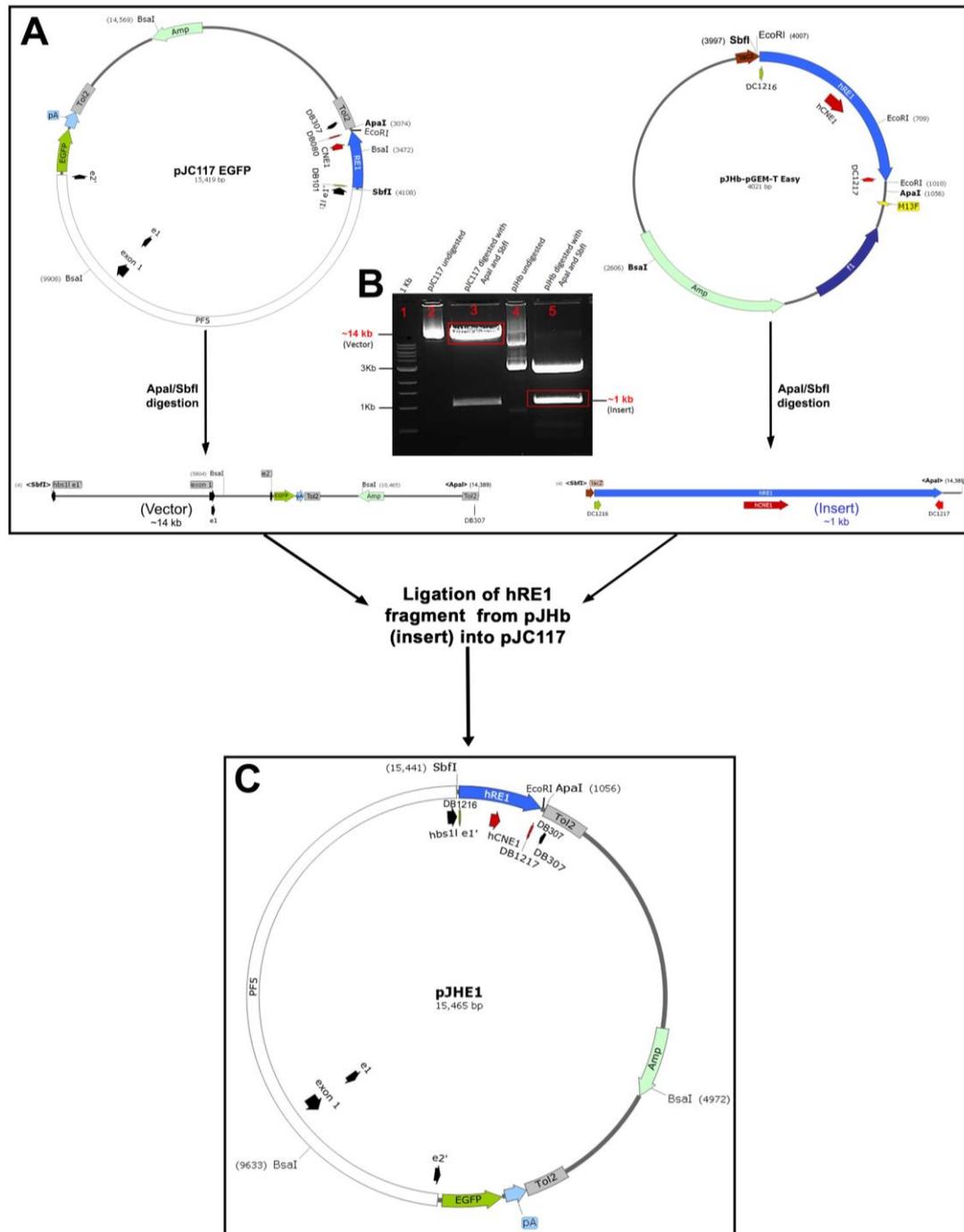


Figure 2.5 Generation of the hRE1-zPF5:egfp reporter transgene – Part 1

(A) Plasmids pJC117 and pJHb were digested using the restriction enzymes *ApaI* and *SbfI*. The diagram shows the restriction maps of pJC117 (vector) and pJHb, as well as the maps of the pJC117 vector and pJHb hRE1 fragments. **(B)** Agarose gel electrophoresis was used to separate the restriction fragments. Lane 1 shows a 1 kb ladder. Lanes 2 and 4 show undigested pJC117 and pJHb plasmids, respectively. Lanes 3 and 5 depict the *ApaI* and *SbfI* digests. Red boxes highlight the ~14 kb pJC117 (vector) and the ~1 kb hRE1 fragment (insert) that were purified from the gel and ligated using T4 DNA ligase. The ligation mix was used to transform *E. coli*.

(C) The diagram displays the restriction map of the expected plasmid pJHE1.

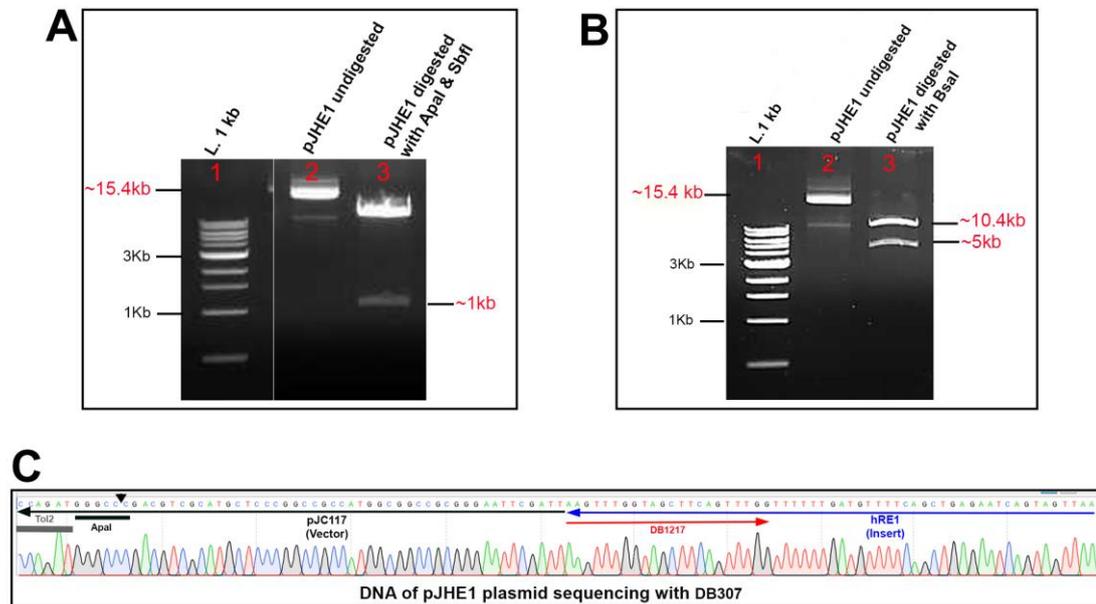


Figure 2.6 Generation of the hRE1-PF5:*egfp* reporter transgene – Part 2

(A) Agarose gel electrophoresis of the *ApaI* and *SbfI* double digest of the newly generated plasmid pJHE1. (B) Agarose gel electrophoresis of the *BsaI* digest of plasmid pJHE1. (C) Sanger sequencing with primer DB307 confirms the presence of hRE1 in pJHE1.

2.2.4 Molecular biology methods

2.2.4.1 Fin clipping of adult zebrafish

To collect fin clip biopsies, adult fish were anaesthetised in a tank that contained 200 ml facility water to which approximately 6 ml of 4 g/L MS222 was added. To make sure that the fish was fully anaesthetised, it was observed for a while until it turns its belly up. The anaesthetised fish was then carefully transferred onto a petri dish or clean surface using a plastic spoon. Next, a fin clip biopsy was obtained using sterilised surgical scissors at a point not more than halfway between the tip of the fin and the point where the scales end. The fin clip was carefully transferred into a 1.5 ml Eppendorf tube, and the tube was placed on ice for further processing. Individually, fully recovered fin clipped fish are housed in a narrow tank. Health monitored daily for one week.

2.2.4.2 Genomic DNA isolation from embryos and fin clip biopsies

Anaesthetised embryos (0.7ml 4 g/l MS222 in 25ml E3 buffer) or fin clip biopsies were moved to tubes. The surplus liquid was removed, and proteinase K working solution (ProK) (see Table 2.1) was added: 20µl to an individual embryo and 50µl per fin clip biopsy. Tissues were incubated overnight at 55°C in a hot block. The next morning, the tubes were vortexed and centrifuged. To inactivate proteinase K, the tube's content was heated at 100°C for 20 minutes. After 20

min, the tubes were cooled for 5 minutes, and sterile distilled water was added to a final volume of 100 μ l. 1 μ l of the gDNA solution was used in a 25 μ l PCR reaction for genotyping.

2.2.4.3 gDNA extraction from embryos after whole-mount RNA in situ hybridisation

After staining in a whole-mount RNA in situ hybridisation experiment (see 2.2.5.3), embryos are fixed and then washed into 80% glycerol/20% PBSTw. Embryos then photographed. Those embryos that were meant to be genotyped were transferred individually into 1.5 ml Eppendorf tubes. The embryos were then slowly washed out of 80% glycerol/PBSTw. Embryos were incubated for 5 min each in 50% and then in 30% glycerol (in PBSTw). Finally, three 5-minute washes in 1X PBSTw followed. The PBSTw was removed after the final wash and embryos were then incubated in 100 μ l of a 300 mM NaCl solution at 65°C for 4 hours. The NaCl was discarded and 25 μ l 1X Base buffer were added to each embryo before the tubes were incubated at 95°C for 20 min. Next, the tubes were cooled for a minute. Then, 25 μ l 1X neutralisation solution was added. At 3000 rpm and for 5 min, the tubes were vortexed and centrifuged at 3000 rpm for 5 min in a microcentrifuge. The gDNA is in the supernatant, and a 2-3 μ l of gDNA used in a 25 μ l PCR reaction.

2.2.4.4 Primer design

Primers for all PCR reactions were designed using Primer 3 web (version 0.4.0) (<http://bioinfo.ut.ee/primer3-0.4.0/>) (Table 2.6) and (Table 2.7).

Table 2.6 List of primers used for genotyping and sequencing of *c-myb* mutants

Mutant	Gering Lab Database (DB) #		Sequence 5'-3'
	Forward Primers	Reverse Primers	
<i>c-myb</i> ^{t25127}	External primers	DB827	TTGGAAGAACTTGAGGGTGAG
		DB828	AGAGCAAGTGGAAATGGCAC
	Internal Primers	DB980	GGAAGAACTTGAGGGTGAGTAAA
		DB981	TGTGTTGCTTGGTGTGTGAT
Sequencing Primer	RP/DB1446	TTGTAGTTCAGGAGGTGAGGTTT	
<i>c-myb</i> ^{qmc193}	External primers	DB1436	AAAACCCCACACACCTGCT
		DB1437	GGGGGAAATACCCTTTTCAA
	Internal Primers	DB871	CACACCTGCTCTCTGTCTGA
		DB899	GCTTTGTGTGCGTCAAGTGT
<i>c-myb</i> ^{qmc194}	External primers	DB1365	TCTTTTCCACTTTGCATCTCC
		DB1437	GGGGGAAATACCCTTTTCAA
	Internal Primers	DB1371	CTTAGTATCCAAATCCATTTTGGAG
		DB732	TGCAGCATTGGATCTTTCAC

Table 2.7 PCR primers of *c-myb* map and locations

A list of all primers used in PCR amplifications and the genomic locations (Ensembl Release 94 October 2018 GRCz11) of all PCR fragments downstream of the *c-myb* gene.

PCR Number on Map	Gering Lab Data	Sequence 5'-3'	Genomic Location (Ensembl Release 94 October 2018 GRCz11)
	Base (DB) # Forward Primers		
	Reverse Primers		
1	FP DB1365 RP DB1366	TCTTTTCCACTTTGCATCTCC TTTGTTTCTATGCGTGTGTGC	chr23 31830040 31830535
2	FP DB885 RP DB1370	ATTATTCAAGGGTTCGGCACA ATAGCAGCAAGGGCAGGATA	chr23 31830547 31831223
3	FP DB930 RP DB931	CTTGTTGAAGCTGCTGTGGT TGTTGCAATCGGTTTCCTCA	chr23 31831665 31832418
4	FP DB887 RP DB888	CACCTGCAGTTCTCTACGA ATGGGCACACATATGAACGC	chr23 31831781 31832866
5	FP DB928 RP DB929	TGTCCAAACACTGACTACGTT ATGCCCTAGTAGAGCTTTATGTC	chr23 31834618 31835606
6	FP DB889 RP DB890	GTCACCTCTGCATTGAAGGC TTGCCAGTGACCCGGTAAAT	chr23 31835863 31836242
7	FP DB891 RP DB892	CATTTGGCGATGCAGGTGT GAGACCCGAGCAGAGTTTCT	chr23 31836409 31837379
8	FP DB932 RP DB933	CAGCAGAAATGATAAAGGCCCA CACACACACACCCTCATTGA	chr23 31836881 31837869
9	FP DB966 RP DB967	GGGTTACAGGAATGGAAAGGC TGGGGTTCGGATTCTCTCC	chr23 31837912 31838471
10	FP DB895 RP DB894	GGAAGAGAATCCGAACCCCA AGTTACGTCTCAGGCTATTG	chr23 31838388 31839226
11	FP DB101 RP DB080	CTGGGAAACATCTACACACACA CTAGTCCAAACCACAGGAAATC	chr23 31840170 31841222
12	FP DB1192 RP DB1193	TATTAGCCGGATGTTGCAGC TGCAGCATTGGATCTTTCACA	chr23 31840991 31841281
13	FP DB1251 RP DB1252	CCAATGCTGCAGTAAACTTGA CATTTTGTGAGCACTTCCGTA	chr23 31841237 31841695
14	FP DB1253 RP DB1254	TTTGTGACATCCCATGTGT CAGGAACTAACAAAGAACGACA	chr23 31842038 31842560
15	FP DB1255 RP DB1256	CCACCTAAGCTCACATACTGAGA TA AAATTATGAGGCCCCCTTTG	chr23 31842265 31843055
16	FP DB1257 RP DB1258	CATTAAGTGGCAGGTTAGGGTAA AGACTGGCTTTGTGAGTTGGA	chr23 31842707 31843365
17	FP DB900 RP DB901	ACCTGGCCAATTGAGAGACA GGCGGGAATTGAATTTGGGT	chr23 31843498 31844473
18	FP DB934 RP DB935	CACACACACTGACCCCAA TAGTTTGAACCACTTACCACT	chr23 31846610 31847598
19	FP DB902 RP DB903	TCCACCCACATTAGACACG CCCAGTGACAGGTTTTGCAT	chr23 31848547 31848804

2.2.4.5 Polymerase chain reaction (PCR)

PCR is a method generally used in genetic screening, for genotyping or to amplify a DNA fragment of interest. In this project, it was carried out in 0.2 ml PCR tubes with a reaction volume of 25–50 μ l using a TECHNE TC-512 thermocycler (Bibby Scientific). The reaction components and cycling conditions used in this PCR are summarised in Table 2.8 and Table 2.9:

Table 2.8 PCR thermocycling conditions

Components	Total volume 25 μl
10X Standard Taq reaction buffer	2.5 μ l
dNTPs (2 mM)	2.5 μ l
Forward primer (10 μ M)	1 μ l
Reverse primer (10 μ M)	1 μ l
DNA template (100 ng/ μ l)	1 μ l
NEB Taq DNA polymerase (1.25 units)	0.5 μ l
dH ₂ O	16.5

Table 2.9 PCR reaction volumes

Thermocycler Steps		Temperature	Time
Initial denaturation		94 °C	2 min
40 Cycles	Melting	94 °C	15 sec
	Annealing	55-65 °C	30 sec
	Extension	72 °C	1 min/1kb
Final Extension		72 °C	5 min
Hold		4 °C	

2.2.4.6 Agarose gel electrophoresis

PCR products were examined by agarose gel electrophoresis. In general, agarose gels were prepared by dissolving agarose powder (Thermo Fisher Scientific, Waltham, MA, USA) (1-2 % based on the size of the expected PCR fragment) in 0.5x (for DNA) or 1.0x (for

RNA) Tris-acetate-EDTA (TAE) buffer in a microwave. Once gel cooled down, 0.5 mg/ml of ethidium bromide (EtBr) was added then gel solution was poured into a gel tray. DNA loading dye (0.1 x volume) was added to all samples before they were loaded into the gel. One kb or 100 bp NEB DNA Ladder were run with samples as size markers. To separate DNAs of different sizes, the electrophoresis performed at 100 V for 30 min in an Advance Mupid®-ex electrophoresis chamber. Preparative gels used to isolate specific DNA fragments were run more slowly at 50 V for 1 hour to achieve better fragment resolution. The DNA fragments were visualized on a UV transilluminator (Bio-Rad Chemi XRS Gel Documentation system).

2.2.4.7 DNA extraction from agarose gel

To obtain DNA, the band of interest was cut out using a scalpel while the gel was placed on a UV illumination box. A GenElute PCR Clean-Up Kit (Sigma cat. #NA1020) was used to extract DNA from the agarose, following the manufacturer's instructions.

2.2.4.8 Ligation

DNA ligations were prepared using NEB T4 DNA ligase (NEB). The DNA of interest was ligated into the vector at a molecular ratio of 1:3 vector to insert as followed Table 2.10:

Table 2.10 The general ligation reaction

Component	10µl Reaction
10x T4 DNA ligase buffer	1µl
Vector	50 ng
Insert DNA	150 ng
T4 DNA ligase	0.5 µl
dH ₂ O	up to 10 µl

The ligation reaction mixture was incubated at room temperature for 4 hours or at 16 °C overnight and used to transform chemically competent *E.coli* cells the following day.

2.2.4.9 Transformation and Mini/Midi-preparation of plasmid DNA

A 50 µl aliquot of competent subcloning efficiency *E.coli* (DH5α, NEB, Cat. #C2987H) or 20µl of *E.coli* (5-alpha *E.coli* high efficiency, NEB, Cat. #C2987I) were transformed with ligation mixes or established plasmid DNAs. Cells were thawed on ice for 15 min. Two to five µl of ligation mix or a small volume of plasmid DNA solution (e.g. 0.5 µl of a 200 ng/µl solution) were added to the competent cells. The tube was flicked carefully 4–5 times to mix cell and plasmid DNA. The suspension was kept on ice for 30 min and then heated shocked at 42°C for 40–90 seconds (based on plasmid size). The mix was placed again on ice for 5 minutes. Subsequently, 500 µl of SOC Medium were added to the bacteria cells. The cell mixture was then incubated at 37°C in a shaking incubator for 1 hour to achieve phenotypic expression. Next, 150–200 µl of the culture was plated out on LB agar plates that contained Kanamycin (50µg/ml) or Ampicillin

(100µg/ml) depending on the resistance encoded by the plasmid vector. The plates were incubated at 37°C overnight. The next day, single colonies were picked on a fresh agar plate and used to inoculate liquid cultures, 3 ml cultures for plasmid mini-preps or 100 ml cultures for plasmid midi-preps. The liquid LB media contained 50 µg/ml kanamycin or 100 µg/ml ampicillin depending on the antibiotic resistance encoded by the plasmid vector. The plasmid DNA was isolated using the Sigma GenElute Plasmid Miniprep Kit (PLN-70) or the Nucleobond Xtra MidiPrep Kit (NucleoBond Xtra Midi Plus, cat. #740412.10) for plasmids of interest.

2.2.4.10 Restriction enzyme digest

DNA Restrictions were typically carried out in a total reaction volume of 20 µl (Table 2.11). All restriction digests were prepared using the manufacturer's recommended buffer. Most reactions were incubated for 2 h at 37°C or at the recommended temperature. All restriction enzymes and buffers were from New England Biolabs (NEB).

Table 2.11 The general restriction digest reaction

Components	Total volume 20 µl
Restriction enzyme	10 units
DNA template	1 µg
1x Buffer	2 µl
dH ₂ O	up to 20 µl

2.2.4.11 Measuring DNA and RNA concentration

The concentrations of purified DNA and RNA were determined using a Nanodrop Spectrophotometer ND-1000 (ThermoFisher Scientific, UK). An A260/280 absorbance ratio was considered to indicate high purity of DNA (≈ 1.8 – 2.2) and RNA (≈ 1.7 – 2.0).

2.2.4.12 DNA Sequencing

For Sanger sequencing, 5 μ l of purified PCR product (1 ng/100 bp) or plasmid DNA (100 ng/ μ l), as well as 5 μ l of the appropriate primer solution (10 μ M) were sent to Source BioScience Ltd. (Nottingham, UK).

2.2.4.13 Cloning of the target site linker in pDR274

The pDR274 plasmid was digested with *BsaI* (NEB, cat #R3535L) at 37°C for overnight. On the following day, the sample was run on 1% agarose gel to check that the DNA had been digested. Then, the vector fragment was isolated from the agarose gel. Next, to make a double-stranded DNA linker, 10 μ l of each of the two DNA oligos (100 μ M) encoding the target site were mixed and heated at 95 °C for 5 min and subsequently cooled to RT to allow the oligos to hybridise to form the linker. The linker was ligated to the *BsaI* vector fragment using T4 DNA ligase overnight at 16°C. The ligation mix is shown in (Table 2.12)

Table 2.12 The ligation mix to insert the linker into the pDR274 vector

Components	10 μl Reaction
Annealed linkers	2 μ l
pDR274 <i>BsaI</i> free	5 μ l
10x ligation buffer	1 μ l
dH ₂ O	1.5 μ l
T4 DNA ligase	0.5 μ l

The next day, the mixture was used to transform subcloning efficiency *E.coli* (DH5 α , NEB, Cat. #C2987H). After the heat shock and after phenotypic expressed the cells were plated on LB/kanamycin (50 μ g/ml) plates as previously described in section 2.2.4.9. Several colonies grown on the plate were picked on a fresh plate and used to inoculate 3 ml LB medium liquid cultures that were incubated overnight in a shaking incubator at 37 °C. From these cultures, plasmid DNA was extracted using the miniprep kit as described previously. The restriction digest was then performed from the extracted plasmid DNA using *DraI* (NEB, cat. #R0129L) and *BsaI* to confirm that the bacterial colonies carry a plasmid with the correct TS sequence. DNA concentration was measured by using the Nanodrop spectrometer. The samples were sent for sequencing using M13F primer. This was to verify the sequence of the TS sequence. Subsequently, plasmid midi preps were performed on bacteria grown in 100 ml overnight cultures as described in the previous section.

2.2.4.14 sgRNAs synthesis for Microinjection

To cut the *DraI* fragment from plasmid, 10 µg of the customized sgRNA vector DNA of interest was linearized using the *DraI* enzyme. To confirm the digestion, 2 µl of the reaction mixture was run on 1% analytical agarose gel. The *DraI*-digested DNA was used to produce sgRNA by *in vitro* transcription using T7 polymerase as shown in (Table 2.13). The reaction was incubated at 37 °C for overnight. At the end of the incubation, 1 µL of TURBO DNase (mMESSAGE mMACHINE, Cat no. AM1344) was added to digest the DNA template then the reaction was incubated at 37 °C for another 15 min. To stop the reaction, 15 µL of 5 M ammonium acetate, 100mM EDTA (mMESSAGE mMACHINE, Cat no. AM1344) was added and then add 35 µL of dH₂O. The RNA product was then purified using phenol:chloroform:isoamyl alcohol (25:24:1) extraction to remove all enzyme and most of the free nucleotides. The mix was then vortexed and centrifuged at 13.000 rpm for 10 minutes. Next, the upper aqueous phase was collected and transferred to a fresh Eppendorf tube. Subsequently, an equal volume of a chloroform:isoamyl alcohol (24:1) mix was added, vortexed and centrifuged at 13.000 rpm for 10 minutes. The upper phase was transferred to a fresh Eppendorf tube and 1 volume of 100% isopropanol was added.

The sample was vortexed and incubated for 30 minutes at -80 °C. Then, the sample was centrifuged for 10 minutes at 13.000 rpm to

pellet the RNA. Next, the supernatant was discarded, and the RNA pellet was washed with 200 μ L of 70% ethanol and then dried. The dried pellet of RNA was resuspended in 15 μ L of nuclease-free water. Finally, the concentration of sgRNA was measured by using Nanodrop spectrometer. To make sure that the sgRNA is not degraded, a mix of 2 μ L of the sgRNA, 5 μ L of RNA loading dye (mMESSAGE mMACHINE, Cat no. AM1344) and 3 μ L of dH₂O was incubated at 65°C for 10 minutes then placed in ice for 1 minute. The mixture was run on a 1.5% agarose gel (in 1.0x TAE buffer) and the product was kept at -80°C.

Table 2.13 The Mixture reaction to make *in vitro* transcription

Components	100 μl
5x transcription buffer	20 μ l
DTT	10 μ l
RNasin (Promega, cat. #N251A)	1 μ l
Unlabeled rNTPs (25mM) (NEB, cat. #N0446S)	2 μ l
RNA polymerase (T7) (NEB, cat. #M0251L)	2 μ l
DNA fragment (1-2 μ g)	X μ l
dH ₂ O	X μ l

2.2.5 Zebrafish methods

2.2.5.1 Maintenance of zebrafish and husbandry

Adult wild-type (WT) zebrafish (*Danio rerio*), transgenic and mutant lines were maintained in the Marine Biotech Recirculating System aquarium facility in the Biomedical Support Unit of the University of Nottingham. The fish were kept at a temperature of 28.5 °C, at dissolved oxygen levels of 7.4 ± 0.5 ppm and at a pH of 7.2 ± 0.5 . The fish were fed twice daily with brine shrimp. They were raised on a 14h light and 10 h dark cycle. Eggs were laid and fertilised after the light came on. Fertilised eggs were collected with a tea strainer from individual crosses and raised in Petri dishes in 1x E3 buffer containing methylene blue to prevent fungal growth (Westerfield, 1993). Under a light microscope, morphological features were used to stage embryos, according to Kimmel and colleagues (Kimmel et al., 1995). All embryos were kept at 28°C, and only phenotypically normal larvae were raised beyond the free-feeding stage and were taken to the fish aquarium at 5 dpf. All project experiments were performed under UK Home Office regulations and were accepted by the local ethical review board under project license numbers; 40/3457, 30/3378 and 30/3356 and personal license number; ID711FEEE.

2.2.5.2 Mutant and transgenic zebrafish lines used in this study

All transgenic fish lines were maintained as described (2.2.5.1). Carriers of the transgenic line *c-myb*^{qmc85} were maintained as heterozygous carriers (provided by Dr Martin Gering). *c-myb*^{t25127} (Soza-Ried et al., 2010) embryos were kept as heterozygous carriers. Once fish reached sexual maturity, they were crossed to each other to obtain 25% homozygous progeny which were identified by genotyping.

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

WISH is a technique widely used to determine the expression pattern of genes on whole-mount zebrafish embryos using 'antisense' digoxigenin (DIG)-labeled RNA probes. If embryos older than 24 hpf were required for WISH, they were treated with PTU (1.5 ml 0.3% PTU in 25 ml E3 buffer) to prevent the development of pigmentation. Embryos were collected at the desired time points. Embryos older than 18 hpf were dechorionated. Embryos were then fixed in 4% paraformaldehyde (PFA) in phosphate-buffered saline tween 20 (PBSTw). Tubes were incubated on a rocker overnight at 4°C (not more than 25 embryos per 1.5 Eppendorf tube). Fixed embryos were washed 3 times for 5 minutes at room temperature (RT) in PBSTw. Next, they were dehydrated and permeabilised slowly through 5 minutes washes in 25%, 50% and 75% Methanol in PBSTw. Finally, the embryos were washed twice in 100% methanol and stored in -

20°C at least overnight until use. All washes were done using 500 µl of solutions and shaking gently on the rocking table (Labnet GyroTwister or Grant-bio, PSM3D).

In situ experiments on zebrafish embryos were carried out over three consecutive days.

On the first day of WISH, embryos that had been stored in MeOH at -20°C were selected for an *in situ* and were rehydrated by washing 5 minutes in 75%, 50% and 25% Methanol in PBSTw followed by 3x5 minute PBSTw washes. Embryos older than 24 hpf were permeabilised further by treating them with 10 µg/ml of Proteinase K in PBSTw at RT according to the age of the embryos (Table 2.14).

Table 2.14 Incubation times in Proteinase K of embryos varied with the age of embryos

Age of embryo	Incubation times in Proteinase K
25-36 hpf	5-7 minutes
2 dpf	15-20 minutes
3 dpf	30-35 minutes
4 dpf	45 minutes
5 dpf	1 hour

Following permeabilisation with Proteinase K, embryos were washed twice for 5 minutes with 100% PBSTw to dispose of the proteinase K. Then the embryos were refixed in 4% PFA for 20 minutes at RT. After 20 minutes, PFA was discarded, and embryos were then rinsed 5 times for 5 minutes in PBSTw. The embryos were then prepared for pre-hybridisation by washing them once in 50% PBSTw and 50% Hybe^{+/+} for 5 minutes. The embryos were then introduced into 100% Hybe^{+/+} and incubated for 1 hour at 65°C. The Hybe^{+/+} was then

replaced by a hybridisation mix containing Hybe^{+/+} with the probe in a 1:200 dilution (Table 2.15). For probe hybridisation, the tubes were incubated in the probe solution overnight at 65°C in a heat block and laid on its side.

Table 2.15 List of WISH probes used in this project

In situ Probes	Reference #Gering lab database
<i>c-myb</i>	In-house (plasmid #77)
<i>Lyz</i>	In-house
<i>GFP</i>	In-house (plasmid #308)
<i>rag-1</i>	In-house (plasmid #21)
<i>srgn</i>	In-house (plasmid #478)
<i>mpx</i>	In-house

On the second day of WISH, the probe was recovered, and the embryos were pre-washed for 10 minutes each in 100%, 75%, 50%, 25% Hybe^{-/-} in 2xSSCTw at 65°C. Then, the embryos were washed in 100% 2xSSCTw for 10 minutes. Following a wash in 0.2xSSCTw, 4X15 minutes, the embryos were incubated at 65°C. Subsequently, and at RT the embryos were washed in 75%, 50%, 25% 0.2xSSC in MABTw, followed by a wash in 100% MABTw, 5 minutes per wash. The embryos were then blocked in MAB Block for at least 1 hour at RT. A solution of the alkaline phosphatase-conjugated anti-DIG antibody (Roche, cat. #11093274910) was diluted 1:5000 in MAB Block. Embryos were incubated in this antibody solution on a rocker overnight in the cold room.

On the third day of WISH, the antibody solution was removed, and any excess antibody was washed off in 8 x 15 min washes with

MABTw at RT. The embryos were then washed 3X in the developing buffer BCL3 for 5 minutes per wash. Next, the embryos were placed in the mix of BM Purple:BCL3 (1:1). BMP Purple is the substrate for alkaline phosphatase (Sigma, cat. #11442074001). Embryos were left on a rocker in the dark at RT until they developed colour. To stop the reaction, embryos were washed in PBSTw containing 20mM EDTA 3X for 5 minutes. The embryos were then fixed in 4% PFA at RT for 20 minutes or at 4°C overnight followed by washing in PBSTw 3 times for 5 minutes. The embryos were then gradually washed into 80% glycerol, by washing once each for 5 minutes with 30%, 50% and finally into 80% glycerol in PBSTw. Embryos were stored in 80% glycerol at 4°C indefinitely.

2.2.5.4 Adult Zebrafish Kidney FACS Analysis

2.2.5.4.1 Whole kidney marrow (WKM) cells isolation

Adult fish to be analysed by FACS (Fluorescence-activated cell sorting) were selected and then euthanised using an overdose of AquaSed as recommended by the manufacturer (one measuring pump (1.5 ml) per litre of water). The kidneys of individual fish were dissected as described by (Gerlach et al., 2011) and placed into 2 ml of cold 1xHBSS Buffer on ice. Next, the cells were flushed out of the kidney by pipetting up and down several times for approximately 3 minutes on ice using a P1000 pipette with a blue tip. Then, 5 ml of FCS/PBS were added. The suspension was strained into 50 ml falcon tubes placed on ice using 40 µm cell strainers (Scientific Laboratory

Supplies, cat. #352340). The tube containing the flushed blood cells was spun at 4°C for 5 minutes at 1500 rpm (Heraeus Multifuge 3SR Plus cat. #75004371). Next, the supernatant was discarded, and the cell pellet was resuspended in 300 µl of FCS/PBS. The kidney cell solution was transferred into 15 ml tubes, which were directly placed on ice and transported to the FACS sorter machine. The cells were analysed on a Beckman Coulter Astrios EQ FACS sorter using the Kaluza software.

2.2.5.4.2 Fluorescence activated cell sorting (FACS)

FACS analysis allows fluorescently labelled cells to be sorted based on their light scattering and fluorescence characteristics. This is done by passing fluorescently labelled cells expressing, for example, an enhanced green fluorescent protein (EGFP) or a red fluorescent protein (dsRed) in a fluid stream through a laser beam. The laser beam (a) is scattered by the cells and (b) excites the fluorescence of the fluorophores. Physical properties of the cell, such as size (forward scatter (FSC)) and density or granularity (side scatter (SSC)), can be used in order to resolve the cell populations (Brown and Wittwer, 2000). Once past the laser, the fluid stream is broken into droplets that are set to each contain a single cell. Application of a small charge allows the subsequent deflection of the drop, allowing the sorting of cells that fall into particular gates that the user defines on the flow cytometer. All gates that were set with the help of the wt non-transgenic control cell samples. The same settings were

subsequently applied to the WKM samples of the transgenic fish to make all data sets comparable. First, in order to completely exclude the dead auto-fluorescing cells from the analysis, the WKM cell preparation (from above section 2.2.5.4.1) was treated with SYTOX (red 3:10) (5 μ M stock). FSC area versus SYTOX red height were used for excluding dead cells and debris (Figure 2.7A). Then, to exclude cell debris and restrict our analysis to live cells a region was manually gated on the FSC/SSC plot of the wild-type WKM sample. To show only live cell population, the wt sample was then gated for live cells and a region was selected to exclude for dead cells (Figure 2.7B). Cell doublets were excluded by comparing SSC height versus SSC area values for every event (Figure 2.7C). Next, single cells were gated for GFP+ cells and to remove cells that displayed auto-fluorescence in multiple channels (Figure 2.7D). Because different blood cell populations of the zebrafish kidney marrow have been shown to display different FSC versus SSC characteristics (Traver et al., 2003b) light scatter characteristics allow us to identify specific cell types of interest such as myelomonocyte, erythrocyte and progenitor cells (Figure 2.7E). Samples were analysed on a Beckman Coulter Astrios EQ operated by Nicola Croxall.

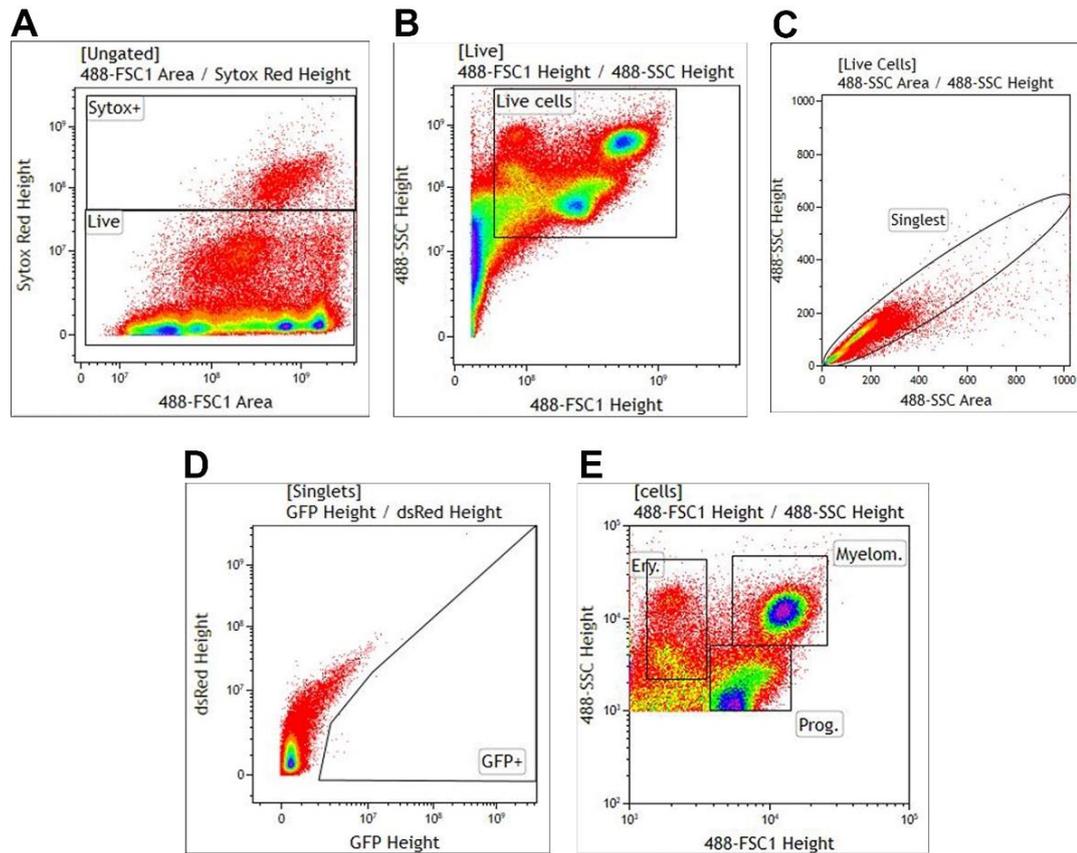


Figure 2.7 Gating strategy for the flow cytometric analysis of adult zebrafish whole kidney marrow (WKM) cells

Gates were set on the FSC area/SYTOX red height, FSC/SSC, SSC height/SSC area and GFP height/ dsRed height plots of the WT sample. (A) The WT sample was plotted on the channel FSC area versus SYTOX red height to exclude dead cells and debris. (B) The FSC/SSC plot allows the exclusion of cell debris and the identification of the live cell population (A). (C) SSC height/SSC area plot allow the detection of singlets and the exclusion of cell doublets. (D) GFP height/ dsRed height plot to select GFP+ cells and exclude autofluorescent cells. (E) The plot shows the light scatter characteristics of all live cells. GFP fluorescence is excited with a 488 nm laser. The emitted light is detected with the help of a 513/26 nm band pass filter. Red fluorescence is excited with a 561 nm laser and detected using a 579/16 nm band pass filter. SYTOX Red is excited at 640 nm and detected with an emission filter at 670/30 nm. The light scatter analysis was done using the 488 nm laser.

2.2.5.5 Microinjection needle preparation

The micropipette puller (Sutter Instrument, Novato, CA) was used to prepare the microinjection needles by pulling the glass capillaries (HARVARD, GC 100F-15, cat. #30-0020) with an internal filament. Because the needle puller settings change with every new puller filament and can change during the lifetime of an existing filament, the reader is advised to consult the manufacturer's instructions to adjust the puller settings. The injection needles were backloaded using Microloader pipette tips (Eppendorf, cat. #E270445J). Needle tips were broken (directly before use) with a clean tweezer to enable the delivery of the injected solution. Drop sizes were adjusted on the microscope with the help of a graticule placed in one of its eyepieces.

2.2.5.6 Microinjection

In total volume of 5 μ l of the injection's mixture was prepared by mixing of 1.5 μ l Cas9 protein (BioLabs Kit, cat. #M0646T), 0.5 μ l buffer (1x Cas9 Nuclease Reaction Buffer) and 1.5 μ l mini ruby (33mg/ml were used diluted 1:6) then 80-90 ng of the pair of prepared sgRNAs were added. The mixture then was incubated for 5 minutes at 37°C to reconstitute active Cas9 protein and sgRNAs (Burger et al., 2016). Injections were performed into one-cell stage embryos. For this purpose, the collected fertilised embryos were lined up against a microscope slide in a petri dish lid and kept moist with E3 buffer. Embryos were orientated so that their animal poles faced the microinjection needle. Volumes of 1 μ l of the solution were

injected per embryo. All injections were performed on the stage of a stereomicroscope equipped with the microinjector (picospritzer® III, Parker, Ohio, USA).

2.2.6 Imaging

To study the expression of EGFP in zebrafish embryos in the transgenic line, the embryos were anaesthetised by adding 650 µl of MS222 (4g/L) to 25 ml of E3 buffer. Embryos were moved onto a 1% agarose surface in E3 buffer. A tiny hole was introduced into the agarose surface, and the embryo was placed with its yolk into this hole so that the rest of the body comes to lie in a horizontal plane, optimal for visualisation of trunk and tail structures in a single focal plane. Live transgenic embryos were imaged and EGFP expression was observed using a fluorescent dissection microscope (Nikon, model. SMZ1500, Japan) equipped with 100W super high-pressure mercury lamp (Nikon, 100 W, C-SHG1, Japan) and FITC, TRITC and Cyan filter sets to detect fluorescence. All bright images of embryos were taken using a digital camera (Nikon DS-5MC, Japan) attached to a dissecting microscope. Images of *in situ* stained embryos were taken with the same digital camera. Images were acquired with Nikon's ATC-2U software. Images were edited and assembled in Adobe Photoshop Software (version CC 20.0.4).

3 Results

3.1 Single-cell RNA Sequencing (scRNA-Seq) data suggests that all GFP+ cells isolated from whole kidney marrow (WKM) of adult Tg(*c-myb-zRE1:PF5:egfp*)^{qmc85} zebrafish are neutrophilic granulocytes.

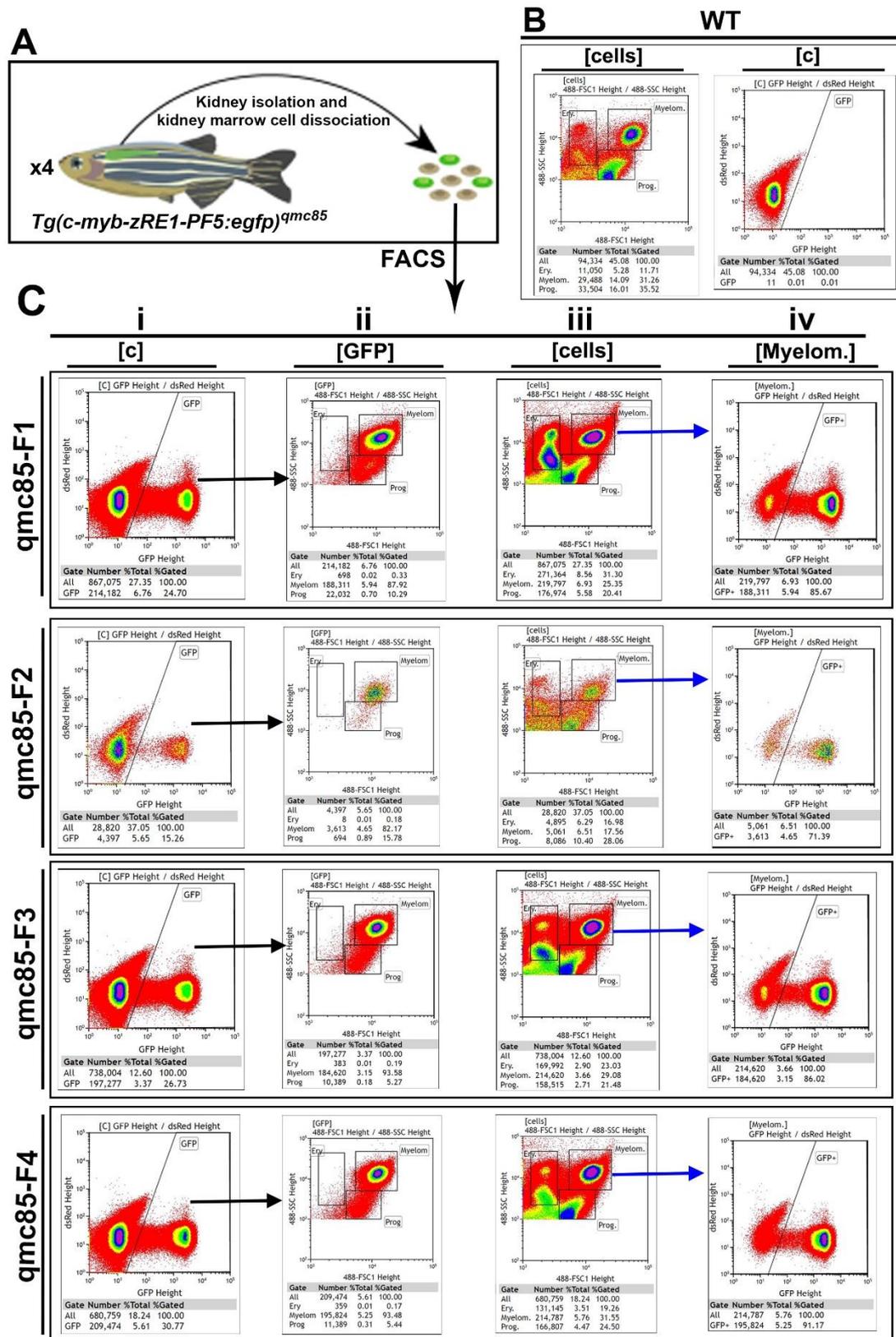
The transgenic line *qmc85* which carries the *gfp* reporter gene downstream of zRE1 and promoter PF5 displays GFP expression in blood cells in the CHT at 3 dpf (Hsu, 2010). Flow cytometric analysis also revealed GFP expression in cells of the adult kidney. Light scatter analysis suggest that these GFP+ cells included myeloid and progenitor cells (Savage, 2012). The exact nature of the GFP+ cells remained to be determined. Here, single-cell RNA Sequencing (scRNA-Seq) was used to gain a deeper insight into the nature of the GFP+ cells.

3.1.1 Isolation of single GFP⁺ cells from the adult WKM of four Tg(*c-myb-zRE1:PF5:egfp*)^{qmc85} fish

In order to sort the cells that express the GFP, Fluorescence-Activated Cell Sorting (FACS) was performed on WKM cells of adult fish of the *qmc85* line. For this purpose, four *qmc85* transgenic and one WT control fish were euthanised and their kidneys dissected (Figure 3.1A). Blood cells were flushed out of the kidneys and were analyzed on a Beckman Coulter Astrios EQ fluorescence-activated cell sorter. Sytox red was employed to exclude dead cells. Gates that

were set with the help of the non-transgenic WT cell sample (Figure 3.1B) were then applied to the WKM samples of the *qmc85* transgenic fish. Light scatter characteristics were examined to distinguish cells from cell debris and to tell single cells from cell doublets (the gating strategy is illustrated in (2.2.5.4.2 and Figure 2.7). Green fluorescent single cells were then sorted from all four *qmc85* transgenic WKM cell samples (Figure 3.1Ci-iv). Figure 3.1Ci show that the enrichment percentage of the GFP+ cells from the four fish was 24.70%, 15.26%, 26.73% and 30.77%, respectively. Forward versus side scatter analyses revealed that the vast majority of these GFP+ cells fell into the myelomonocyte gate (87.92%, 82.17%, 93.58% and 93.48%) of the wt whole KM light scatter profile (Figure 3.1Cii) as defined by Traver and colleagues (Traver et al., 2003a). These cells were characterised by large cell size and high granularity. In Figure 3.1Ciii, forward and side scatter characteristics were performed to know the different distribution patterns of the all cells. The myelomonocytic gated cells were then analysed by green versus red fluorescence (Figure 3.1Civ). This shows that the enrichment percentage of the GFP+ cells within the myelomonocyte from four fish was 85.67%, 71.39%, 86.02% and 91.17%, respectively. The enriched GFP+ cells of all the four fish were then combined and subsequently purified, as shown in Figure 3.1Di. The percentage of the enriched, combined and purified GFP+ cells was 83.57%. Among the pooled and purified GFP+ cells, the

percentage of the myelomonocytic gated cells was 90.10% (Figure 3.1Dii). A small fraction (6.31%) of GFP+ cells displayed the light scatter characteristics of progenitor cells. The GFP+ cells within the myelomonocyte cells after pooled and purified from the four fish was 98.67%, as shown in Figure 3.1Div. Figure 3.1E shows the dot plot of the percentage of the gated myelomonocyte and GFP+ cells before and after enrichment in each of the four *qmc85* zebrafish transgenic biological replicate. Overall, approximately 1.5 million single GFP+ cells were isolated from the adult WKM of four *Tg(c-myb-zRE1:PF5:egfp)^{qmc85}* fish. The cells were stained with Trypan blue. Live cells (>82%) are unstained, while dead cells are blue and LUNA recognise that. Approximately 6000 cells of the sample were then used in the scRNA-Seq analysis (Table 2.2).



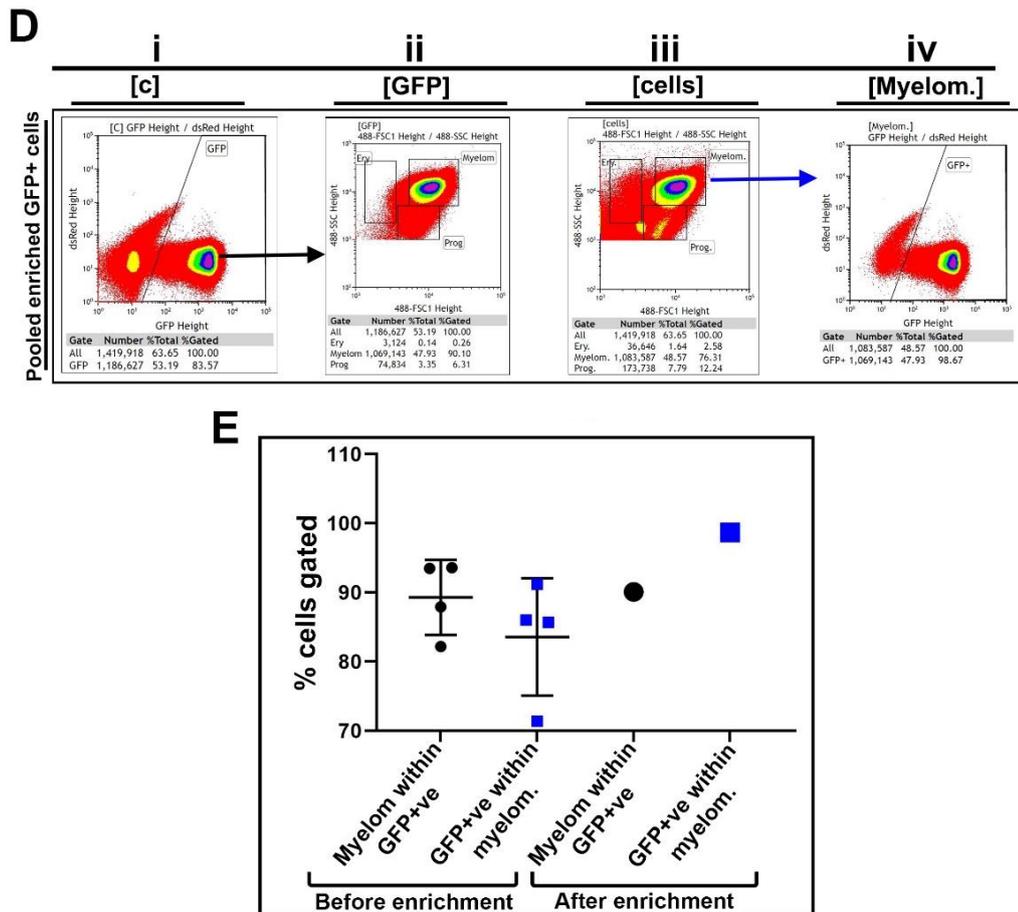


Figure 3.1 Isolation of single GFP⁺ cells from the adult kidneys of four Tg(*c-myb-zRE1:PF5:egfp*)^{qmc85} fish

(A) The scheme shows the preparation of the kidney from an adult zebrafish. Kidney marrow (KM) cells were flushed out of the kidneys. FACS was used to sort single GFP⁺ cells from each of the four kidneys. **(B)** Flow cytometric analysis of adult wt fish (non-transgenic). The right panel shows the light scatter characteristics of all cells for the wt control fish. The left panel shows a density plot that displays green versus red fluorescence in the KM cells of wt fish. **(Ci-iv)** Flow cytometric analysis of adult kidneys of four *qmc85* transgenic fish. **(Ci)** Panels show density plots that display green versus red fluorescence as detected in the KM cells of each of the four fish analysed. **(Cii)** Panels show the light scatter characteristics of all cells for the GFP⁺ cells identified in the four transgenic fish. **(Ciii)** Panels show the light scatter characteristics of all cells in the four transgenic fish. **(Civ)** Panels show density plots that display green versus red fluorescence detected in the myelomonocytic gated cells of the KM cells of each of the four fish analysed. Following an initial enrichment of the GFP⁺ cells of each four transgenic fish, the GFP⁺ cells from all four fish were combined and subsequently purified. **(Di-iv)** Plots reveal the extent of enrichment achieved in the purification step. **(Di)** Panel shows density plot that displays green versus red fluorescence as detected in the KM cells of all pooled four transgenic fish analysed. **(Dii)** Panel shows the light scatter characteristics of all cells for the GFP⁺ cells identified in the pooled four transgenic fish. **(Diii)** Panel shows the light scatter characteristics of all cells in the pooled four transgenic fish. **(Div)** Plot of pooled purified GFP⁺

cells reveals that most of the myelomonocytes in *qmc85* transgenic fish are GFP+. **(E)** Dot plot graph showing the percentage of gated myelomonocytes and GFP+ cells in each enriched *qmc85* zebrafish transgenic sample before and after enrichment. The percentage of the combined and purified myelomonocytic cells within the GFP+ cells was over 90% (big black dot). The percentage of the combined and purified GFP+ cells within the myelomonocytes was over 98% (big blue square). **Shapes:** black dots represent wt, and blue squares represent *qmc85* (1, 2, 3 and 4). The overall number of cells and those in the particular gates and the corresponding percentages are shown under each panel. Gates from left to right: [c]= time AND live AND cells; [GFP]= time AND live AND cells AND GFP. [Cells]= time AND live AND cells; [Myelom]= time AND live AND cells AND Myelom. The cell sorting was performed on a Beckman Coulter Astrios EQ operated by Nicola Croxall.

3.1.2 All single GFP⁺ kidney marrow cells of Tg(*c-myb-zRE1:PF5:egfp*)^{qmc85} fish express neutrophil genes

The 10x Genomics Chromium system was used to determine the gene expression profile of the single GFP⁺ cells that were collected from the adult WKM of the four Tg(*c-myb-zRE1:PF5:egfp*)^{qmc85} fish. The cells were separated and lysed, and the cDNA was prepared by marking with cell-specific barcodes and transcripts were labelled with unique molecular identifiers (UMIs). The cDNAs were then pooled and amplified, and libraries were generated and paired-end sequenced on Illumina sequencer. The sequence analysis was performed using the Cell Ranger pipeline. More than 191 million sequence reads were collected, of which over 90% mapped to the zebrafish genome (GRCz11). The number of UMIs per barcode was determined, and all barcodes that displayed at least 0.1 times as many different UMIs as the top 1 percentile of barcodes were subsequently considered cells (Figure 3.2). This reduced the number of analysed barcodes to 2246 cells. For these cells, a mean number of 85,099 reads were collected, and a median number of 937 genes were found to be represented (Figure 3.2).

The gene expression profiles of the 2246 cells that were analysed through the Cell Ranger bioinformatics pipeline were visualised in the Loupe Cell Browser (version 3.1.0). The 10x Genomics' Loupe Cell browser displays the data in the t-SNE plot that is a two-dimensional representation of the multi-dimensional data set. In

Loup Cell browser, t-SNE plot reveals that most cells fall into a large elongated cell cloud. There are also a few small additional cell clusters. Graph-based cluster analysis, which uses the Louvain clustering algorithm, identifies seven subclusters within the large cloud (Figure 3.3). Projecting overall gene transcript accumulation onto the t-SNE plot reveals the position of those cells that had the highest and lowest UMI counts (Figure 3.4).

To determine whether the large cloud and clusters represent different blood cell lineages, the gene expression levels of the individual lineage-specific marker was projected onto the t-SNE plot. These results reveal that all of the 2246 cells (Figure 3.5A) expressed the neutrophil granulocyte gene lysozyme. None of the genes that are specific for other haematopoietic cell types, such as macrophage (*mfap4*, *mpeg1.1*), red blood (*gata1*) and T cells (*rag1*) were expressed (Figure 3.5B-E). Since other neutrophil markers, like *lect2l*, *npsn*, *srgn*, *cpa5* and *abcc13* were also widely expressed in our cell cohort (Figure 3.5F-J), it can be concluded that all of the *qmc85*:GFP⁺ belong to the neutrophil cell lineage.

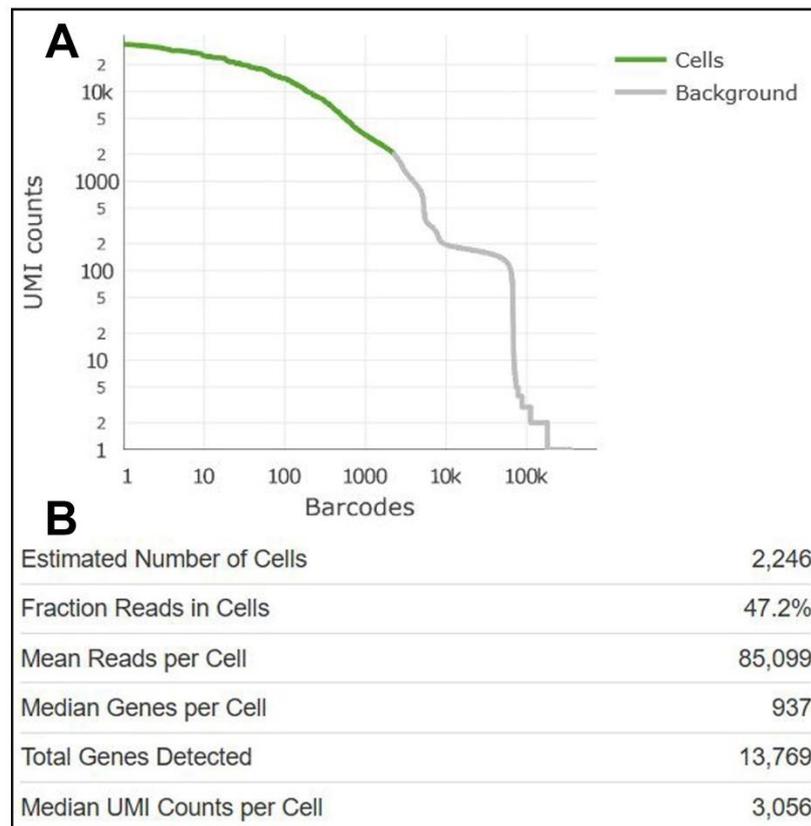


Figure 3.2 Calling cell barcodes.

(A) The plot illustrates the number of UMI counts mapped to each barcode. Barcodes determine cell-associated based on their UMI counts or by their RNA profiles. UMI counts are shown on the y-axis ranging from 0 to over 20k in log scale. On the x-axis, barcodes are ranked in descending order based on the number of UMIs associated with them. Barcodes attributed to cells are highlighted in green. The table shown in **(B)** displays the number of cells detected, the fraction of reads in cells, the mean number of reads per cell, the median number of genes detected per cell, the total number of genes detected overall and the median UMI counts per cell.

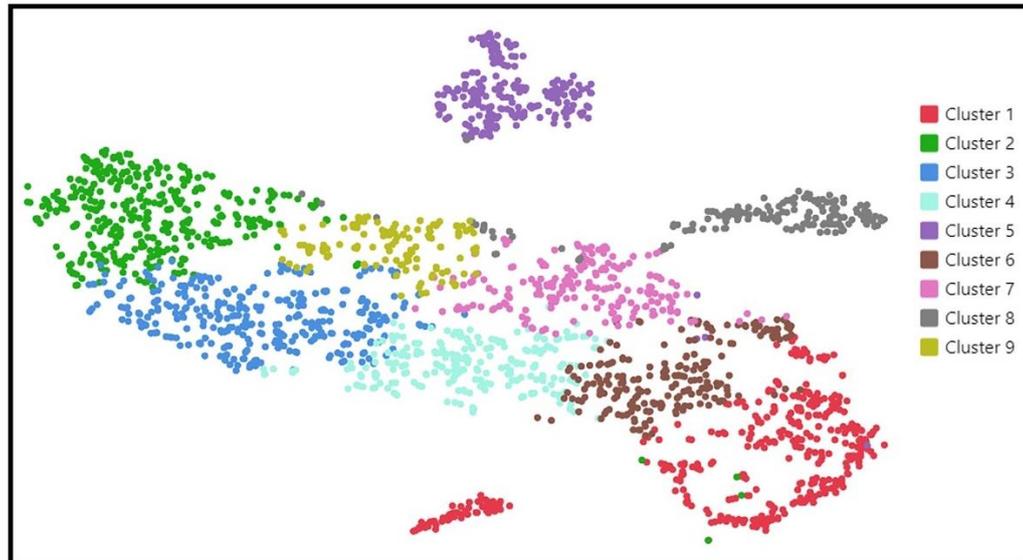


Figure 3.3 Single-cell RNA Sequencing of zebrafish whole kidney marrow neutrophils of $Tg(c-myb-zRE1:PF5:egfp)^{qmc85}$ identified nine neutrophil cell clusters.

Two-dimensional visualization (t-SNE) plot of 2246 kidney marrow neutrophils of $Tg(c-myb-zRE1:PF5:egfp)^{qmc85}$ shows nine distinct clusters across the Loupe Cell Browser visualization. Coloured squares highlight specific cell clusters.

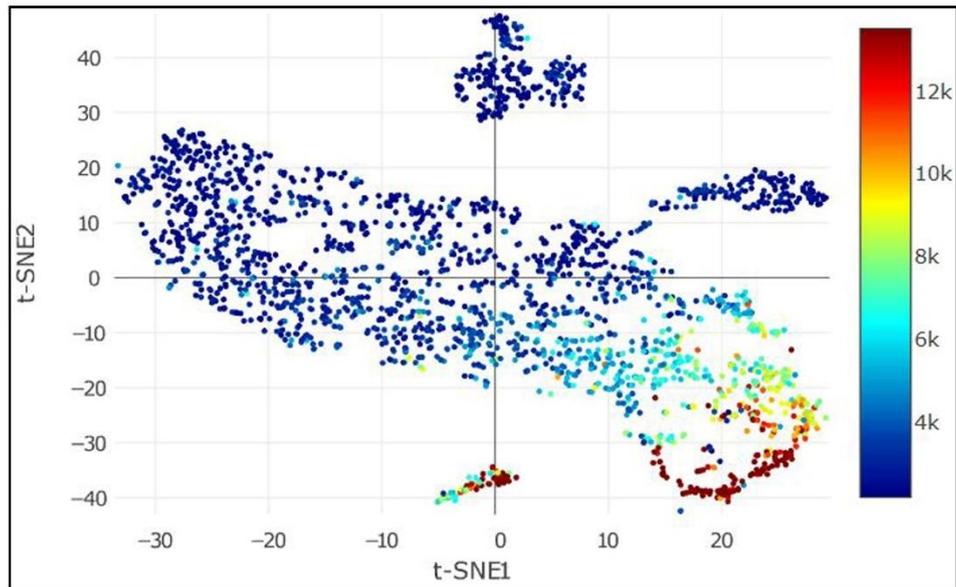


Figure 3.4 t-SNE visualization of single GFP+ kidney marrow cells of $Tg(c-myb-zRE1:PF5:egfp)^{qmc85}$ coloured by the number of UMI counts per cell

The plot shows the RNA content of the single GFP+ kidney marrow cells of $Tg(c-myb-zRE1:PF5:egfp)^{qmc85}$. The heatmap reflects the number of UMIs associated with the cells.

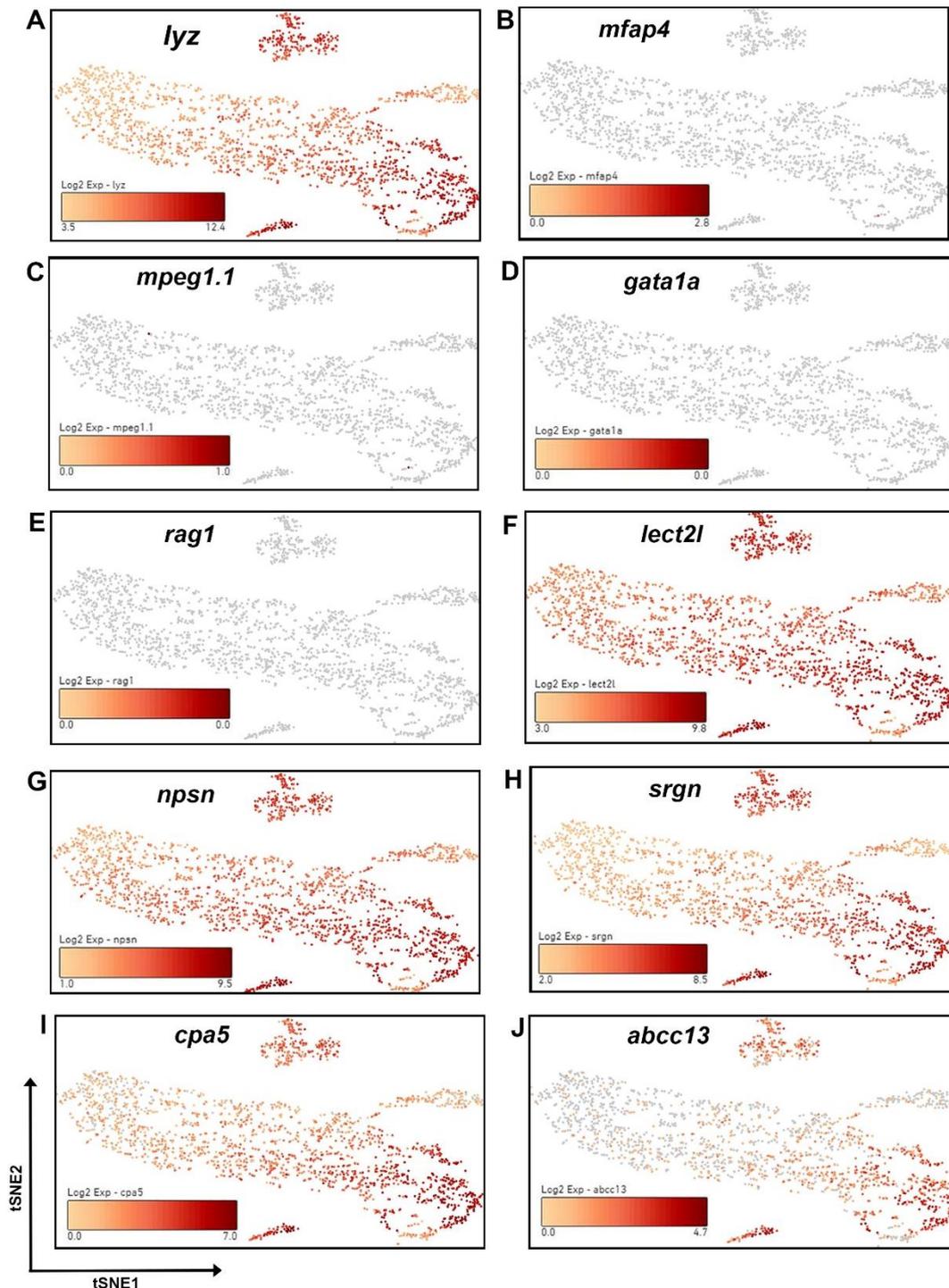


Figure 3.5 Single-cell RNA Sequencing reveals that all GFP⁺ cells that were isolated from the kidney marrow of *Tg(c-myb-zRE1:PF5:egfp)^{qmc85}* fish express neutrophil genes

The single-cell RNA-Seq data for 2246 cells were analyzed by a clustering algorithm, the t-stochastic neighbour embedding (t-SNE) algorithm, that clusters cells based on their gene expression profile and provides a two-dimensional representation of the multi-dimensional data set. A heat map representation of log₂ values of the number of transcripts (UMI) counts for individual genes is projected for every single cell depicted in the t-SNE plot.

Panel **(A)** shows the expression of the *lyz* gene. The panels show in **(B-J)** display the expression of a number of selected genes. The selected genes were typically expressed in blood cells such as macrophage (*mfap4*, *mpeg1.1*), red blood (*gata1a*), T cells (*rag1*) and neutrophil (*lect2l*, *npsn*, *srgn*, *cpa5* and *abcc13*).

Summing up, all UMI counts in the entire cell cohort revealed that five neutrophil-specific genes were among the top ten genes with the highest transcript counts (Table 3.1). Another three genes encoded actin or actin-binding proteins, the expression of which is most likely related to the migratory behaviour of neutrophils. Another seven neutrophil marker genes had decent overall transcript counts while erythroid, macrophage, T-cell and thrombocyte genes displayed hardly any. There is a clear correlation between the level of *Lyz* expression and the level of expression of other neutrophil markers, as shown in Table 3.1. Interestingly, our data revealed that there is a highly positive correlation between *Lyz* and some of the neutrophil markers, for instance, *lect2l* and *npsn*, suggesting that the neutrophil markers were widely expressed in our *qmc85*:GFP+ cell cohort of WKM sample (Figure 3.6). Our cells also express other genes that are specific for neutrophil lineage, such as *cxcr4b*, *rac2*, *wasb*, *cebpb*, *cfl1*, *illr4*, *mpx* and *ncf1* (Appendix figure 6.2). Altogether these data suggest that the GFP-positive cells isolated from the kidney marrow of the Tg(*c-myb*-zRE1:PF5:*egfp*)^{*qmc85*} fish express neutrophil genes and most likely represent cells of the neutrophil granulocyte lineage.

Table 3.1 Total transcript (UMI) counts in the entire cohort of 2246 GFP⁺ kidney marrow cells isolated from adult Tg(*c-myb-zRE1:PF5:egfp*)^{qmc85} fish

Transcript (UMI) counts for every gene were added up over the entire cohort of 2246 *qmc85*:GFP⁺ kidney marrow cells. Genes were ranked in descending order of their overall UMI counts. The table shows the top 10 genes, a set of 7 other neutrophil markers, as well as a set of genes that are typically expressed in other blood cell types. The boxes coloured with yellow depict the genes whose expression correlates with that of *lyz* (linear regression analysis: $r^2 \geq 0.70$).

Rank	Gene ID	Gene name	total umi	Protein	r ²
1	ENSDARG00000057789	lyz	853850	Lysozyme	1.00
2	ENSDARG00000037870	actb2	465598	Actin 2b	0.29
3	ENSDARG00000033227	lect2l	313353	Chemotaxin 2	0.70
4	ENSDARG00000099970	CR383676.1	244445	lncRNA	0.12
5	ENSDARG00000077777	tmsb4x	209597	Thymosin	0.49
6	ENSDARG00000010423	npsn	143801	Nephrosin	0.84
7	ENSDARG00000088091	pfn1	134905	Profilin	0.46
8	ENSDARG00000101479	BX908782.2	123247	CD59 Glycopr.-like	0.29
9	ENSDARG00000092807	si:dkey-151g10.6	111472	Ribosomal Pr.	0.19
10	ENSDARG00000077069	srgn	101973	Serglycin	0.83
16	ENSDARG00000042725	cebpb	57268	Cebpb TF	0.32
86	ENSDARG00000021339	cpa5	28279	Carboxypeptidase A5	0.79
121	ENSDARG00000019521	mpx	17211	Myeloperoxidase	0.52
146	ENSDARG00000000767	spi1b	11588	Pu.1 TF	0.13
293	ENSDARG00000036074	cebpa	4241	Cebpa TF	0.28
338	ENSDARG00000062519	abcc13	3583	Abcc13	0.78
1322	ENSDARG00000053666	myb	610	Myb TF	0.17
9199	ENSDARG00000056407	irf8	10	Irf8 TF	<0.001
9373	ENSDARG00000087646	runx1	9	Runx1 TF	<0.001
9565	ENSDARG00000090783	mfap4	8	Microfibril-assoc. pr. 4	<0.001
10272	ENSDARG00000019930	scl/tal1	5	Scf/Tal1 TF	<0.001
11529	ENSDARG00000055290	mpeg1.1	2	Macrophage-expr. 1.1	<0.001
13770	ENSDARG00000013477	gata1a	0	Gata1a TF	<0.001
13770	ENSDARG00000113599	hbbe1.1	0	β -embryonic Globin	<0.001
13770	ENSDARG00000052122	rag1	0	Rag1 Recombinase	<0.001
13770	ENSDARG00000102986	csf1ra	0	CSF1-Receptor a	<0.001
13770	ENSDARG00000018687	itga2b - cd41	0	Integrin a2b - CD41	<0.001

Genes that are known to be expressed in:

Neutrophils ■ Macrophages ■ Erythrocytes ■ Lymphocytes ■ Thrombocytes ■

Genes that encode actins or actin-interacting proteins ■

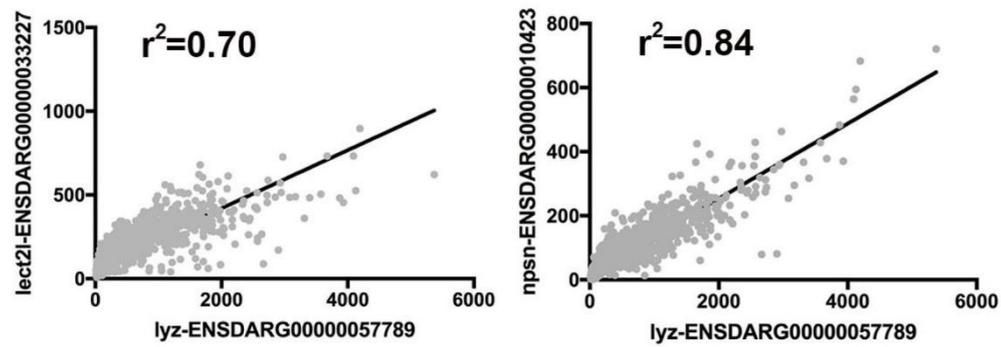


Figure 3.6 The correlation of selection of neutrophil markers

A strong positive correlation between *Lyz* and *lect2I*, and *Lyz* and *npsn* genes expressions in the entire cohort of 2246 *qmc85:GFP+* kidney marrow cells.

3.1.3 The single-cell RNA Sequencing (scRNA-Seq) analysis of *qmc85*:GFP⁺ kidney marrow cells reveals distinct neutrophil maturation stages

In the previous section, the FACS reveal that the majority of *qmc85* GFP positive cells are found in the myelomonocyte gate (Figure 3.1). In our lab, it has been found that the *qmc85* GFP⁺ cells, which were sorted by FACS, cytopun and Giemsa stained, are neutrophils in different maturation stages (segmented and not segmented nucleus) using a light microscope (data not published). In our scRNA-Seq study, the analysis of *qmc85*:GFP⁺ kidney marrow cells showed that all 2246 cells were neutrophils (Figure 3.5). Further analysis revealed that these cells were grouped into a large cloud (seven clusters) and other small clusters. These structured clusters could mean that the cells are in different maturation stages. Here, because the vast majority of our cells (1816) are within the large cloud (Figure 3.4), we focused on this cloud. Within this cloud, different maturation stages could occur in all the subclusters. The two extreme end clusters, one and two, were selected for further analysis to understand the possible maturation development among all clusters. To do this, we looked at the gene expression levels for the different maturation stages of the cells in clusters one and two, then compared with similar previous studies on different maturation stages. In zebrafish, there is a lack of antibodies against some of the surface markers (Moore et al., 2016). This has made it difficult to

sort the different neutrophil maturation stages. There are no data on the gene expression profiles of neutrophils in different maturation stages in zebrafish to compare with our scRNA-Seq findings. In contrast, several mouse and human studies have used antibodies to separate cells into different neutrophil maturation stages based on the expression of neutrophil surface markers (Dinh et al., 2020, Evrard et al., 2018, Grassi et al., 2018, Zhu et al., 2018). We used the most recent study on the human bone marrow to compare with our scRNA-Seq findings (Dinh et al., 2020). In their study, human bone marrow neutrophils were grouped into four different subsets based on their different level of maturation. To group these subsets, they used the surface marker expression, morphology, proliferating assays (use to characterise neutrophil progenitor) and bulk RNA-sequencing analyses. The earliest neutrophil progenitor (ePreN) was grouped as the first subset while the most matured neutrophil (Neuts) as the last subset (Dinh et al., 2020). They found out that the neutrophil maturation was gradually developing along with the identified subsets (Dinh et al., 2020). The RNA-seq data of the ePreN and Neuts from Dinh and colleagues' study were re-analysed (2.2.1.3) in order to compare them with our scRNA-Seq data from clusters one and two, respectively (Figure 3.4). The human RNA-seq data of each subset were then imported into the SeqMonk program (v1.47.1) using Homo sapiens GRCH38_v97. In the SeqMonk program, these data were then organised for visualisation and

analysis. Each gene subsets of the RNA-seq data then were mapped to an MA plot (Figure 3.7A). In SeqMonk, DESeq analysis ($p < 0.05$) was conducted to identify genes that were differentially expressed between the ePreN and Neuts subsets data. The genes that were displayed in the $+\text{Log}_2$ difference of the MA plot were more highly expressed in ePreN (immature), while those at the $-\text{Log}_2$ difference of the MA plot were more highly expressed in Neuts (mature) (Figure 3.7B).

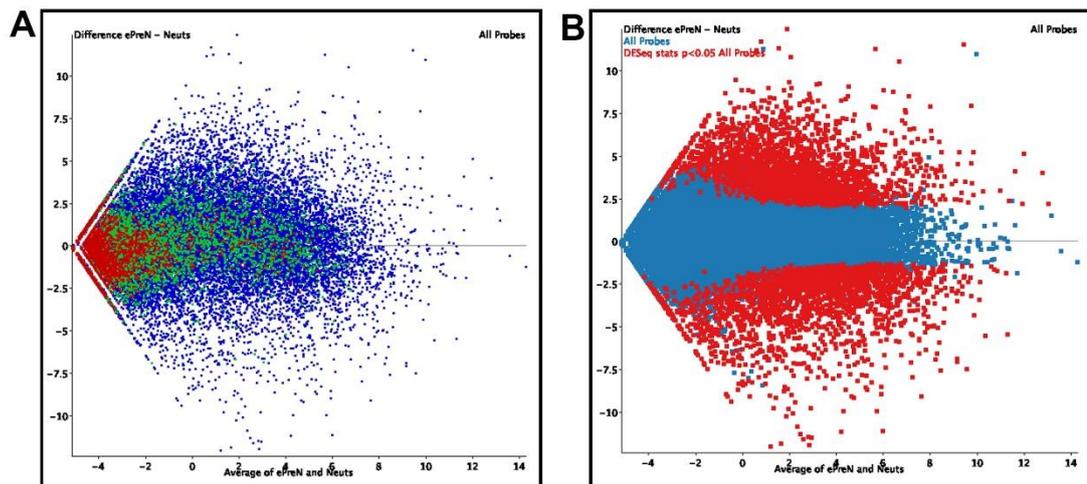


Figure 3.7 MA plots comparing human early neutrophil progenitors (ePreN) and neutrophils (Neuts) (Dinh et al., 2020).

The MA plots represent each gene with a dot. **(A)** The M (Y-axis) represents the log₂ fold change between ePreN and Neuts subsets, and A (X-axis) represents the average expression of the overall samples. Plot **(B)** was obtained from the plot **(A)**. **(B)** The plot illustrates the differentially expressed genes ($p < 0.05$) within the two groups: ePreN (top) and Neuts (bottom), of gene expression counts. Red dots represent differential expressed genes, while blue dots represent non-differentially expressed genes ($p < 0.05$). The black horizontal line at zero provides genes with similar expression values in ePreN and Neuts groups.

To compare Dinh and colleagues' data subsets (ePreN and Neuts) with our clusters one and two data, human orthologues of both clusters-specific zebrafish genes of scRNA-Seq data were used. To achieve this, the Loupe browser of our zebrafish scRNA-seq data was used to select the set of genes that were identified to be differentially expressed in cells of cluster one relative to cells in cluster two ($p < 0.05$) (Appendix table 6.1). The same process was repeated by selecting the genes of cluster two relative to those in cluster one (Appendix table 6.2). For both sets, the human orthologues of both cluster-specific zebrafish genes were determined in BioMart database (ENSEMBL 102) (Appendix table 6.3 and Appendix table 6.4). The set of genes (72) of our human orthologues of cluster one-specific zebrafish genes were then visualised in SeqMonk. Sixty-seven (67) (Appendix table 6.5) out of 72 genes were found in the data set of Dinh and colleagues' human neutrophil genes (Figure 3.8A). From the 67 genes, 40 (~64%) were differentially expressed within the ePreN (immature) and Neuts (mature) of Dinh and colleagues' data (Appendix table 6.6), while 24 (~36%) were non-differentially expressed genes. Interestingly, most of the differentially expressed genes (40 out of 43) were significantly upregulated in human ePreN relative to Neuts, and three were found in the human Neuts group of Dinh and colleagues' data (Figure 3.8B). The majority of these genes were genes encoding ribosomal proteins (Appendix table 6.6). In the same manner, the set of genes

(205) of our human orthologues of cluster two-specific zebrafish genes were visualised in the SeqMonk. One hundred and ninety-five (195) (Appendix table 6.7) out of 205 genes were found in the data set of the human neutrophil genes (Figure 3.8A). From the 195 genes, 75 (~39%) were differentially expressed within the Neuts (mature) and ePreN (immature) of Dinh and colleagues' data (Appendix table 6.8), while 120 (~62%) were non-differential expressed genes. Remarkably, most of the differentially expressed genes (58 out of 75) were significantly upregulated in human Neuts relative to ePreN, and 17 were found in the human ePreN group of Dinh and colleagues' data (Appendix table 6.8) (Figure 3.8C). Several genes were found in cluster two that were specific for mature neutrophils. Overall, this result suggests that the comparison of the genes of clusters one and two with that of Dinh and colleagues' data (ePreN and Neuts) reveals that the different cell clusters of our scRNA-Seq could represent different stages of neutrophil maturation (Figure 3.8D & E). Together, our results highlight that clusters one and two within the large cloud represent the immature and mature neutrophils, respectively. This could mean that there is a gradual maturation across the clusters from extreme cluster one to extreme cluster two.

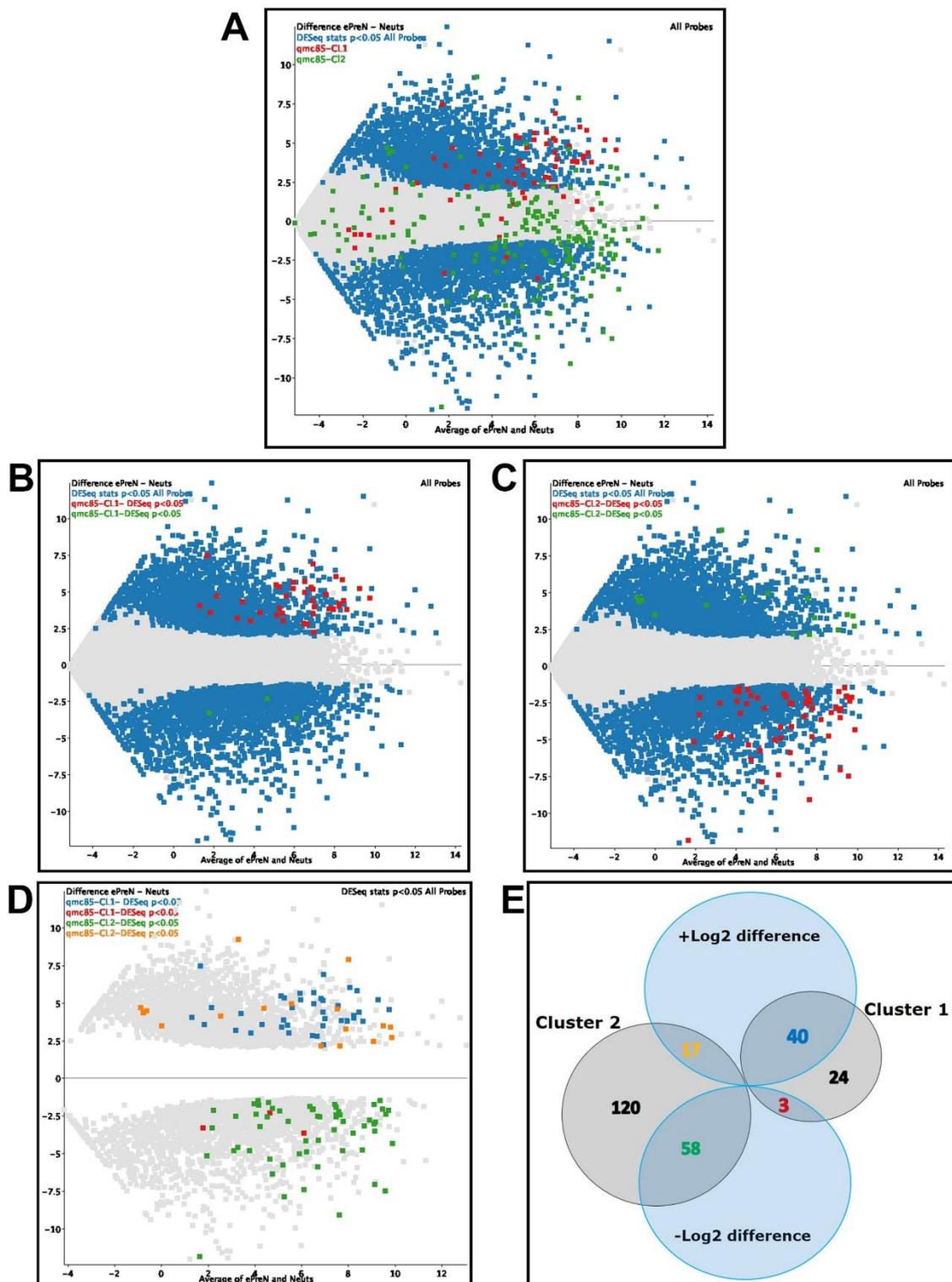


Figure 3.8 Comparison of Dinh and colleagues' data with our human orthologues of zebrafish clusters one and two genes.

The (A, B, C and D) MA plots were generated by comparing MA plot (figure 3.7B) of Dinh and colleagues' data with our human orthologues of zebrafish clusters one and two genes of scRNA-seq of zebrafish whole kidney marrow neutrophils of $Tg(c-myb-zRE1:PF5:egfp)^{qmc85}$. The MA plots represent each gene with a dot. The M (Y-axis) represents the log2 fold change between

ePreN and Neuts groups, and A (X-axis) represents the average expression of the overall samples. **(A)** MA plot of the Log₂ fold change of all human orthologues of zebrafish clusters one and two genes. The red dots indicate human orthologues of zebrafish genes of cluster one (67 genes), and the green dots indicate human orthologues of zebrafish genes of cluster two (195 genes). MA plots (B & C) were obtained from the plot (A). **(B)** The plot depicts the differentially expressed human orthologues of zebrafish genes of cluster one. The red dots indicate the genes in ePerN (top) (Log₂ fold change ≥ 2.5), and the green dots indicate the genes in Neuts (bottom) (Log₂ fold change ≤ -2.5). **(C)**. The plot shows the differentially expressed human orthologues of zebrafish of cluster two. The red dots indicate the genes in Neuts (bottom) (Log₂ fold change ≤ -2.5), and the green dots indicate the genes in ePerN (top) (Log₂ fold change ≥ 2.5). Blue dots represent differentially expressed genes, while grey dots represent non-differential expressed genes ($p < 0.05$) (plots A, B and C). **(D)** The plot depicts the differentially expressed human orthologues of zebrafish genes of both clusters one and two. The blue and red dots indicate the genes of cluster one in ePerN (top) (Log₂ fold change ≥ 2.5) (40 genes) and in Neuts (bottom) (Log₂ fold change ≤ -2.5) (3 genes), respectively. The green and orange dots indicate the genes of cluster two in Neuts (bottom) (Log₂ fold change ≤ -2.5) (58 genes) and in ePerN (top) (Log₂ fold change ≥ 2.5) (17 genes), respectively. Grey dots represent differentially expressed genes, while the White space around the horizontal line represents non-differential expressed genes ($p < 0.05$). The black horizontal line at zero provides genes with similar expression values in ePreN and Neuts groups. **(E)** Venn diagram shows the number of genes of the overall differentially expressed genes identified in (D).

To further demonstrate the presence of neutrophil maturation stages in our clusters one and two, Grassi and colleagues' data (Grassi et al., 2018) of human neutrophils were examined to see whether they were in agreement with our Dinh-based cell stage allocation data (Dinh et al., 2020). In Grassi colleagues' study, five stages of human neutrophil differentiation were analysed to gain insight into their functions and regulation. These were the promyelocytes and myelocytes (P/Ms), metamyelocytes (MMs), band neutrophils (BNs), segmented neutrophils (SNs) and the blood circulating polymorphonuclear neutrophils (PMNs) (Grassi et al., 2018). Bulk RNA-seq was used to characterise and compare the gene expression profiles of different neutrophil maturation stages. The gene expression analysis result was integrated with epigenetic analysis during five neutrophil maturation stages. The result revealed that there was a strong correlation between epigenomic and transcriptomic profiles at their differentiation stages from progenitors to mature cells (Grassi et al., 2018). The website (<https://blueprint.haem.cam.ac.uk/neutrodiff/>), which houses the analysed result used in Grassi and colleagues' data, was used to validate some of the genes that were found to be differentially expressed in our clusters one and two. In our study, with regards to cell cluster maturation stages, two of Grassi colleagues' cell types were selected for comparison with our Dinh-based cell stage differentiation data, i.e., the P/Ms and SNs. P/Ms were chosen to

present the gene expressions as immature neutrophils in the human bone marrow to compare with our Dinh-based cluster one genes as an immature group in zebrafish kidney marrow. In the same manner, the SNs group was selected as mature neutrophils in the human bone marrow to compare with our Dinh-based cluster two genes as a mature group in zebrafish kidney marrow. Comparing the Grassi and colleagues' data with our Dinh-based data revealed that the most significant differentially expressed genes in cluster one were more highly expressed in P/Ms. This could represent an immature stage of neutrophil differentiation, including ribosomal genes (Figure 3.9). Likewise, a comparison of the two studies revealed that the majority of significant differentially expressed genes in cluster two were more highly expressed in SNs, including some of specific neutrophil genes i.e., chemokine receptor CXCR1. This could also represent the mature stage of neutrophil differentiation (Figure 3.10). Our findings concerning the differentiation stage allocation of up-regulated genes in clusters one and two concur with Grassi and colleagues published data. In summary, our analysis of scRNA-Seq data suggest that different cell clusters represent different neutrophil differentiation stages. Together, our results further highlight that clusters one and two could represent the immature and mature neutrophils, respectively. This could mean that the change in gene expression from one cluster to another suggests that the clusters of

zebrafish scRNA-Seq data could show different gradual maturation stages.

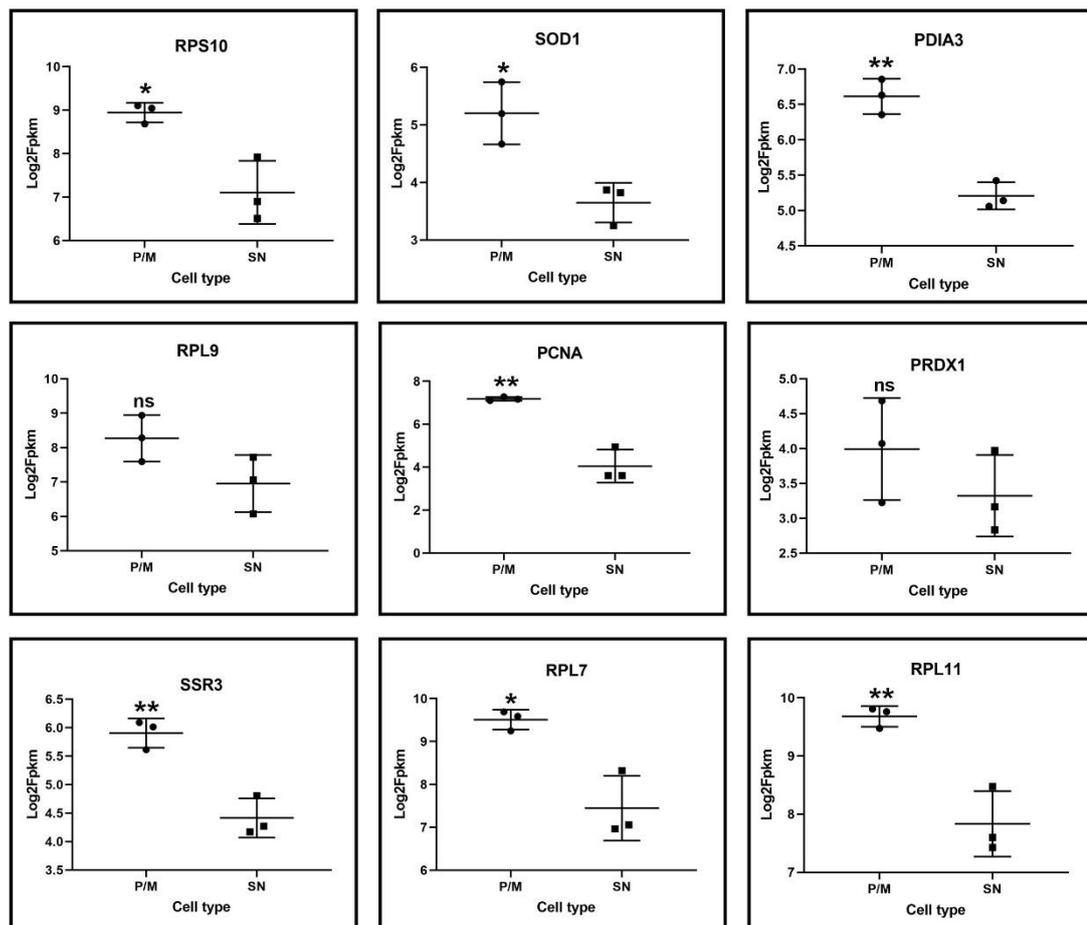


Figure 3.9 Up-regulation of the gene expression levels of human orthologues of zebrafish genes of cluster one in comparison of promyelocytes and myelocytes (P/Ms) relative to the segmented neutrophils (SNs)

Graphs show the mean levels of genes in P/Ms and SNs, alongside the standard deviations for the biological replicates. (*t*-test corrected *p* values are indicated in the figure: ns, nonsignificant, **p* < 0.05 and ***p* < 0.01).

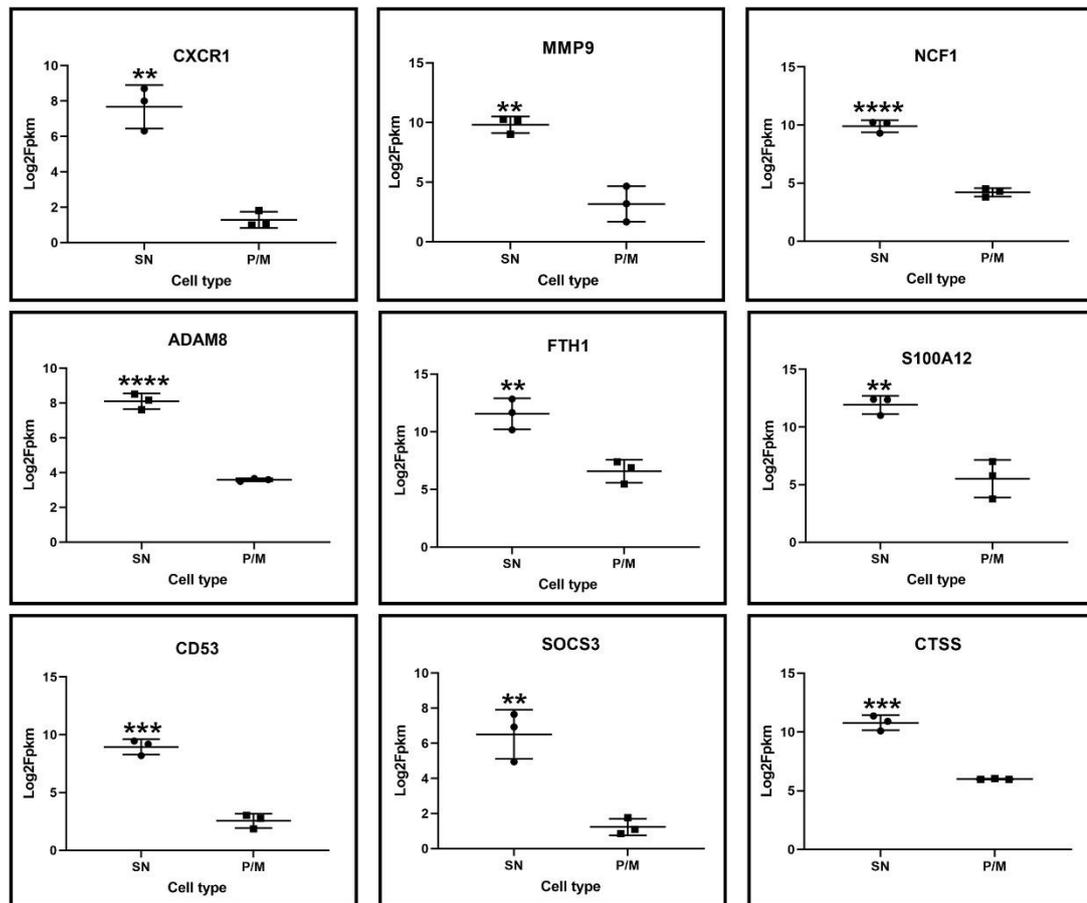


Figure 3.10 Up-regulation of the gene expression levels of human orthologues of zebrafish genes of cluster two in comparison of the segmented neutrophils (SNs) relative to the promyelocytes and myelocytes (P/Ms).

Graphs show the mean level of genes in SNs and P/Ms, alongside the standard deviations of the biological replicates. (*t*-test corrected *p* values are indicated in the figure: ***p* < 0.01, ****p* < 0.001 and *****p* < 0.0001).

3.2 Verification of the DNA sequence of the region downstream of the *c-myb* gene

To find out whether zRE1 is necessary for endogenous *c-myb* expression in haematopoietic cells, we decided to delete the zRE1 from the genomic DNA downstream of the *c-myb* gene. For this purpose, Cas9 target sites were defined, and guide RNAs were designed and generated. Before identifying these sites, the DNA sequence downstream of the *c-myb* was verified in our Nottingham zebrafish. The wt *c-myb* gene lies on chromosome 23 of the (*Danio rerio*) zebrafish genome and consists of sixteen coding exons (Ensembl: ENSDARG00000053666) (Figure 3.11). To verify the sequence, large stretches of DNA in a region of approximately 17 Kb downstream of the *c-myb* gene were amplified in small pieces using PCR on wild-type genomic DNA isolated from fin clips of six wild type fish. The small pieces covered the sequences around potential Cas9 target sites. All PCR products (1-19 fragments) shown in (Table 2.7) were sequenced. Their sequence was compared to that published in ENSEMBL (GRCz11, October 2018) in order to identify the presence of any single nucleotide polymorphisms (SNPs) in the *c-myb* locus of our Nottingham wild-type fish. Several SNPs were identified in the region downstream of *c-myb* gene (Appendix figure 6.3). There were, however, no SNPs in most of the target sites of interest except for TS2. The SNP was found in TS2 in the position of the third nucleotide upstream of the PAM sequence. The sequence of TS2 was

modified by changing the nucleotide T of the SNP to A. All the primers that were used to amplify the whole of the downstream region of the *c-myb* are listed in (Table 2.7).

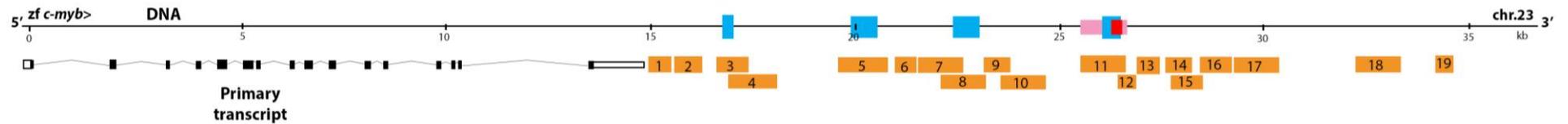


Figure 3.11 Verification of the DNA sequence downstream of *c-myb*

Genomic map of the wild-type zebrafish *c-myb* locus (the black line on the top of the figure) showing the primary *c-myb* transcript. Within the transcript, boxes represent the 16 exons of the *c-myb* gene. Coding sequences are depicted as black boxes, 5' and 3' UTRs as white boxes. The red and pink boxes represent the zCNE1 and zRE1 sequences, respectively. The light blue boxes show the potential enhancers MP40036-, MP40036, MP40037 and MP40037+. Fragments that were amplified and sequenced are shown as orange rectangles numbered from 1-19. The fragments were amplified using the primers listed in (Table 2.7).

3.3 Deletion of zRE1 from the wild-type zebrafish genome using CRISPR/Cas9

Having shown that zRE1 (includes zCNE1), which acts as zebrafish regulatory element, is sufficient to direct *c-myb* promoter activity and therefore reporter gene expression to definitive haematopoietic cells in transgenic zebrafish (Hsu, 2010), it remained to be determined whether this zRE1 was also necessary for endogenous *c-myb* expression in these cells. To address this question, the CRISPR/Cas9 was used to delete zRE1 from the zebrafish genome.

A volume of 1 μ l of a mix of gRNAs TS2 and TS7, Cas9 protein and the fluorescent dye mini ruby was injected into 1-cell stage wild-type zebrafish embryos. At approximately 5 hpf, the embryos were checked by fluorescent microscopy to confirm the success of the injection (Figure 3.12A). Only successfully injected embryos were grown on. At 24 hpf, the genomic DNAs of 12 single embryos, one uninjected and 11 injected embryos, were extracted and then used as a template for nested PCR amplification. The primers (external pair (DB1436 and DB1437) and internal pair (DB871 and DB899)) were designed to allow the amplification of a DNA fragment across the fusion site between TS2 and TS7 on the deletion allele (Figure 3.12B & C). The expected nested PCR fragment was \sim 470 bp in size. The expected PCR fragment of \sim 470 bp was detected in 10 out of 11 injected embryos. As expected, a similar PCR product could not be generated on the wild-type control DNA. To confirm that the PCR product amplified on the genomic DNA

of the injected embryos corresponded to the sequence of the expected deletion allele of *c-myb*, the PCR product that was amplified on the genomic DNA of embryo number 2 was isolated, purified and sequenced using primer DB899. The sequencing results showed that the sequence did indeed correspond to the sequence downstream of the *c-myb* gene. Closer analysis of the sequence revealed that in this sequence, the first 13 bp of TS2 were fused to the last two nucleotides of the PAM sequence of TS7, suggesting that exactly 2976 bp had successfully been deleted between the two TSs (Figure 3.12D). Because the sequencing suggested that guide RNAs 2 and 7 could successfully target Cas9 to their target sites and delete the sequence between them, the rest of injected embryos were raised to adulthood.

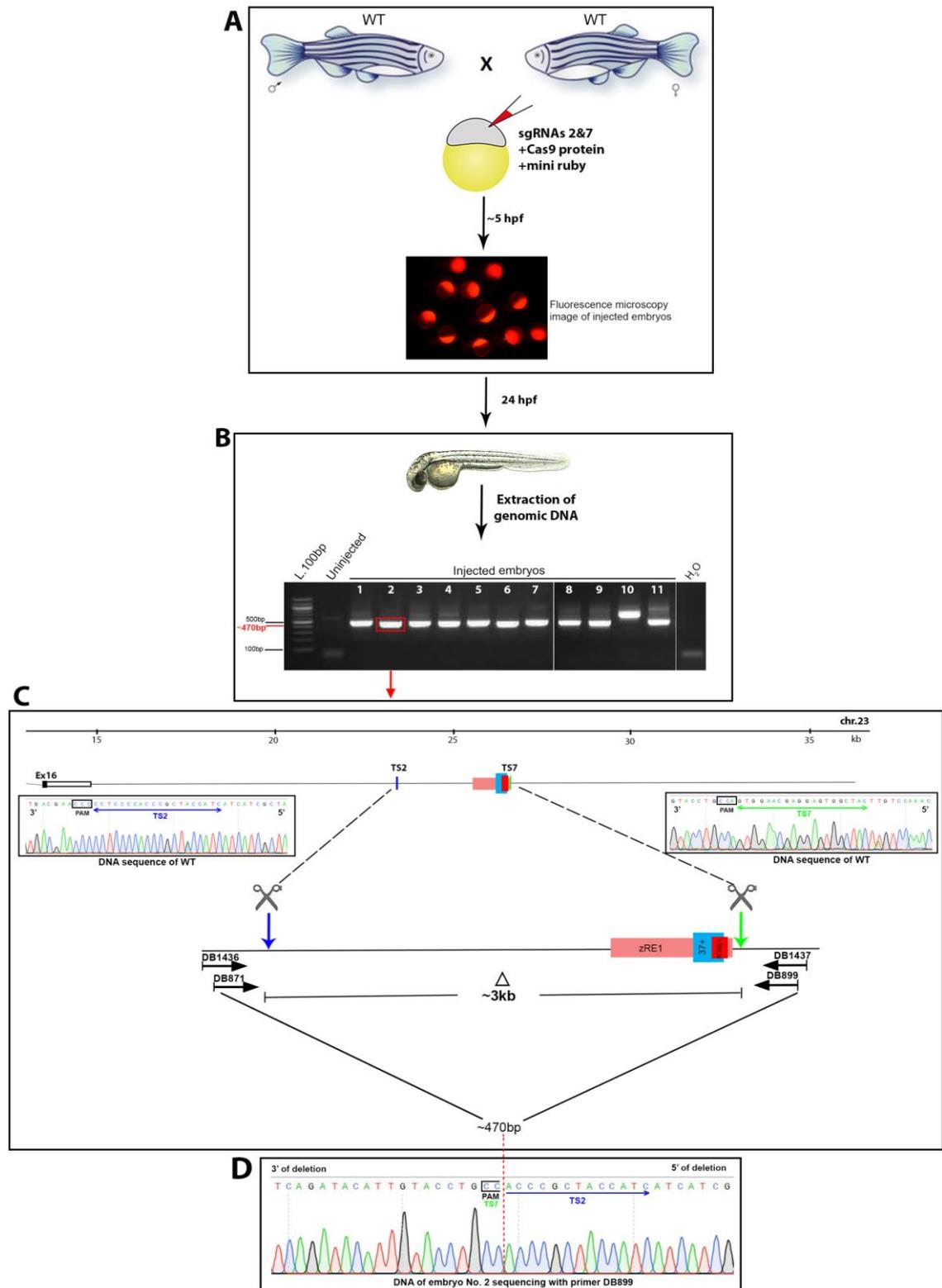


Figure 3.12 *c-myb* 2&7 sgRNAs successfully produced deletion in the target sites 2 and 7

(A) Wild-type (wt) zebrafish were incrossed, and one cell stage zebrafish embryos were microinjected with the sgRNAs 2 and 7, the Cas9 protein and the fluorescent dye mini ruby. At 5 hours post fertilisation (hpf), fertilised red

fluorescent embryos were sorted. **(B)** At 24 hpf, genomic DNA was isolated from individual uninjected and successfully injected embryos and was used as a PCR template. PCR amplification on genomic DNAs was carried out using primers (external pair (DB1436 and DB1437) and internal pair (DB871 and DB899)). It was expected to yield a ~470bp fragment on genomic DNA that had the sequence between TS2 and 7 deleted. The red rectangle refers to the embryo that was genotyped. **(C)** Schematic illustrating the genomic DNA editing by the Cas9/TS2 and 7 sgRNA complexes. The inserts in the top show Sanger sequencing of TS2 and 7 within the DNA of wt zebrafish. Grey scissors indicate CRISPR/Cas9 deletion breakpoints. Blue and green arrows show TS2 and 7, respectively. The red box represents a conserved noncoding element one (zCNE1). The light blue box shows potential enhancer (PE) MP40037+ (Aday et al., 2011). A pink box shows regulatory element one (zRE1). Position of the (DB1436-DB1437) and (DB871- DB899) as (forward-reverse) primers used in the PCR reaction are shown with the size of the fragment amplified. **(D)** Sanger sequencing of the DNA of embryo number 2 with primer DB899 shows the deletion breakpoints (indicated by the red dashed line) in the mutant allele identified in the genomic DNA of injected embryo number 2. The retained sequences of TS2 is shown between the brackets. The retained two nucleotides of the PAM sequence of TS7 is shown between the black box. The PCR products were analysed on a 1.5% agarose gel.

At 3-4 months, F0 adult fish grown from injected embryos with sgRNAs 2 and 7 were screened for transmission of the deletion allele through the germline. To this end, putative founders were outcrossed to wild-type fish (Figure 3.13A). From the individually successful cross, a group of 11 single embryos were randomly selected and genotyped as described before using nested PCR with external oligos (DB1436-DB1437) and internal oligos (DB871-DB899) at 24 hpf. PCR amplification showed the presence of the approximately 470 bp band in 10 out of 11 embryos that derived from the outcross of one of the putative founders. The PCR product of sample number 5 was gel-extracted and sequenced. The Sanger sequencing result verified that 2974 bp of sequence between the two TSs were deleted. In addition, an 8 bp insertion (CATCCAGT) and a single nucleotide polymorphism (A>C) were observed in the sequence. (Figure 3.13B). This deletion allele was called *qmc193*. The male that carried the deletion allele in its germline was crossed to a wild-type female. Their progeny established the *qmc193* line.

When the F1 *qmc193* fish were fertile adults, potential heterozygotes were anaesthetised and fin-clipped. Nested PCR as described before, was used to determine the heterozygous individuals. Four out of five F1 fin-clipped fish carried the *qmc193* mutation, and DNA sequence of fish number 5 confirmed that (Figure 3.13C).

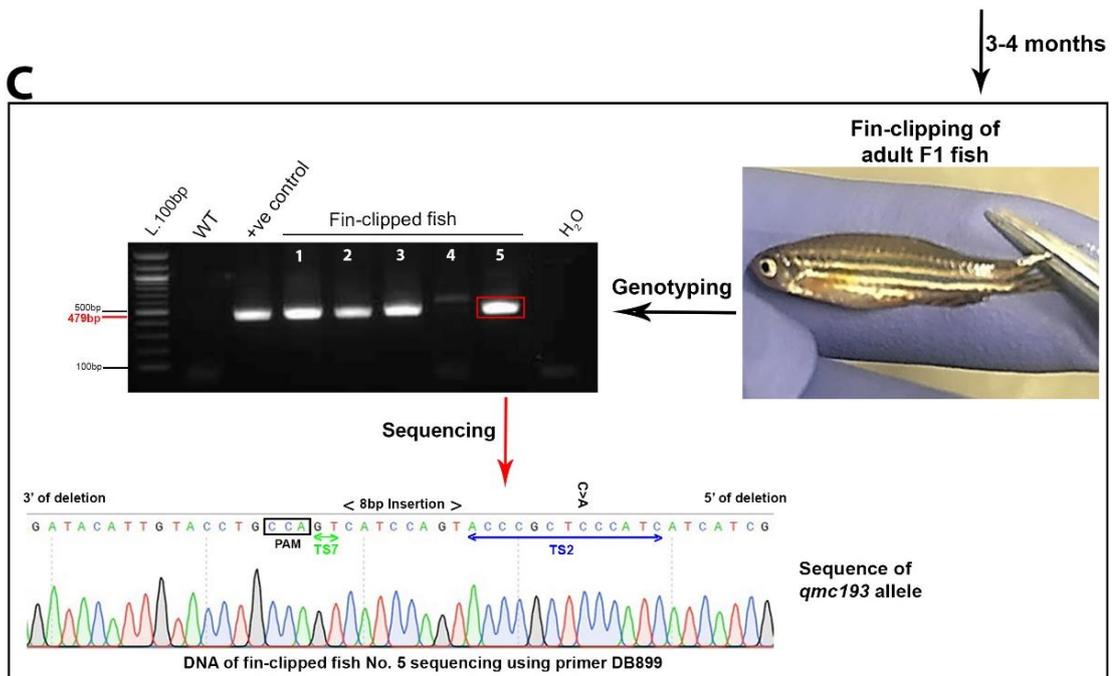
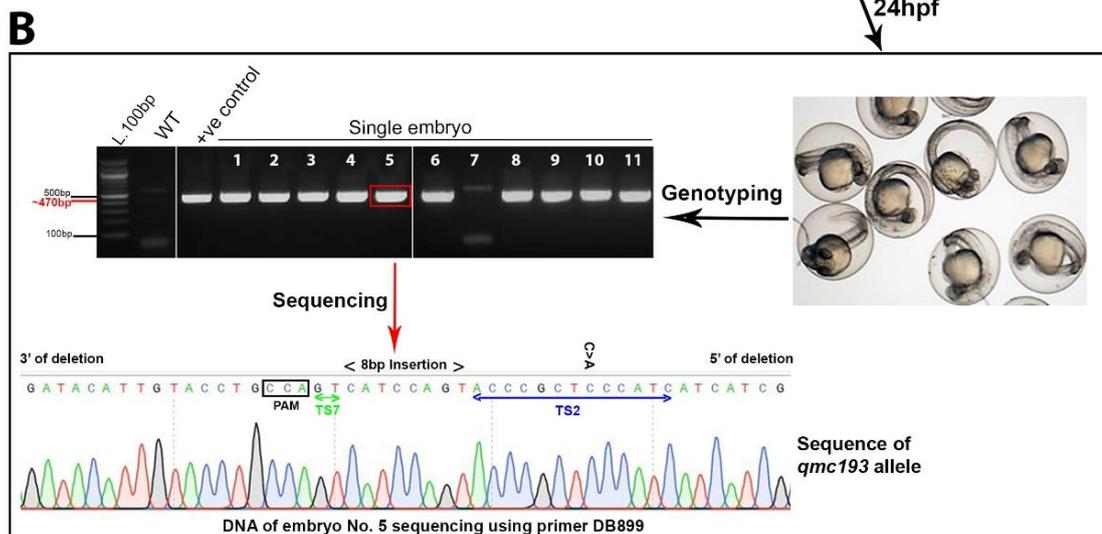
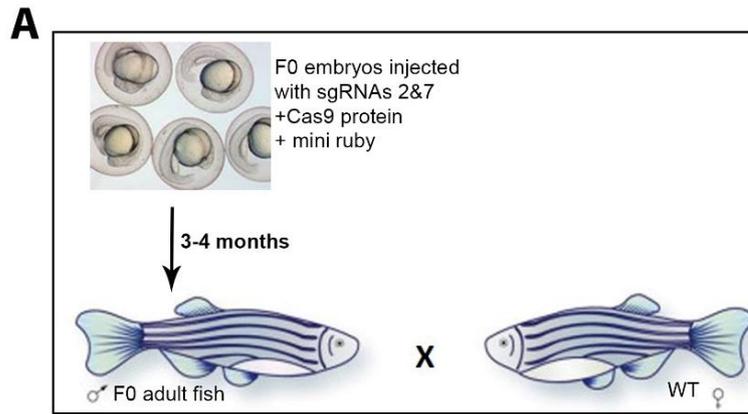


Figure 3.13 Germline transmission of the *c-myb* TS2 and TS7 zRE1 deletion allele

(A) Embryos injected with *c-myb* sgRNAs 2 and 7 were raised to adulthood and then outcrossed to wild-type fish. **(B)** At 24 hpf, eleven F1 embryos were collected for DNA extraction and genotyped using nested PCR as described in Figure 3.12 using primers (external pair (DB1436 and DB1437) and internal pair (DB871 and DB899)). Genomic DNA of one of F0 injected embryos was used as a positive control. The expected 479 bp is found in nine out of ten of F1 embryos. Sanger sequencing was used to characterise the deletion in embryo number five, an eight base pair insertion (CATCCAGT) and one single nucleotide polymorphism (A>C) were detected in the sequence. **(C)** The remaining embryos of F1 progeny are raised to adulthood. Caudal fin clips were taken from five adult F1 fish. Fin clip No.5 was genotyped and sequenced to identify adult heterozygous F1 carries the mutant *c-myb*^{qmc193} allele. The retained sequences of TSs 2 and 7 are shown between the brackets. The retained sequences of PAM of TS7 is shown between the black box. The PCR products were analysed on a 1.5% agarose gel.

To generate F2 *qmc193* homozygous fish and to determine the ratio of the wt and *qmc193* alleles in the progeny, heterozygous *qmc193* adult fish were incrossed. At 24 hpf, a group of 14 single embryos were collected and genotyped. The genomic DNA was extracted and used as a template to identify wt, *qmc193* heterozygous and *qmc193* homozygous embryos. Here, two PCR amplification experiments were performed. The first PCR experiment used primers DB1190 and DB1191 that hybridise to the genomic sequence upstream of TS7. It yields a 271 bp PCR fragment that can only be generated on the wt allele. The second, PCR amplification was carried out to detect the 479 bp mutant allele, using the same primer pairs that were used before with F0 and F1 progeny (Figure 3.14A & B). Analysis of the genotyping data from 14 embryos revealed that the incross of two *qmc193* heterozygous fish yielded 4 (28%) wt embryos, 5 (36%) heterozygous and 5 (36%) homozygous embryos, confirming that expected genotypes were present in the progeny. The rest of the F2 progeny was raised to adulthood.

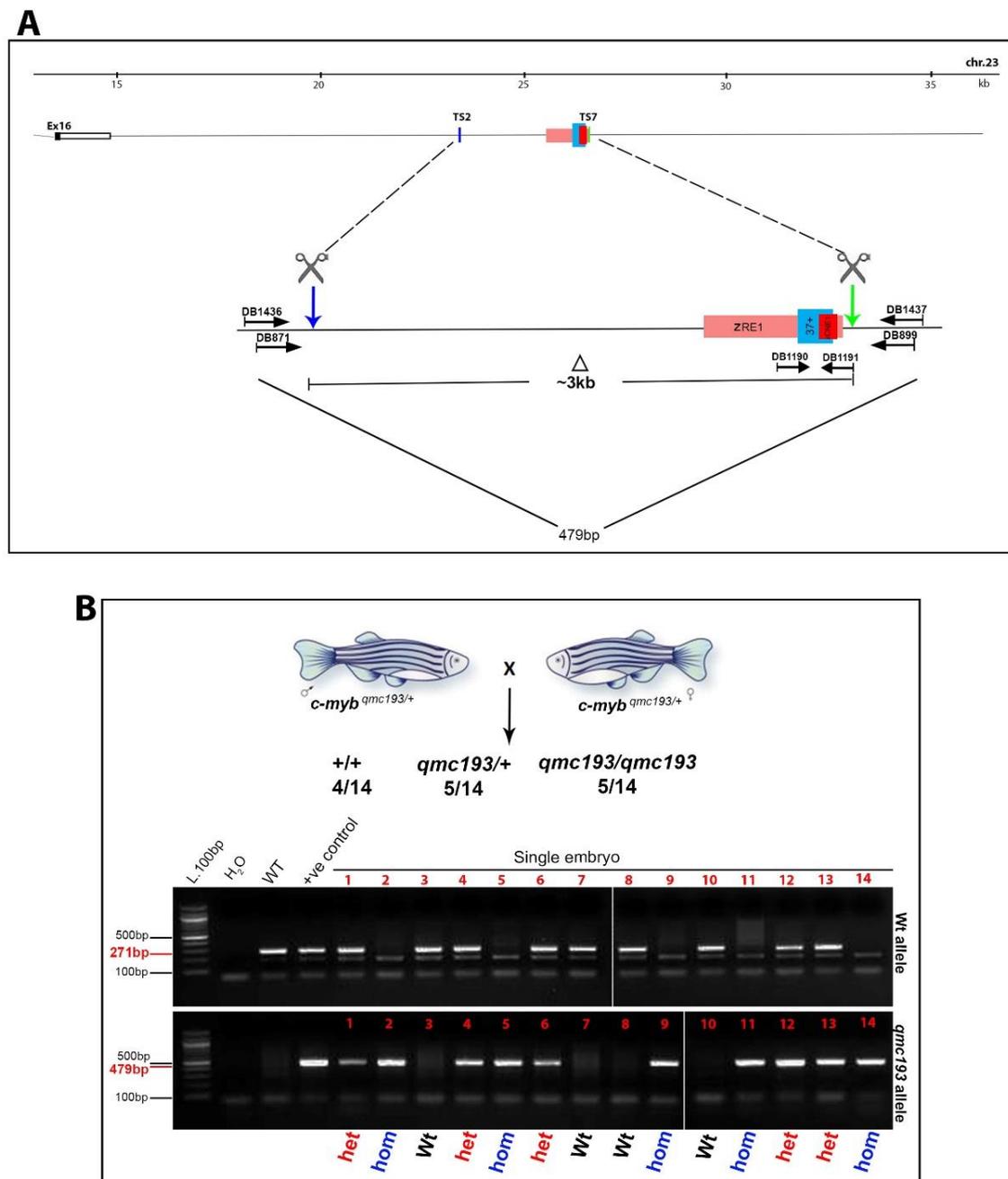


Figure 3.14 Identify homozygous mutants in F2 embryos

(A) Schematic illustrating the genomic DNA editing by the Cas9/TS2 and TS7 sgRNA complexes as shown previously in (Figure 3.12C). Position of the primers (external pair (DB1436 and DB1437) and internal pair (DB871 and DB899)) as forward and reverse primers, used in the PCR reaction, are shown with the size of the mutant allele fragment amplified. The wt allele was amplified using primers DB1190 and DB1191. **(B)** Diagram of a *c-myb*^{qmc193} fish incross yielding 28% wt embryos and 36% for both heterozygous and homozygous embryos. PCR amplification on genomic DNA extracted from 24 hpf embryos obtained from the incross, using (DB1190- DB1191) primers to amplify wt allele (top gel) and primer pairs DB1436 and DB1437 as well as DB871 and DB899 for the mutant allele (bottom gel). The PCR products were analysed on a 1.5% agarose gel.

At ~3-4 months, fin-clips were taken, and genomic DNA was used to identify the *qmc193* F2 fish using PCR amplification under the same conditions as described before. Among the 26 fish analysed, 5 (19%) were wt fish, 14 (54%) were heterozygous for *qmc193*, and 7 (27%) were homozygous for *qmc193* (Figure 3.15A). The results suggest that all genotypes were present in the expected Mendelian ratios (χ^2 -test; $P < 0.9$). In addition and to confirm that the homozygous *qmc193* are fertile and carry *qmc193* alleles, one of the identified homozygous *qmc193* was outcrossed to a wt fish, and then 10 of the offspring were genotyped to check the presence of wt and *qmc193* alleles. All the progeny were found to be heterozygous for *qmc193* (Figure 3.15B). The results suggest that homozygous *qmc193* were fertile and fully viable. The identified fish showed no phenotypic abnormalities.

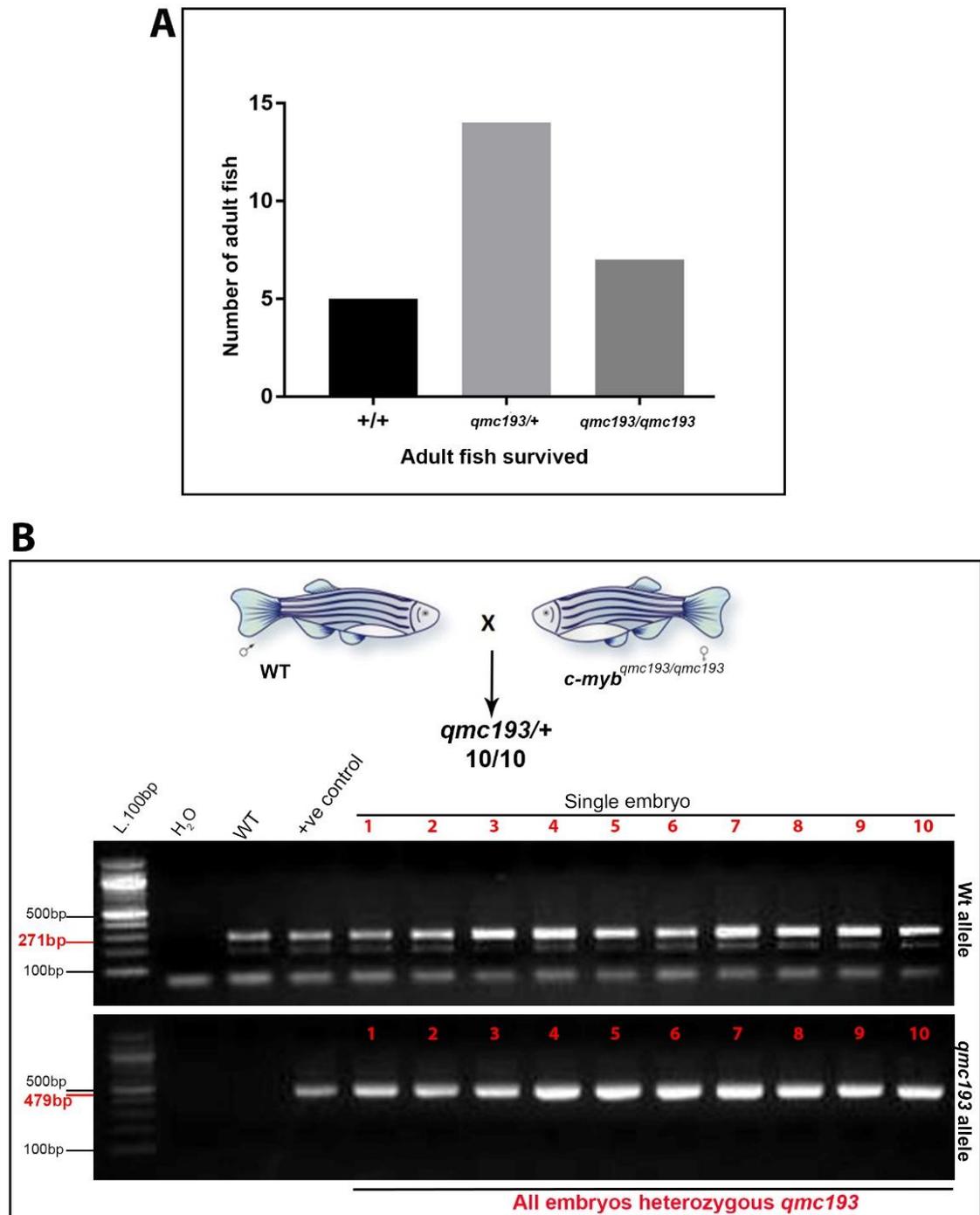


Figure 3.15 Identification of F2 homozygous *qmc193* mutants' adult fish

(A) The bar shows the number of genotyped adult F2 fish that had been derived from an incross of heterozygous *qmc193* carriers. **(B)** Diagram of *c-myb*^{*qmc193/qmc193*} fish outcross yields 100% heterozygous embryos. The genomic DNAs of individual embryos were tested for the presence of wt and *qmc193* mutant alleles. The PCR experiments used primers DB1190 and DB1191 for the wt allele (top gel) and primer pairs DB1436 and DB1437 as well as DB871 and DB899 for the mutant allele (bottom gel). The PCR products were analysed on a 1.5% agarose gel.

Next, whole-mount *in situ* hybridisation experiments were performed to study neutrophil development in *qmc193* homozygous embryos. Loss of *c-myb* expression has recently been shown to interfere with neutrophil development in the early zebrafish embryo (Jin et al., 2016). As zRE1 directed *c-myb* promoter activity to embryonic and adult neutrophil progenitors, it was conceivable that loss of zRE1 interfered with neutrophil development. To address this question, 34-36 hpf embryos derived from an incross of heterozygous *qmc193* carriers were stained in a WISH experiment with a probe against *lyz* mRNA. This experiment did, however, not reveal any differences in staining in a batch of 20 embryos. Genotyping 8 of the embryos identified a homozygous *qmc193* carrier that displayed the same level of *lyz* mRNA that was present in wt and heterozygous *qmc193* siblings (Figure 3.16Ai-iii and Appendix figure 6. 4), suggesting that embryonic neutrophil development is normal in the early homozygous *qmc193* embryos.

In the absence of an obvious phenotype, it remained to be seen whether the loss of zRE1 had any effect on the level of *c-myb* expression in the blood cells of the developing embryo. To address this question, the expression of *c-myb* was studied in the primitive blood cells in embryos homozygous for the *qmc193* allele. For this purpose, whole-mount *in situ* hybridisation experiments were performed on embryos obtained from a *qmc193* heterozygous incross at 15 somites, 30 hpf and 2 dpf (Appendix figure 6.5, Appendix figure 6. 6 and

Appendix figure 6. 7). Remarkably, *c-myb* expression was retained in all embryos at all the different stages. This suggests that the zRE1 may not be required for the normal level expression of the *c-myb* gene in primitive blood cells. Next, WISH experiments were performed to study the level of *c-myb* expression in definitive blood cells of the caudal haematopoietic tissue. For this purpose, 18 embryos derived from an incross of heterozygous *qmc193* carriers were stained using a probe against *c-myb* mRNA. This experiment did not reveal any obvious differences in the level of *c-myb* expression at 4 dpf compared to the wild type embryos that had the normal expression of the *c-myb* gene at the same time point (Figure 3.16Bi-iii and Appendix figure 6.8). Genotyping 9 of these embryos identified 3 homozygous *qmc193* carriers that displayed no obvious reduction in *c-myb* expression, suggesting that while zRE1 is sufficient to direct *c-myb* promoter activity to haematopoietic cells of the CHT, this element is not essential for the expression of the *c-myb* gene in these cells.

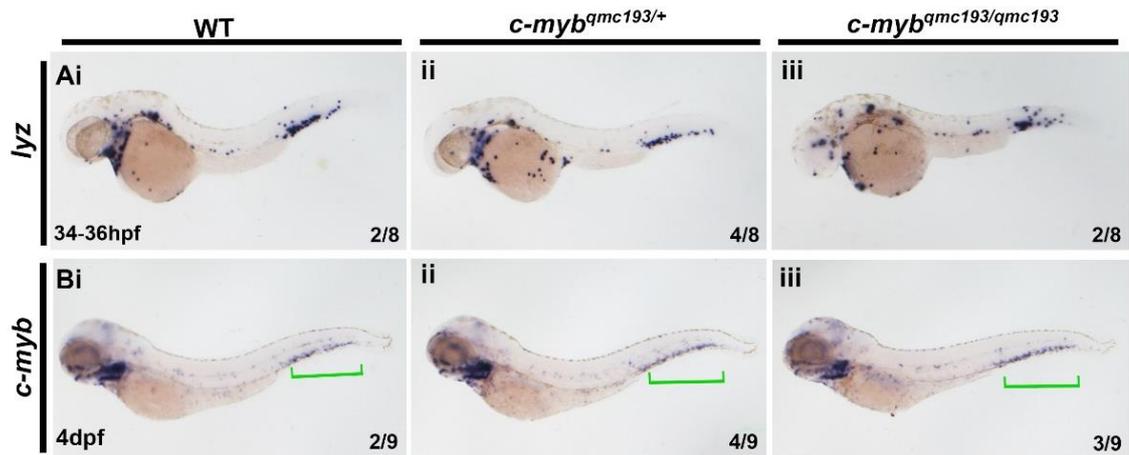


Figure 3.16 *Lyz* and *c-myb* expression were maintained in 34-36 hpf and 4 dpf *qmc193/qmc193* embryos

WT, *qmc193* heterozygous, and homozygous embryos are showing *lyz* and *c-myb* mRNA expression detected by WISH. Embryos were obtained from an incross of *qmc193* heterozygous fish. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT).

3.4 Attempted CRISPR/Cas9-mediated deletion of three potential enhancers (PEs) downstream of zebrafish *c-myb* gene

In addition to eliminating the zRE1 from the wild-type zebrafish genome, we attempted to delete a potential *cis*-regulatory region of the genomic DNA downstream of the *c-myb* gene in order to find out whether the PEs MP40036-, MP40036 and MP40037 (see Figure 1.12 and Table 1.1) were required for endogenous *c-myb* expression in haematopoietic cells. For this purpose, Cas9 target site 16 (TS16) with a score 42 was identified using the CRISPRscan track. Then, its guide RNA was designed and generated using the previously described procedure (see Figure 2.3). The sequence of TS16 is located on the reverse strand 325 bp downstream of the exon 16 as shown in (Figure 2.1). Deletion of the sequence between this target site and TS2 that had been used before to generate *qmc193* mutant allele was expected to remove the region between exon 16 of *c-myb* gene and zRE1.

At a one-cell stage, embryos that were collected from a wt incross were injected with 1 µl of a mixture containing gRNAs against TS16 and TS2, Cas9 protein and the fluorescent dye mini red ruby (Figure 3.17A). The presence of red fluorescence in the embryonic blastomeres confirmed the successful injection of the embryos. At 24 hpf, ten injected embryos and one uninjected embryo were randomly collected for genotyping to examine whether the CRISPR/Cas9 system had efficiently deleted the region of interest. The genomic DNA was

extracted from the embryos and was then used as a template for nested PCR amplification using first the external primers DB1371 and DB967 and then the internal primers DB1418 and DB968. The nested PCR fragment was expected to have a size of ~265 bp (Figure 3.17B & C). The results showed that the expected PCR fragment of approximately 265 bp was detected in 7 out of 10 injected embryos. The PCR product of embryo number nine was isolated, purified and sequenced using primer DB1371. Sanger sequencing analysis verified that the amplified DNA was derived from the *c-myb* locus and confirmed the absence of the PEs MP40036-, MP40036 and MP40037+ downstream of *c-myb* in the deletion allele (Figure 3.17D). The mutant allele carried a deletion of 7,853 bp with the breakpoints located in TS16 and TS2. The PAM sequence of TS16 on the reverse strand was fused to the last 2 bp upstream of the PAM sequence of TS2 on the forward strand. The sequencing result verified that the CRISPR/Cas9 system successfully delete all three PEs from the *c-myb* locus, and as a result, the siblings of the genotyped embryos were raised to adulthood.

When the F0 injected fish were sexually mature, the fish were outcrossed with wild-type fish to find a founder that carried the deletion allele in its germ cells (Figure 3.17E). At 24 hpf, gDNA of individual embryos was then extracted and genotyped using established nested PCR protocol. Unfortunately, no founder was detected within 33 putative founders that were genotyped.

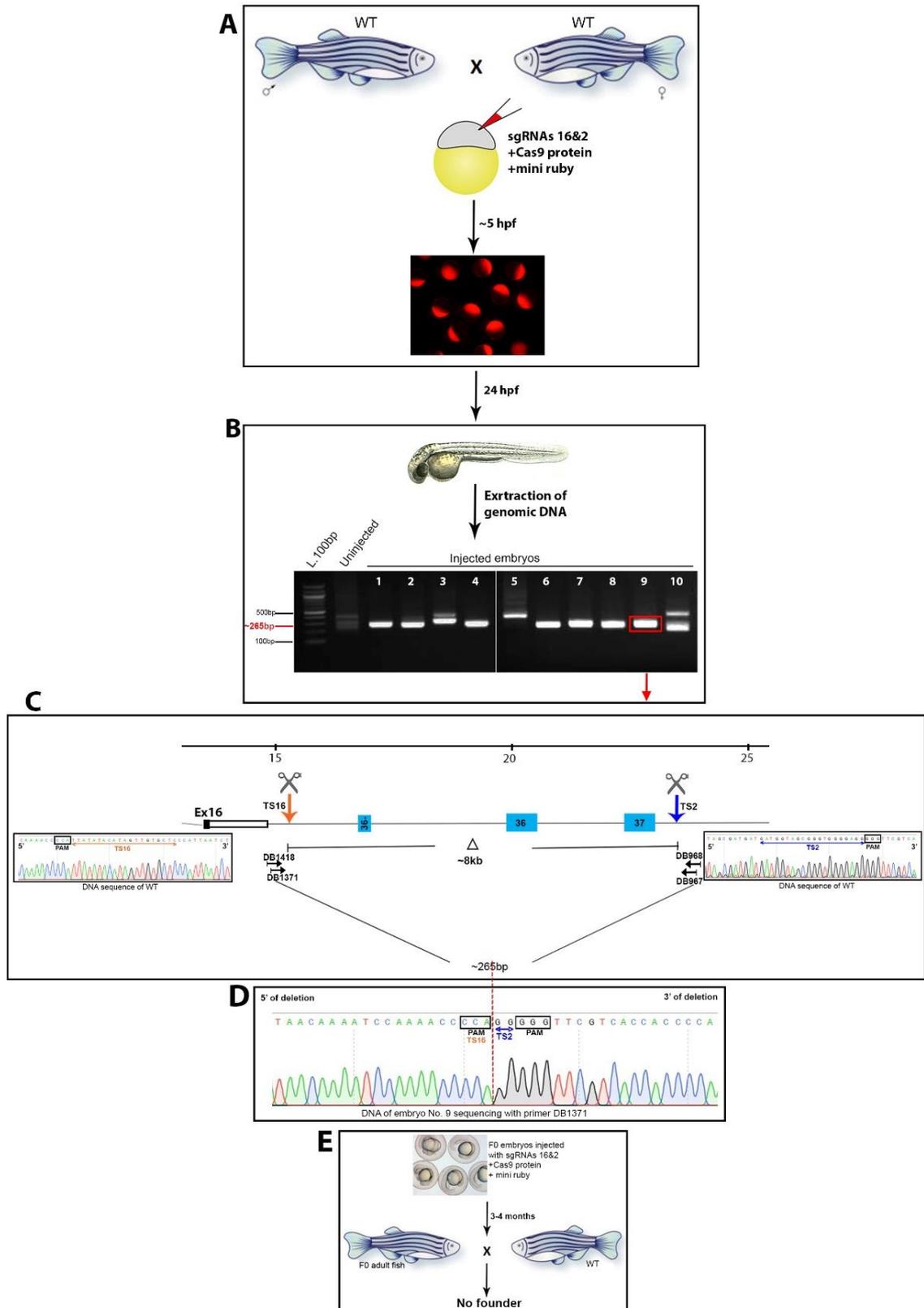


Figure 3.17 *c-myb* 16&2 sgRNAs successfully produced deletion in the target sites 16 and 2

(A) Wild-type (wt) zebrafish were incrossed, and one cell stage zebrafish embryos were microinjected with the sgRNAs 16 and 2, the Cas9 protein and

the fluorescent dye mini ruby. At 5 hpf, fertilised red fluorescent embryos were sorted. **(B)** At 24 hpf, genomic DNA was isolated from individual uninjected and successfully injected embryos and was used as a PCR template. PCR amplification on genomic DNAs was carried out using primers (external pair (DB1418 and DB968) and internal pair (DB1371 and DB967)). The PCR products were analysed on a 1.5% agarose gel. PCR amplification was expected to yield a ~265bp fragment on genomic DNA that had the sequence between TS16 and 2 deleted. The red rectangle highlights the PCR fragment that was extracted from the gel and was sequenced. **(C)** Schematic illustration of the genomic DNA editing caused by the successful injection of the Cas9/TS16 and 2 sgRNA complexes. The inserts in the top show Sanger sequencing of TS16 and 2 within the DNA of wt zebrafish. Grey scissors indicate CRISPR/Cas9 deletion breakpoints. Orange and blue arrows show TS16 and 2, respectively. The light blue boxes show potential enhancers (PEs) MP40036-, MP40036 and MP40037. The positions of the forward and reverse primers used in the PCR reaction are shown together with the sizes of the expected PCR fragments. **(D)** Sanger sequencing of the forward strand of the PCR product of embryo number 9 with primer DB1371 revealed the deletion breakpoints (indicated by the red dashed line) in the mutant allele. The retained sequence of TS2 is shown between the brackets. The PAM sequences of TS16 and TS2 are shown in the black boxes. **(E)** At ~3-4 months, progeny were scanned for germ transmission by outcrossing the F0 injected fish to wt fish. The progeny were genotyped using the same nested PCR protocol to check for the presence of a deletion allele. None of the F0 adults had transmitted the mutation to their progeny.

3.5 Successful CRISPR/Cas9-mediated deletion of four potential enhancers downstream of *c-myb* from the wild-type zebrafish genome

As we did not detect any phenotypic abnormalities in the *qmc193* homozygous fish, it remained to be seen whether the three PEs and the zRE1 (Hsu, 2010) might act as a super-enhancer (Whyte et al., 2013) and whether their combined deletion had any effect on the expression of the *c-myb* gene during zebrafish embryogenesis. To answer this question, CRISPR/Cas9 genome editing was used to delete the entire region from the genomic DNA downstream of the *c-myb* gene. A new TS (TS17) was defined using the CRISPRscan. Its guide RNA was generated using the previously described protocol (see Figure 2.2 and Figure 2.3). TS17 is located 965 bp downstream of the exon 16 (Figure 2.1) and has a score value of 51. By targeting this TS and TS7 that had previously used to generate the *qmc193* mutant allele, the deletion of the sequence between the two TSs was expected to delete an approximately 10 kb fragment that encompasses the three PEs as well as zRE1.

A volume of 1 μ l of a mix of guide RNAs 17 and 7, Cas9 protein and the fluorescent dye mini ruby was injected into the one-cell stage wild-type embryos (Figure 3.18A). Under the fluorescent dissecting microscope, successfully injected embryos were identified by the red fluorescence. To determine whether the region of interest (~10 kb) was deleted successfully, nine injected embryos and one uninjected

embryo were randomly collected at 24 hpf. Genomic DNA was extracted from individual embryos and used for PCR amplification using first the external primers DB1365 and DB1437 and, subsequently, the internal primers DB1371 and DB732 in nested PCR reactions. When the nested PCR products were analysed by agarose gel electrophoresis, there were PCR fragments of different sizes visible in the agarose gel, including a ~880 bp fragment which potentially correspond to the expected mutant *c-myb* allele (Figure 3.18B). Seven out of nine injected embryos showed the ~880 bp fragment. The PCR product for embryo number one was extracted from the agarose gel and sequenced using primer DB732 (Figure 3.18C & D). The analysis showed that the fragment represented a *c-myb* deletion allele. The deletion encompassed 10,181 bps. The sequencing revealed that the nucleotides of the PAM sequence of TS7 were fused to the last three nucleotides of TS17 that were followed by the PAM sequence of the TS17. The result indicated the successful deletion of the potential *cis*-regulatory region downstream of *c-myb* locus from the wildtype zebrafish genome. Therefore, the remaining injected embryos were grown to adulthood.

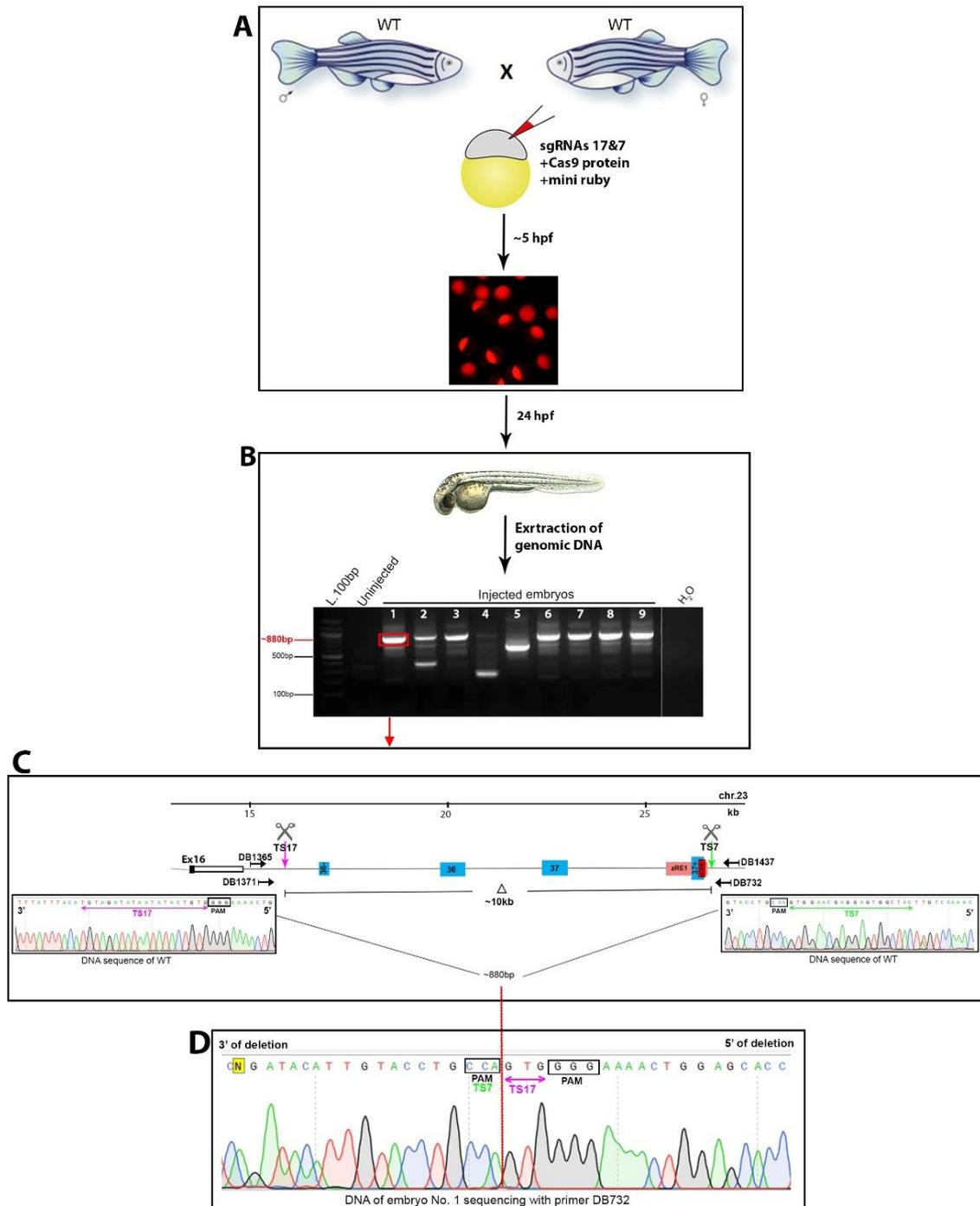


Figure 3.18 Successful CRISPR/Cas9-mediated deletion of three putative enhancer elements as well as zRE1 downstream of the *c-myb* gene

(A) Wild-type (wt) zebrafish were incrossed and one cell stage zebrafish embryos were microinjected with the sgRNAs 17 and 7, the Cas9 protein and the fluorescent dye mini ruby. At 5 hpf, fertilised red fluorescent embryos were sorted. **(B)** At 24 hpf, genomic DNA was isolated from individual uninjected and successfully injected embryos and was used as a PCR template. PCR amplification on genomic DNAs was carried out using first external (DB1365 and DB1437) and then internal primers (DB1371 and

DB732)). The PCR reaction was expected to yield a ~880 bp fragment on genomic DNA that had the sequence between TS17 and 7 deleted. The PCR products were analysed by electrophoresis on a 1.5% agarose gel. The red rectangle refers to the embryo that was genotyped. **(C)** Schematic illustrating the genomic DNA editing by the Cas9/TS17 and 7 sgRNA complexes. The sequencing chromatograms show Sanger sequencing reads of the DNA around the TSs 17 and 7 determined on wt zebrafish genomic DNA. Grey scissors indicate CRISPR/Cas9 deletion breakpoints. Pink and green arrows show the TSs 17 and 7, respectively. The red and pink boxes represent the zCNE1 and zRE1 sequences, respectively. The light blue boxes show the potential enhancers MP40036-, MP40036, MP40037 and MP40037+. The positions of the primers DB1365, DB1437, DB1371 and DB732 are indicated. The sizes of the expected PCR fragments are shown. **(D)** The chromatogram illustrates the result of the Sanger sequencing experiment on the nested PCR product amplified on the genomic DNA of embryo number 1. Primer DB732 was used in the sequencing experiment. The red line shows the deletion breakpoints in the mutant allele identified in the genomic DNA of injected embryo number 1. The retained sequences of TS17 is shown above the double arrow. The PAM sequences of TS17 and TS7 are shown in the black boxes.

To identify a putative founder, adult F0 fish were screened to identify those that carried the deletion allele in their germline. For this purpose, the F0 fish were outcrossed to wild-type fish, and their offspring were analysed for the presence of the mutant allele (Figure 3.19A). At 24 hpf, a group of 10 single embryos were collected, and their gDNAs were extracted for genotyping. Nested PCR amplification revealed the presence of an ~880 bp fragment in 6 of the 10 embryo samples obtained from a male founder. Fragment number six was isolated from the agarose gel and was sequenced. The sequence corresponded to *c-myb* sequence and showed that 10,181 bp were deleted between the two TSs. Closer analysis of the sequence showed that in this sequence, one extra nucleotide had been added (C nucleotide) between the remaining TS7 and TS17 sequences (Figure 3.19B). This deletion allele that lacked the PEs MP40036-, MP40036, MP40037 and MP40037+ downstream of *c-myb* was called *qmc194*. The *qmc194* male founder was crossed to a wt female, and their progeny were raised to adulthood.

Once the F1 *qmc194* fish reached sexual maturity, fin-clips were taken from the putative heterozygotes to identify the *qmc194* allele carrier. Nested PCR, as described before, was used to find the heterozygous individuals. Five out of 10 F1 fin-clipped fish carried the *qmc194* mutation, and DNA sequence of fish number one verified that (Figure 3.19C).

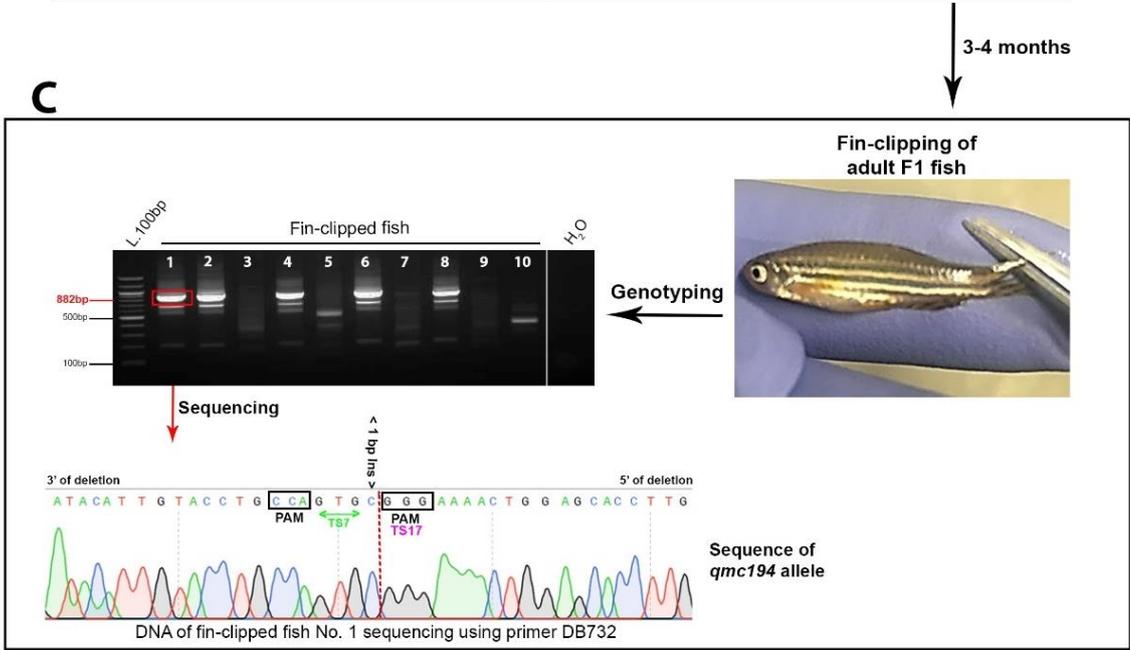
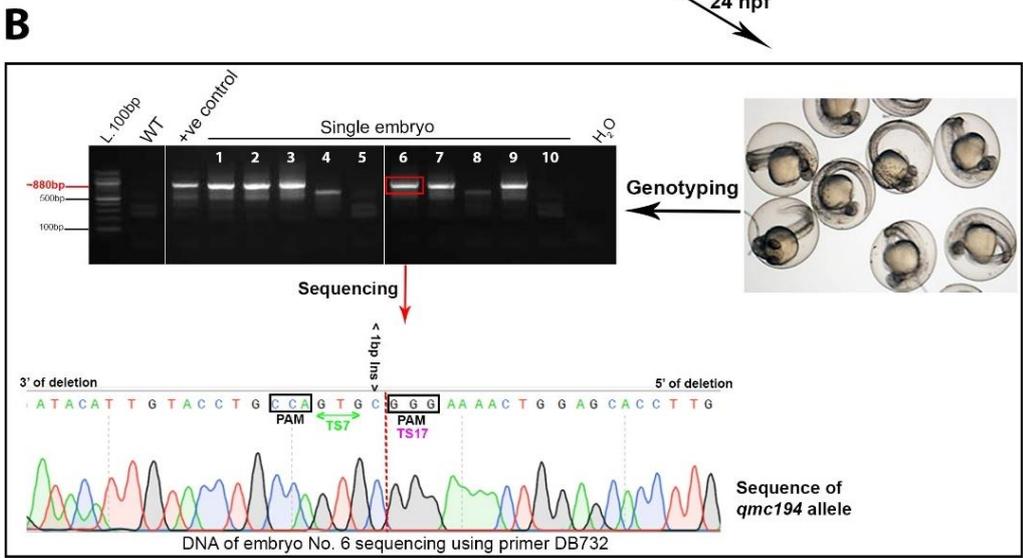
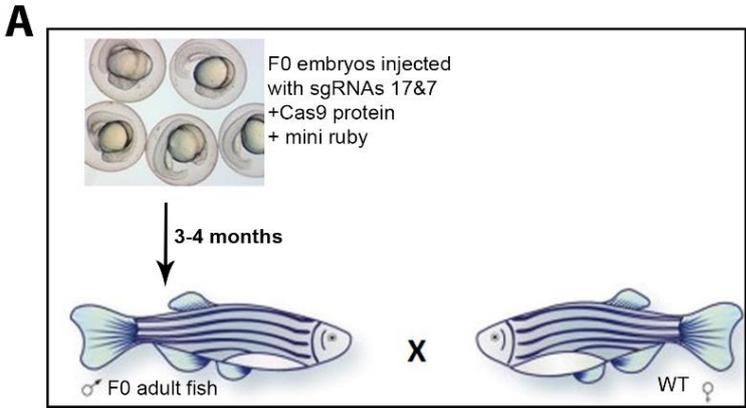


Figure 3.19 Establishment of a zebrafish line that carries the deletion allele *qmc194* that lacks three potential enhancer elements as well as zRE1 downstream of *c-myb*

(A) Embryos injected with *c-myb* sgRNAs 17 and 7 were raised to adulthood and then outcrossed to wild-type fish. **(B)** At 24 hpf, ten F1 embryos were collected for DNA extraction and genotyped using nested PCR as described in Figure 3.18. Genomic DNA of one of the F0 injected embryos was used as a positive control. The expected 882 bp was found in six out of ten F1 embryos. The chromatogram illustrates the outcome of the Sanger sequencing experiment performed with primer DB732 on the PCR product amplified on the embryo 6 genomic DNA sample. The mutant allele was named *qmc194*. **(C)** The remaining siblings were raised to adulthood. Caudal fin clips were taken from ten adult F1 fish. Genomic DNA was prepared from each sample and used in a nested PCR reaction. The PCR products were analysed by electrophoresis in a 1.5% agarose gel. The PCR product of the fin clip from fish number 1 was sequenced to confirm that the fish carried the mutant *c-myb*^{*qmc194*} allele. The retained TS7 sequence is shown above the double-headed arrow. The nucleotides of the PAM sequences of TS17 and TS7 are shown inside the black boxes.

In order to see whether *qmc194* homozygous fish were viable, F2 *qmc194* homozygous embryos were derived from an incross of adult fin-clipped heterozygous *qmc194* carriers. A group of 20 embryos was collected and genotyped at 24 hpf. For this purpose, gDNA was extracted and used as a template to amplify the wt and mutant alleles to identify wt, *qmc194* heterozygous and *qmc194* homozygous embryos. To this end, two PCR amplification experiments were carried out. The first PCR experiment was performed with primers DB1190 and DB1191 to generate a 271 bp PCR fragment that can only be amplified on the wt *c-myb* allele. The second PCR experiment was done to amplify the 882 bp mutant allele fragment, using the same primers pairs that were used for genotyping F0 and F1 progeny (Figure 3.20A & B). The results of the genotyping PCR experiments showed that among the 20 genotyped embryos, wt, *qmc194* heterozygous and homozygous progeny were present at roughly the expected Mendelian ratios. The rest of the F2 progeny was raised to adulthood.

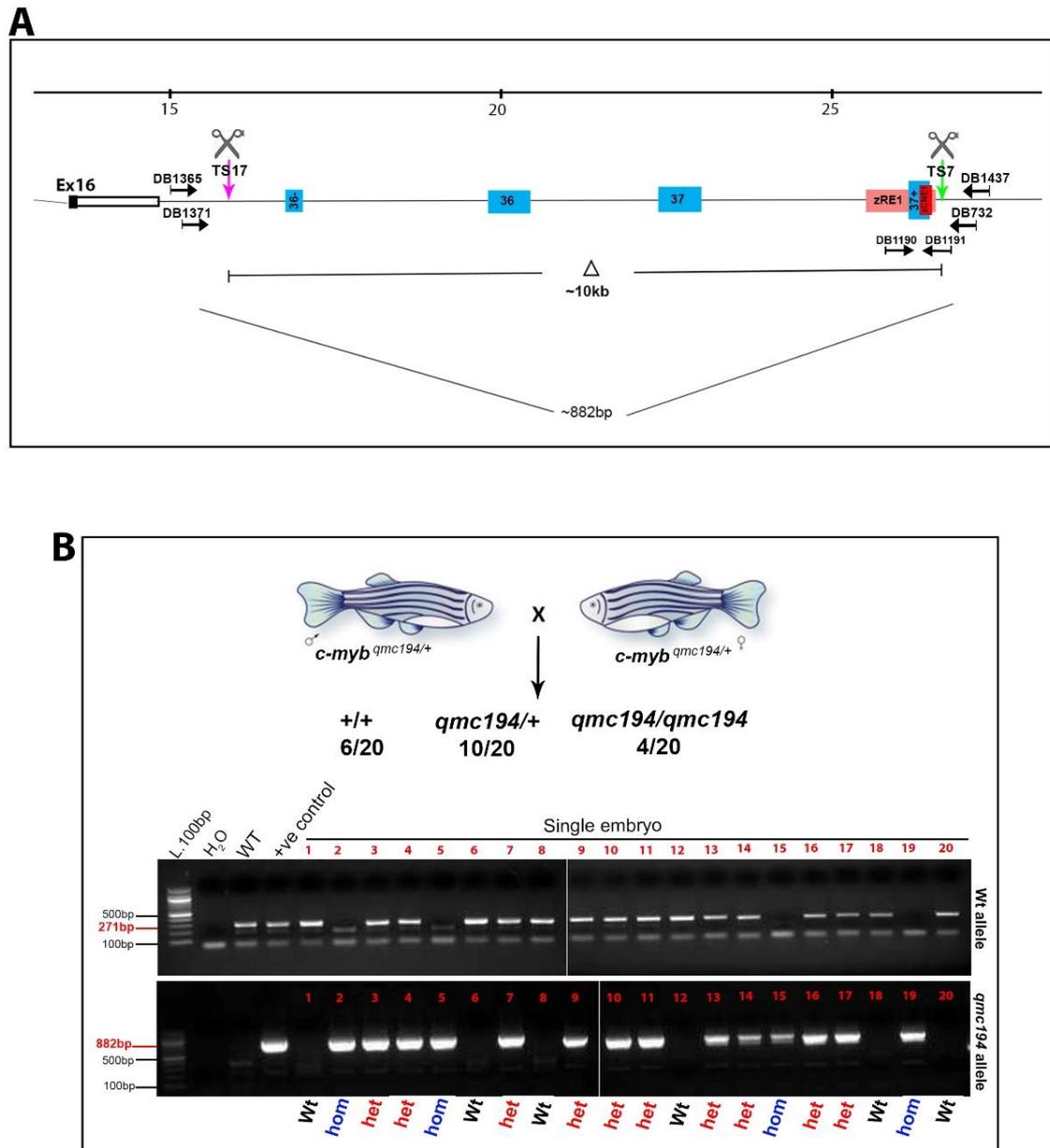


Figure 3.20 Identification of homozygous *qmc194* mutant embryos in the F2 progeny of two heterozygous *qmc194* F1 carriers

(A) Schematic illustrating the genomic DNA editing by the Cas9/TS17 and TS7 sgRNA complexes as shown previously in (Figure 3.18C). The position of the external and internal primers used in the nested PCR reaction are shown. The size of the expected mutant PCR fragment is indicated. The wt allele was identified by PCR amplification of a 271 bp fragment with primers DB1190 and DB1191. **(B)** Diagram of a $c\text{-myb}^{qmc194/+}$ fish incross. Twenty embryos were genotyped by PCR amplification of the *qmc194* mutant and the wt PCR fragments. The diagram summarises the number of wt, heterozygous and homozygous embryos identified. The images of the two agarose gels below illustrate the outcome of the wt and *qmc194* mutant allele-specific PCR experiments that were performed on the genomic DNA that was isolated from the twenty embryos at 24 hpf. The PCR products were analysed on a 1.5% agarose gel.

To check whether *qmc194* homozygous fish can survive to adulthood, siblings of the 20 embryos that were genotyped at 24 hpf were fin-clipped at ~3-4 months. Their genomic DNAs were used as templates in PCR reactions in which the wt and *qmc194* mutant DNAs were amplified as described above. Among the 26 fish analysed, 8 (31%) were wt fish, 10 (38%) were heterozygous, and 8 (31%) were homozygous for *qmc194* (Figure 3.21A), suggesting that all genotypes were present in near Mendelian ratios (χ^2 -test, p -value 0.7). Besides, and to confirm that the homozygous *qmc194* are fertile, one of the adult fish that was found to carry two *qmc194* alleles was outcrossed with wt, and then 10 embryos were genotyped to check the presence of wt and *qmc194* alleles. All the 10 progeny were identified as heterozygous for *qmc194* (Figure 3.21B). The results suggest that the homozygous *qmc194* fish survive to adulthood and are fertile. The identified fish revealed no phenotypic abnormalities.

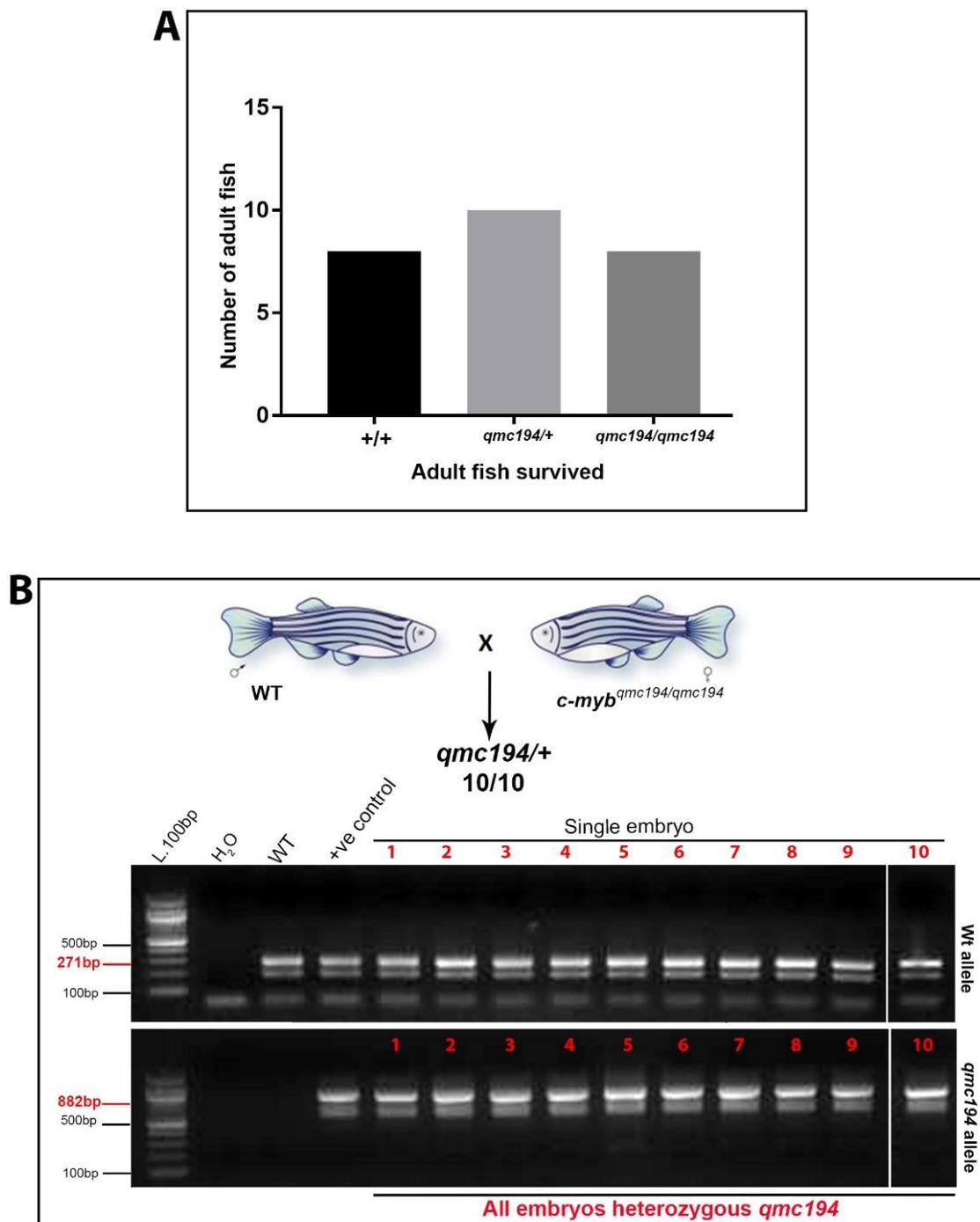


Figure 3.21 Identification of adult homozygous *qmc194* carriers in the progeny of two *qmc194* heterozygous fish

(A) Heterozygous *qmc194* F1 carriers were crossed, and the progeny was raised to adulthood. The adult F2 fish were fin-clipped and genotyped. The bar chart gives the number of *wt*, *qmc194* heterozygous and *qmc194* homozygous fish in the cohort of 26 fish. (B) The diagram illustrates an outcross of *c-myb*^{*qmc194/qmc194*} fish with a *wt* fish. The genomic DNAs of individual embryos derived from this cross were tested for the presence of *wt* and *qmc194* mutant alleles. The PCR experiments used primers DB1190 and DB1191 for the *wt* allele (top gel) and primer pairs DB1365 and DB1437 as

well as DB1371 and DB732 for the mutant allele (bottom gel). The PCR products were analysed on a 1.5% agarose gel.

Next, neutrophil development was studied in embryos homozygous for the *qmc194* allele. For this purpose, whole-mount *in situ* hybridisation experiments were carried out on embryos derived from a *qmc194* heterozygous incross. A digoxigenin-labelled antisense RNA probe against the *lyz* mRNA was used to stain these embryos at 34-36 hpf. The *lyz* gene has previously been shown to be expressed in developing neutrophils at this stage (Jin et al., 2016). WISH experiment reveals that there was no change in staining in a batch of 21 embryos that were derived from an incross of heterozygous *qmc194* carriers. Genotyping 12 embryos revealed that 3 of them were homozygous *qmc194* carriers that showed the same level of *lyz* mRNA that was present in wt and heterozygous *qmc194* siblings (Figure 3.22Ai-iii and Appendix figure 6.9). This suggests that embryonic neutrophil development is normal in the early homozygous *qmc194* embryos.

Then, WISH was performed to study the expression of *c-myb* in primitive and definitive blood cells of the developing of homozygous *qmc194* embryo obtained from a *qmc194* heterozygous incross. First, embryos were collected at 15 somites, 30 hpf and 2 dpf. These embryos were then stained using a probe against *c-myb* mRNA. The WISH experiments showed that the expression of *c-myb* was unaffected in all embryos at all 3 stages (Appendix figure 6.10, Appendix figure 6.11 and Appendix figure 6.12). To examine the level of *c-myb* more specifically in definitive blood cells of the CHT at the time when zRE1 is able to direct *c-myb* promoter activity to blood cells

in the transgenic reporter lines, an additional WISH experiment was performed on 20 embryos at 4 dpf. This experiment too, did not show any noticeable differences in the levels of *c-myb* expression between the stained embryos (Figure 3.22Bi-iii and Appendix figure 6.13). When 10 of these 20 embryos were genotyped, 2 turned out to be homozygous *qmc194* carriers, confirming that wt, *qmc194* heterozygous and homozygous embryos showed similar levels of *c-myb* mRNA expression in the CHT at 4 dpf. Thus, there was no obvious reduction in *c-myb* expression in the tested primitive and definitive blood cells in homozygous *qmc194* embryos. These results suggest that while zRE1 is sufficient to direct *c-myb* promoter activity to definitive blood cells in the CHT, zRE1 alone and in combination with the upstream PEs may not essential for the expression of the *c-myb* gene in these cells.

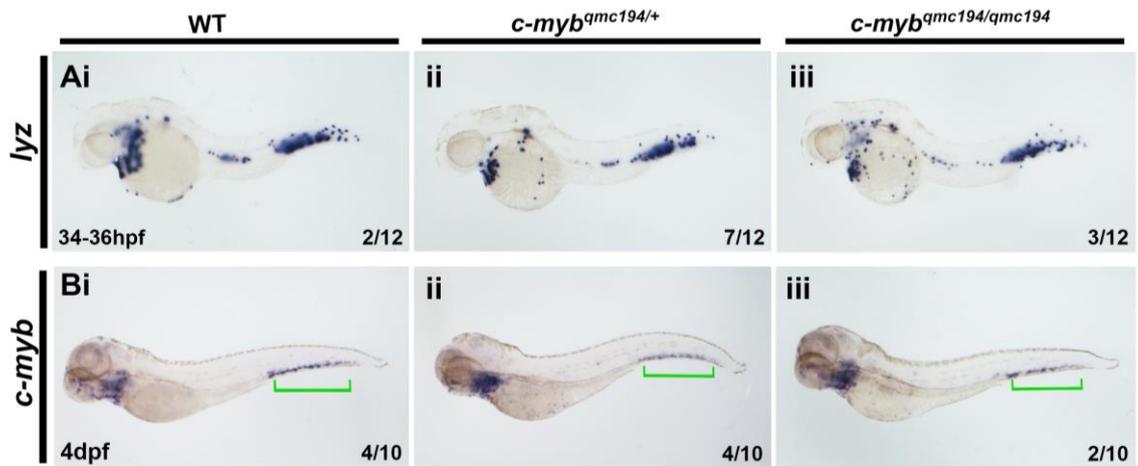


Figure 3.22 *Lyz* and *c-myb* expression are unaffected in 34-36 hpf and 4 dpf *qmc194/qmc194* embryos, respectively

Embryos were derived from an incross of *c-myb*^{qmc194/+} carriers. Embryos were collected at different stages and stained by in situ hybridisation using probes against *lyz* and *c-myb* mRNA. All embryos are presented in a lateral view with anterior facing to the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT).

3.6 *Lyz* and *c-myb* genes expression levels were unaffected in the compound heterozygous *c-myb*^{t25127/qmc193} or *c-myb*^{t25127/qmc194} mutant zebrafish embryos

The WISH experiments on the progeny of incrosses between heterozygous carriers of *qmc193* and *qmc194* fish suggested that the levels of *lyz* and *c-myb* mRNA were unaffected in blood cells of homozygous *qmc193* and *qmc194* embryos. To see whether a complete loss of one copy of *c-myb* would cause *lyz* and *c-myb* mRNA expression in the presence of the *qmc193* or *qmc194* alleles and, thereby, reveal a role of the enhancer elements, the *qmc193* and *qmc194* mutant fish were crossed to a heterozygous carrier of the loss-of-function allele *t25127* of *c-myb* (Soza-Ried et al., 2010). The expression levels of *lyz* and *c-myb* were studied in the progeny.

The *c-myb*^{t25127} mutant allele carries a thymidine/adenine (T/A) transversion at nucleotide position 15 of exon 6 of the *c-myb* locus on zebrafish chromosome 23. This mutation leads to an amino acid change from isoleucine (Ile) to asparagine (Soza-Ried et al., 2010) (Figure 3.23A). This missense mutation affects the DNA binding domain of zebrafish c-Myb protein. The mutant protein cannot interact with the c-Myb binding site on the DNA (Soza-Ried et al., 2010). In zebrafish, primitive red blood cell formation is not affected in *c-myb*^{t25127/t25127} embryos (Appendix figure 6.14, Appendix figure 6.15 and Appendix figure 6.16). By contrast, homozygous fish are incapable of surviving

to adulthood because of a loss of blood cells during definitive haematopoiesis (Hess et al., 2013).

Here, to confirm the previously published expression patterns, embryos derived from an incross of heterozygous *t25127* carriers were stained by WISH using probes against *lyz* and *c-myb* mRNA at 30-32 hpf and 4 dpf, respectively. At 30-32 hpf, *lyz* mRNA was lost in primitive neutrophil cells in *c-myb*^{t25127} homozygous embryos (Figure 3.23Bi-Bii). In addition, *c-myb* expression was greatly diminished in the caudal haematopoietic tissue (CHT) at 4 dpf in the *c-myb*^{t25127} mutants, indicating a reduction in the number of *c-myb*⁺ HSPCs (Figure 3.23Ci-Cii). By contrast, in siblings, normal expression of *lyz* and *c-myb* mRNAs were detected in neutrophils and in definitive blood cells in the CHT, respectively (Figure 3.23Bi and Ci). From 17 *lyz* and 18 *c-myb*-stained embryos, 7 embryos each were chosen for genotyping based on the level of gene expression to confirm the presence of homozygous carriers of the *c-myb*^{t25127} allele. To amplify the DNA sequence mutated in the *t25127* allele, nested PCR was performed using the external primers DB827 and DB828 in the first and the primers DB980 and DB981 in the second step to amplify a 280 bp fragment that was subsequently sequenced using primer DB1446 (Figure 3.23D). The Sanger sequencing confirmed that the two embryos with reduced staining that were chosen from each batch were the homozygous carriers of the *c-myb*^{t25127} mutation.

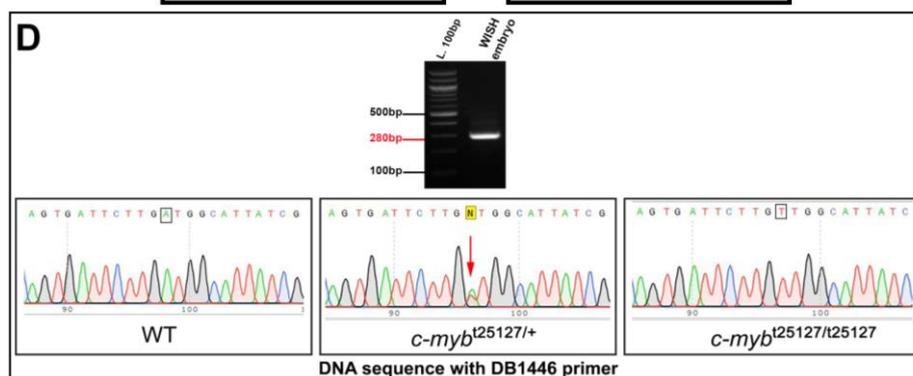
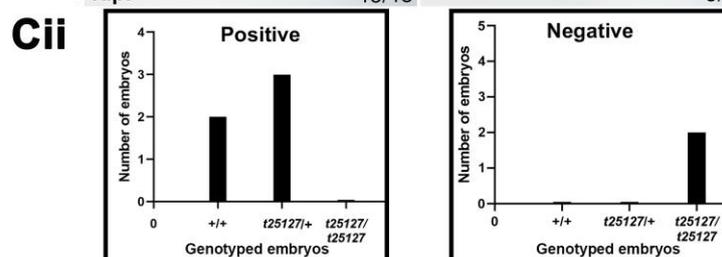
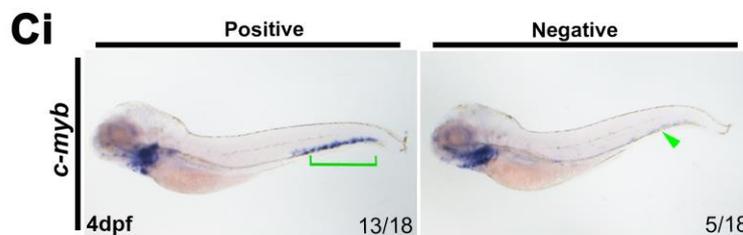
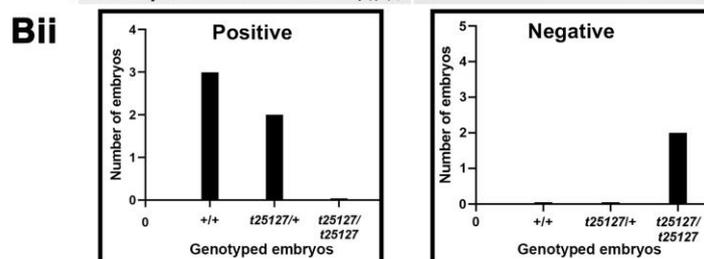
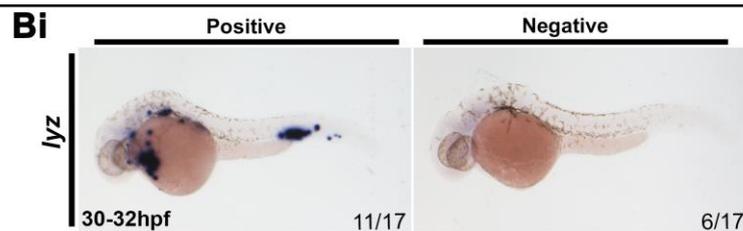
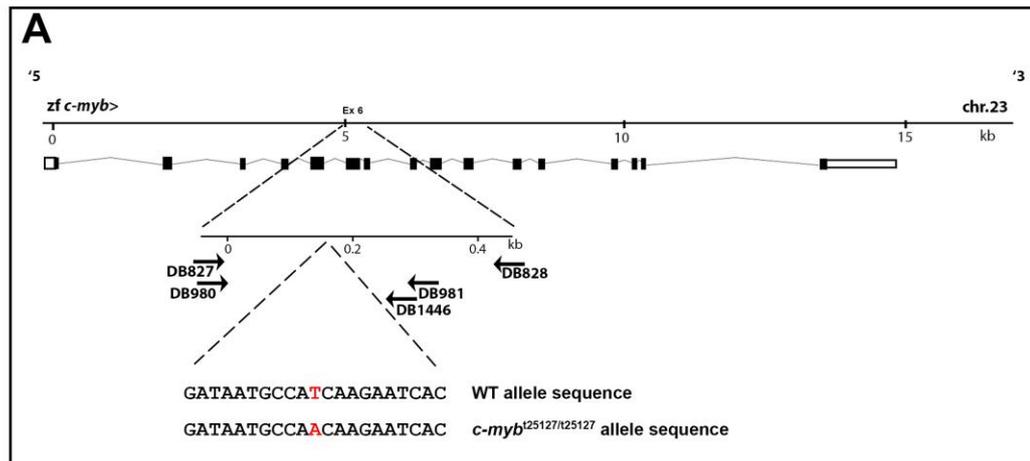


Figure 3.23 The expression of *lyz* and *c-myb* are missing in the *c-myb*^{t25127/t25127} embryos

(A) Schematic representation of the *c-myb*^{t25127} allele at the exon 6 of *c-myb* locus on zebrafish chromosome 23. The pairs of primers around the mutation site (external and internal primers (DB827- DB828 and DB980-DB981, respectively)) using to identify the *c-myb*^{t25127} mutant allele. Red letters indicate the transversion of thymidine to adenine (T>A) in *c-myb*^{t25127} allele (Soza-Ried et al., 2010). **(Bi)** Lateral view of embryos obtained from an incross of *c-myb*^{t25127} heterozygous and stained in a WISH using *lyz* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up. **(Bii)** The bars show the numbers of genotyped embryos with phenotypes over the total numbers of embryos analysed in (Bi). **(Ci)** Lateral view of embryos obtained from an incross of *c-myb*^{t25127} heterozygous and stained in a WISH using *c-myb* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up. The numbers of embryos with phenotypes over the total numbers of embryos analysed are provided on the panels. Green bracket: caudle haematopoietic tissue (CHT) and green arrowhead: missing expression in CHT. **(Cii)** The bars show the numbers of genotyped embryos with phenotypes over the total numbers of embryos analysed in (Ci). **(D)** PCR amplification of the *c-myb*^{t25127} mutant allele. The PCR product was analysed on a 1.5% agarose gel. Genotyping of some of the WISH embryos was done to distinguish wt, *c-myb*^{t25127} and *c-myb*^{t25127/t25127} alleles of *c-myb*. The panels show sanger DNA sequencing reads of the template strand of the wt, heterozygous and homozygous alleles of *c-myb*. The sequences have been read using a DB1446 primer. The transversion of T to A in *c-myb*^{t25127} allele (indicated by a small black box). Red arrow depicts the presence of T and A within the *c-myb*^{t25127} heterozygous sequence of WISH embryos.

In addition to the *lyz* gene, the expression of two additional neutrophil markers was analysed at 36 hpf. These markers were the *srgn* and *mpx* genes. The WISH experiments revealed that 4 out of 18 embryos stained with a probe against *srgn* mRNA had lost anterior *srgn* expression in cells over the yolk cell. The same embryos also displayed reduced diffuse staining in the PBI, similar to what had previously been reported by (Jin et al., 2016) (Figure 3.24). Similarly, 4 out of 20 embryos displayed reduced levels of *mpx* expression. Their expression levels were also drastically decreased in *c-myb*^{t25127} mutant embryos. In addition, the expression of the lymphoid marker *rag-1* was absent in the developing thymus in *c-myb*^{t25127} mutant embryos at 5 dpf (Figure 3.24). These differences between siblings and *c-myb*^{t25127} homozygous mutant embryos were consistent with previously published data (Soza-Ried et al., 2010, Zhang et al., 2011).

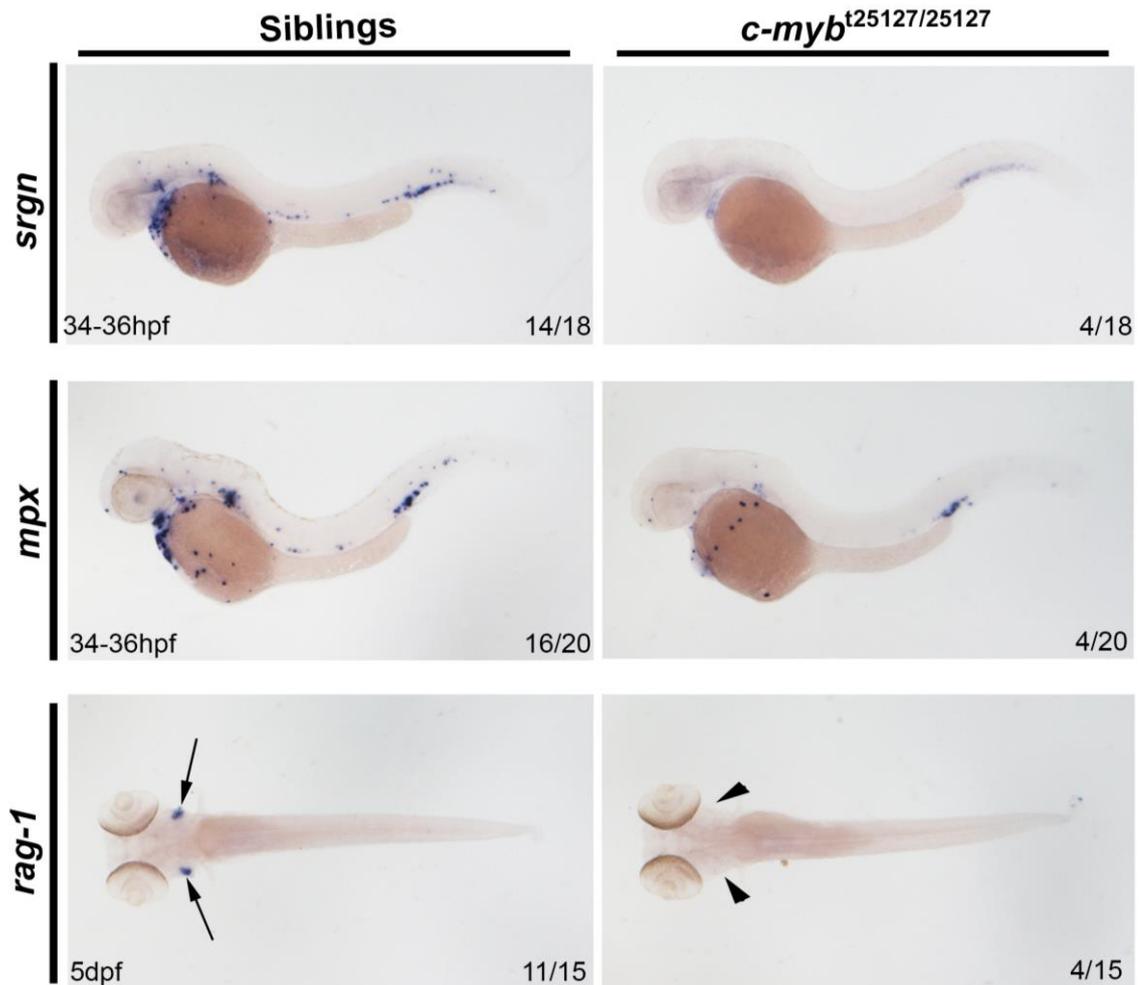


Figure 3.24 Reduced /or loss of the expression of some genes in the *c-myb*^{t25127/t25127}

Lateral view of embryos obtained from an incross of *c-myb*^{t25127} heterozygous and stained in whole-mount RNA in situ hybridizations (WISH) using different probes of markers. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Black arrow: thymus and black arrowhead: missing expression in thymus. The numbers of embryos with phenotypes over the total numbers of embryos analysed are provided on the panels.

3.6.1 *Lyz* and *c-myb* genes expression levels were normal in the compound heterozygous *c-myb*^{t25127/qmc193} mutant zebrafish embryos

In order to investigate whether the deletion of zRE1 in *qmc193* has a significant effect on blood cell development in embryos that also carry the *c-myb*^{t25127} loss-of-function allele, the expression of *lyz* and *c-myb* were examined by WISH in compound heterozygous *c-myb*^{t25127/qmc193} zebrafish embryos. Embryos were obtained by crossing a *qmc193* heterozygous carrier to a *t25127* heterozygous zebrafish (Figure 3.25A). After WISH staining, embryos displayed variable levels of *lyz* expression at 34-36 hpf (Figure 3.25Bi and Appendix figure 6. 17). Eleven of the embryos displayed high levels of *lyz* mRNA. Nine embryos showed lower levels of staining. Genotyping 4 of the well and 3 of the weakly stained embryos showed that there was no clear correlation between the observed phenotype and the genotype of the embryos (Figure 3.25Bii).

Next, *c-myb* expression was examined at 3 dpf in 15 embryos obtained from the same previous cross using WISH analysis. The WISH experiment shows that all embryos showed staining in the CHT (Figure 3.25Ai and Appendix figure 6.18). In 12 embryos, this staining was higher than in the remaining 3 embryos. Genotyping 6 of the strongly stained and all weakly stained embryos revealed that there was no clear correlation between the phenotype and the genotype (Figure 3.25Aii). Both results suggest that the *lyz* and *c-myb* expression were

maintained at relatively normal levels even in those embryos that carried the two *c-myb* alleles *qmc193* and *t25127*. The differences in staining that were observed in the WISH experiments may be due to technical issues or may be caused by stage differences between embryos. Despite, all embryos were collected at the desired time points using morphological features under a light microscope (Kimmel et al., 1995), some of these embryos developed at different embryonic stages.

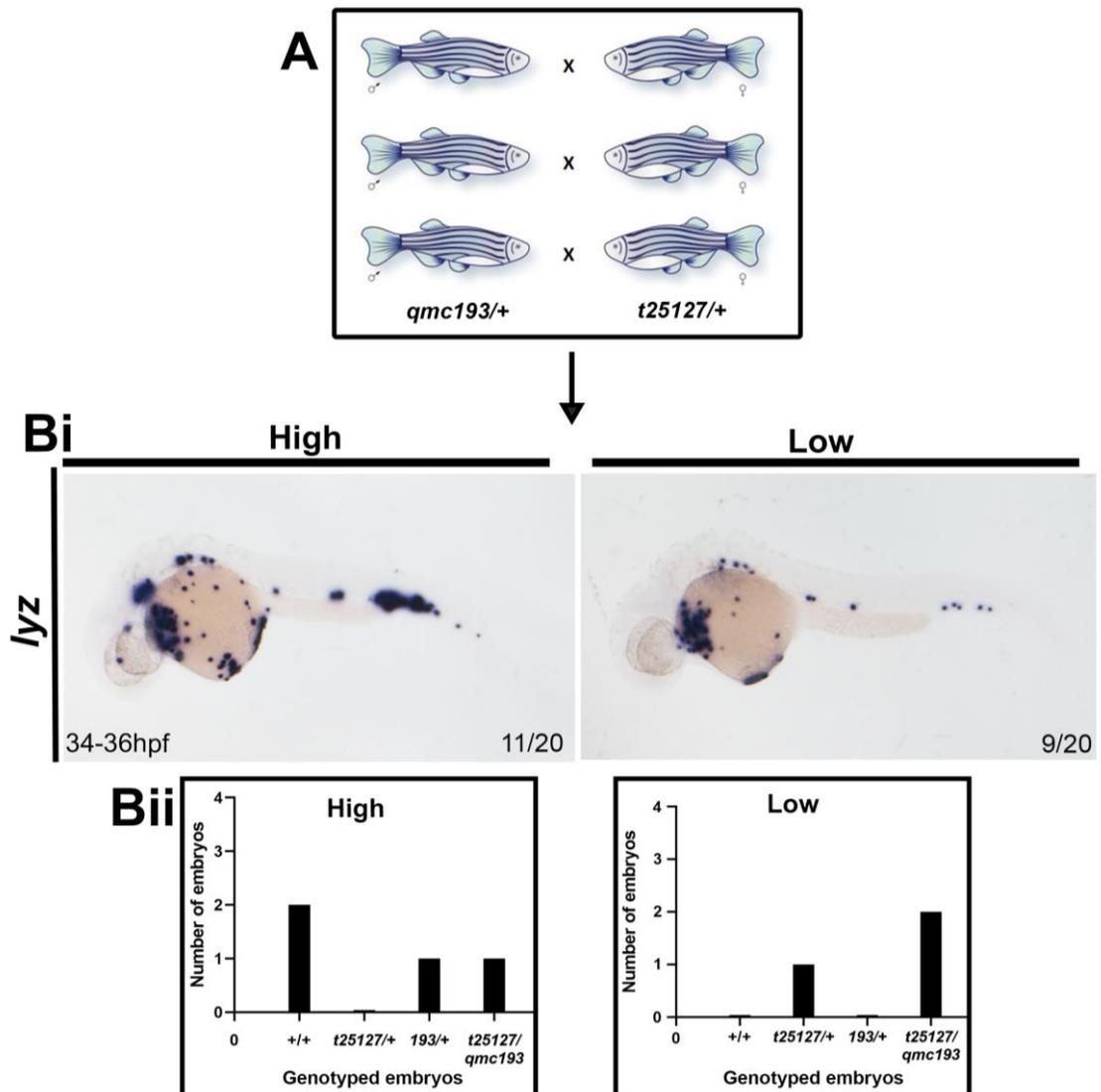


Figure 3.25 *Lyz* expression was maintained in 34-36 hpf in *c-myb*^{t25127/qmc193} embryos

(A) A crossing of a three *qmc193* heterozygous carrier males to a three *t25127* heterozygous females zebrafish using breeding beach style tank. (Bi) Lateral view of embryos obtained from (A) stained in a WISH using *lyz* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up. (Bii) The bars show the numbers of genotyped embryos with phenotypes over the total numbers of embryos analysed in (Ai).

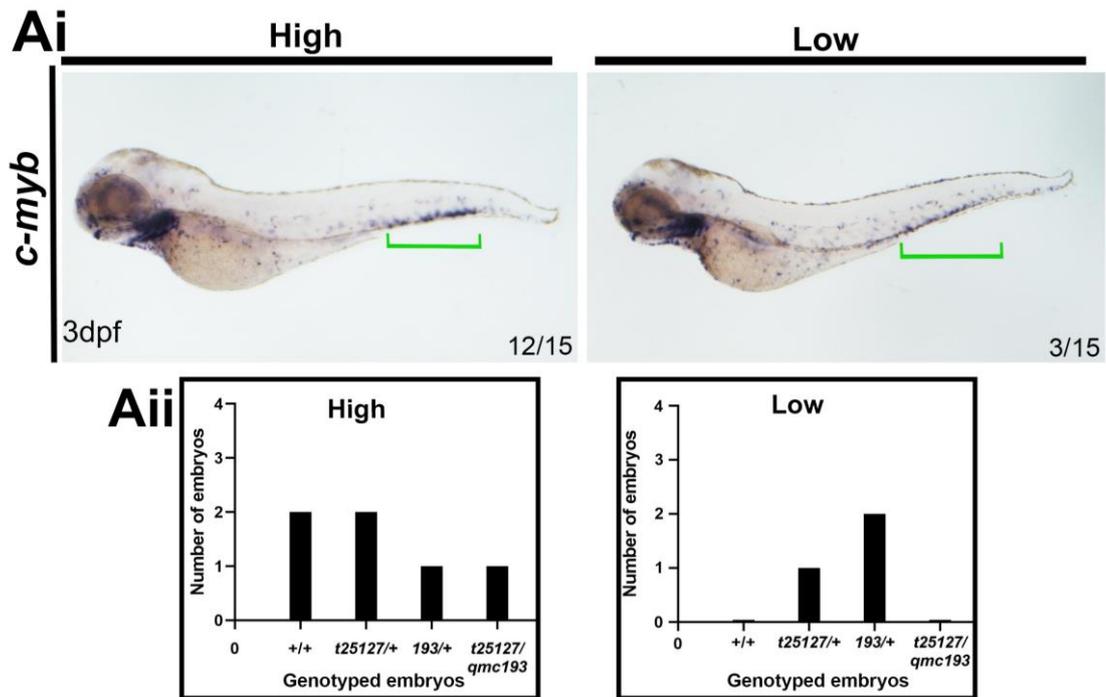


Figure 3.26 *c-myb* expression was maintained in 3 dpf *c-myb*^{t25127/qmc193} embryos

(Ai) Lateral view of embryos obtained from outcrosses of *c-myb*^{t25127} and *c-myb*^{qmc193} heterozygous stained in a WISH using *c-myb* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT). **(Aii)** The bars show the numbers of genotyped embryos with phenotypes over the total numbers of embryos analysed in (Ai).

3.6.2 The expression level of *lyz* and *c-myb* genes were unchanged in the compound heterozygous *c-myb*^{t25127/qmc194} mutant zebrafish embryos

In order to find out whether the deletion of all potential enhancers MP40036-, MP40036, MP40037 and MP40037+ in *qmc194* have a notable effect on blood cell development in embryos that also carry the *c-myb*^{t25127} loss-of-function allele, the expression of *lyz* and *c-myb* were examined by WISH in a compound heterozygous *c-myb*^{t25127/qmc194} zebrafish embryos. Embryos were obtained by crossing a *qmc194* heterozygous carrier to a *t25127* heterozygous zebrafish (Figure 3.27A). *Lyz* expression was examined at 34-36 hpf using WISH analysis. WISH-stained embryos presented variable levels of *lyz* in a batch of 20 embryos (Figure 3.27Bi and Appendix figure 6.19). Eleven embryos were presented with high levels of *lyz* mRNA. Nine embryos displayed reduced levels of staining. Genotyping 4 of the highest and 3 of the weakest stained embryos showed that there was no obvious correlation between the detected phenotype and the genotype of the embryos (Figure 3.27Bii).

Next, the *c-myb* expression was analysed at 4 dpf in 17 embryos which were obtained from the same previous cross using the WISH experiment. The WISH analysis revealed that all embryos showed staining in the CHT (Figure 3.28Ai and Appendix figure 6.20). Within 17 embryos, 12 embryos were slightly higher staining than in the remaining 5 embryos. Genotyping all embryos indicate that there was

no clear correlation between the phenotype and the genotype (Figure 3.28Aii). Both results suggest that the expression of *lyz* and *c-myb* were unaffected at relatively normal levels even in those embryos that carried both the *qmc194* and *t25127 c-myb* alleles. The high number of *qmc194* and compound heterozygous, *c-myb*^{t25127/qmc194} embryos that were present in the stained embryos could be due to the way that the embryos were randomly picked from the breeding beach style tank.

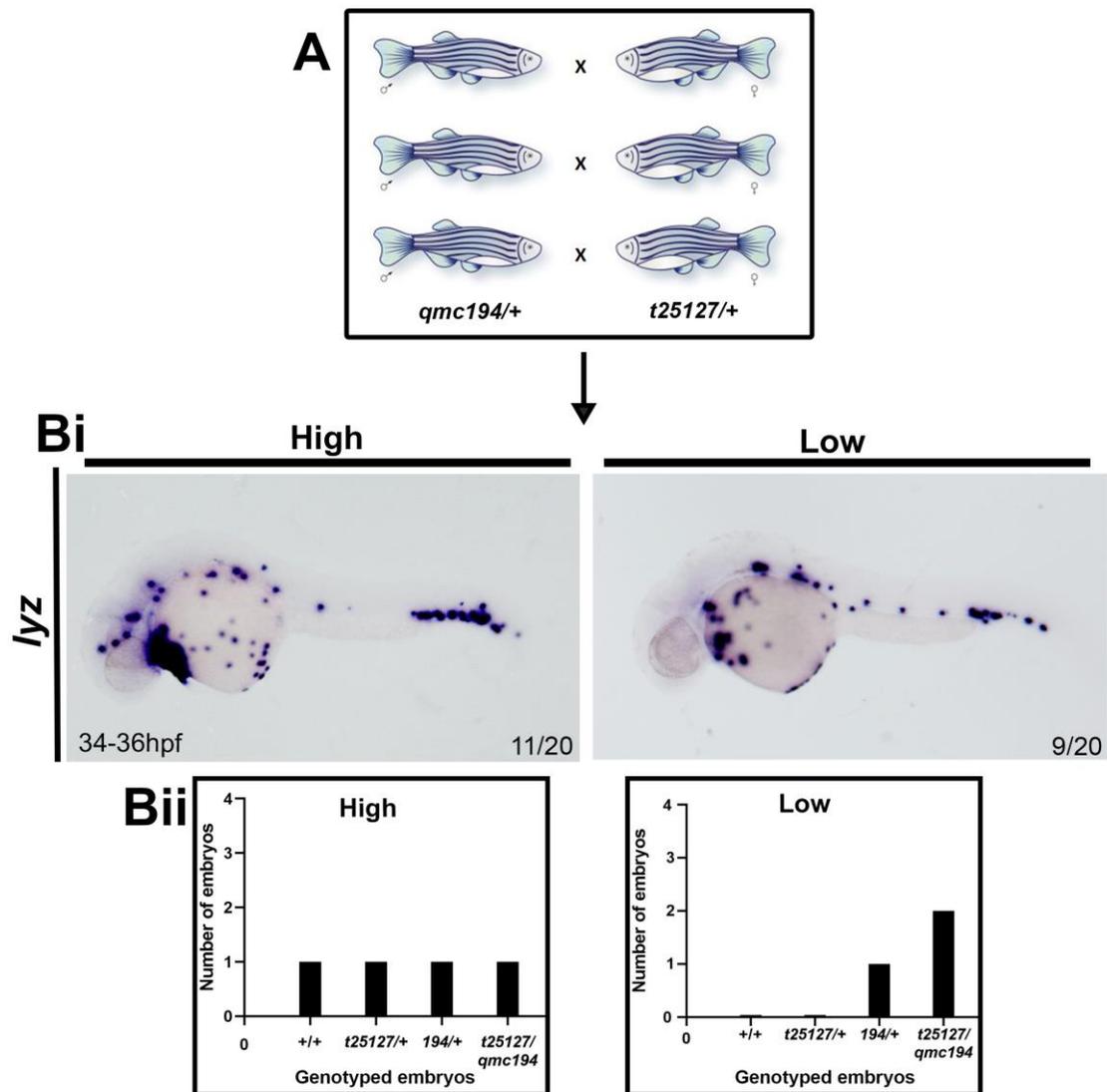


Figure 3.27 The expression level of *lyz* in 34-36 hpf was not changed in the compound heterozygote of *c-myb*^{t25127/qmc194} embryos

(A) A crossing of a three *qmc194* heterozygous carrier males to a three *t25127* heterozygous females zebrafish using breeding beach style tank. **(Bi)** Lateral view of embryos obtained from (A) stained in a WISH using the *lyz* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up. **(Bii)** The bars show the numbers of genotyped embryos with phenotypes over the total numbers of embryos analysed in (Ai).

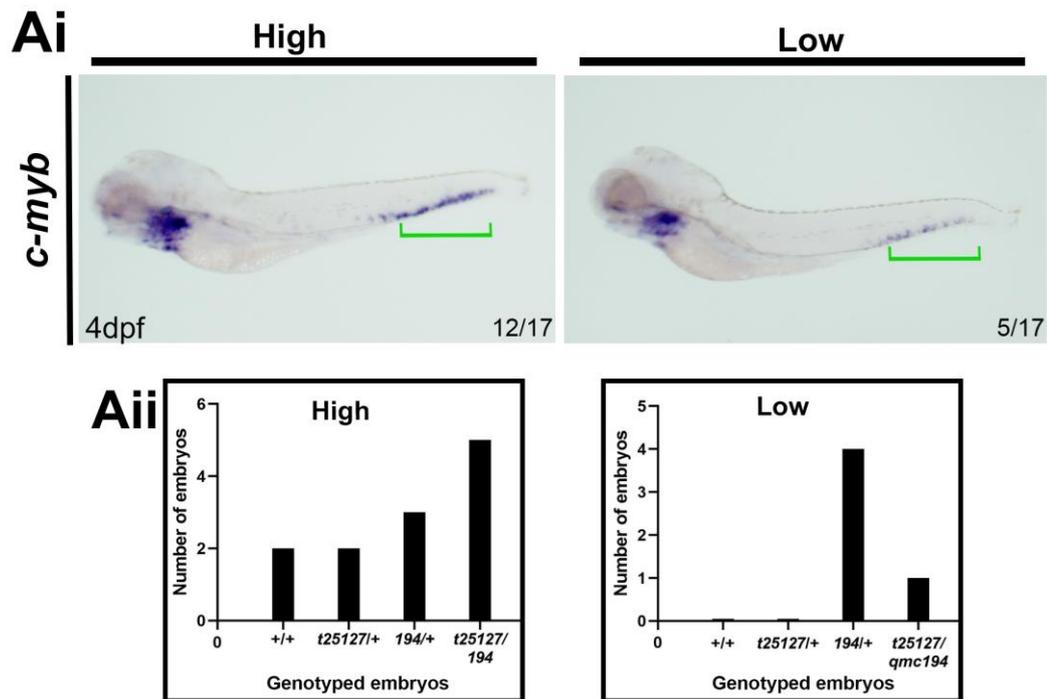


Figure 3.28 The expression level of *c-myb* at 4 dpf was not altered in the compound heterozygote of *c-myb*^{t25127/qmc194} embryos

(Ai) Lateral view of embryos obtained from outcrosses of *c-myb*^{t25127} and *c-myb*^{qmc194} heterozygous stained in a WISH using the *c-myb* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT). **(Aii)** The bars show the numbers of genotyped embryos with phenotypes over the total numbers of embryos analysed in (Ai).

3.7 Human regulatory element one (hRE1) is able to drive GFP expression in the zebrafish adult kidney

The regulatory element one of zebrafish (zRE1) and mouse (mRE1) drives the expression of GFP in the caudal haematopoietic tissue (CHT) at 3 dpf in zebrafish embryos and the adult kidney marrow when combined with the zebrafish *c-myb* promoter fragment five (PF5). It was interesting to find out whether the human regulatory element that includes the human conserved non-coding element one (hCNE1) can also direct GFP expression in haematopoietic cells in zebrafish embryos and adult kidney marrow cells. A 1 kb size of the human regulatory element (hRE1) sequence was tested in this study in order to compare it with the zRE1 of a previous study (Hsu, 2010) where a sequence of 1 kb size including the zCNE1 was able to direct the expression of GFP in haematopoietic tissues when combined with the zebrafish *c-myb* promoter. The investigation of regulatory information deposited in ENSEMBL GRCh 37.p13 revealed that a DNaseI hypersensitive site (DNaseI HS) had previously been identified in the human myelogenous leukaemia cell line K562 (<https://www.encodeproject.org/annotations/ENCSR611VZD/>) that is located downstream of the human *C-MYB* gene and overlapped with hCNE1 (Figure 3.29A & B). K562 cells have some proteomic similarity to both undifferentiated granulocytes and erythrocytes (Aktuna, 2018, Andersson et al., 1979). Therefore, primers DB1216 and DB1217 were designed to amplify a DNA fragment called hRE1 that included both the

hCNE1 and the DNaseI HS sequences in a 1 kb sequence (Figure 3.29B).

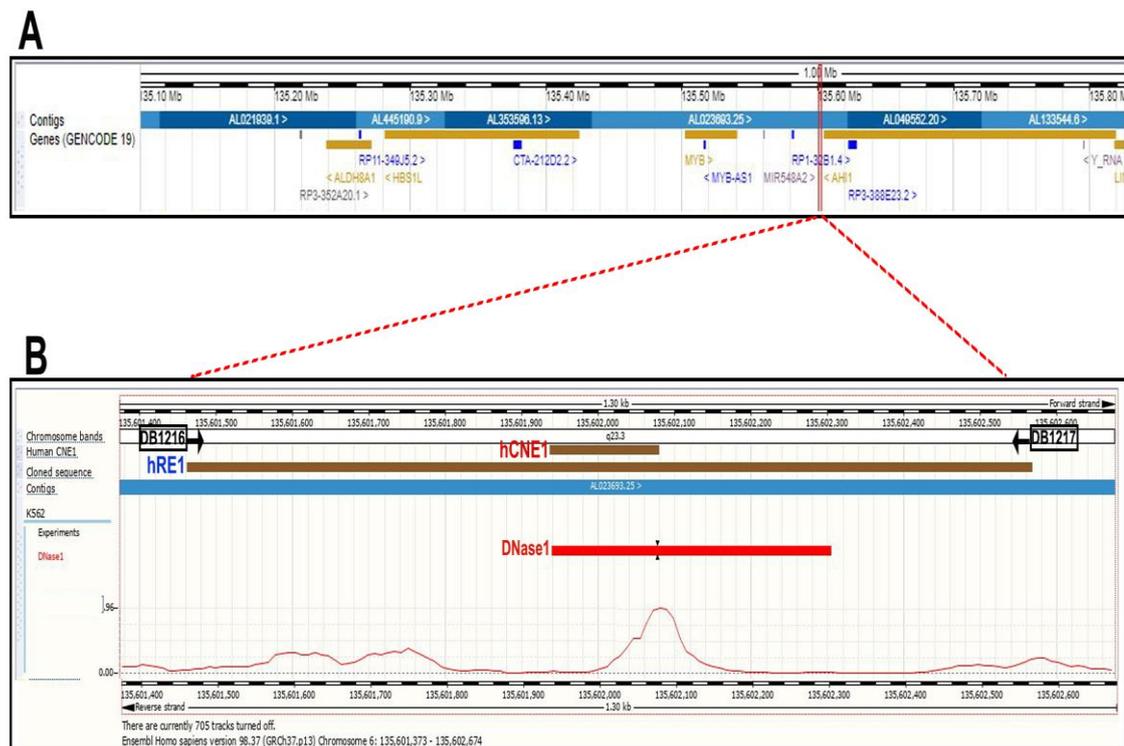


Figure 3.29 Localisation of the conserved non-coding element one downstream of human *C-MYB* gene.

(A) The panel shows an overview of human *C-MYB* (protein-coding) that is depicted as a golden bar along chromosome 6. The vertical red line presents the location of the interesting region downstream of human *C-MYB* gene. **(B)** A close up of (A). The diagram depicts the human conserved non-coding element one (hCNE1) within human regulatory element one (hRE1) on chromosome 6 in human, as shown on the ENSEMBL database with custom tracks. It refers that it contains a DNaseI hypersensitive site (red box) within the human genome. Ensembl Homo sapiens version 98.37 (GRCh 37.p13) Chromosome 6: 135,601,373-135,602,674.

3.7.1 Establishment of stable human regulatory element one (hRE1) zebrafish transgenic line

To test whether the newly generate transgene *cmyb*-hRE1-PF5:*egfp* showed expression in haematopoietic cells in zebrafish, plasmid pJHE1 was co-injected with *Tol2* transposase mRNA into one-cell stage wt zebrafish embryos (Figure 3.30A). The next day, injected and uninjected embryos were examined under a fluorescence microscope for mosaic GFP fluorescence in their developing tissues. Fluorescent images of injected and uninjected embryos were taken between 1-3 dpf (Figure 3.30B-E). Besides auto-fluorescent pigment cells that were present in both injected and uninjected embryos, injected embryos displayed green fluorescence in cells of the retina and in the olfactory placodes, tissues in which promoter fragment PF5 drives reporter gene expression. At 3 dpf, green fluorescence was also detected in ectopic places such as the muscle fibres (Figure 3.30D & E). Embryos that displayed transient GFP expression were grown up to adulthood in the hope that they would carry the Tg(*c-myb*-hRE1-PF5:*egfp*) transgene in their germline.

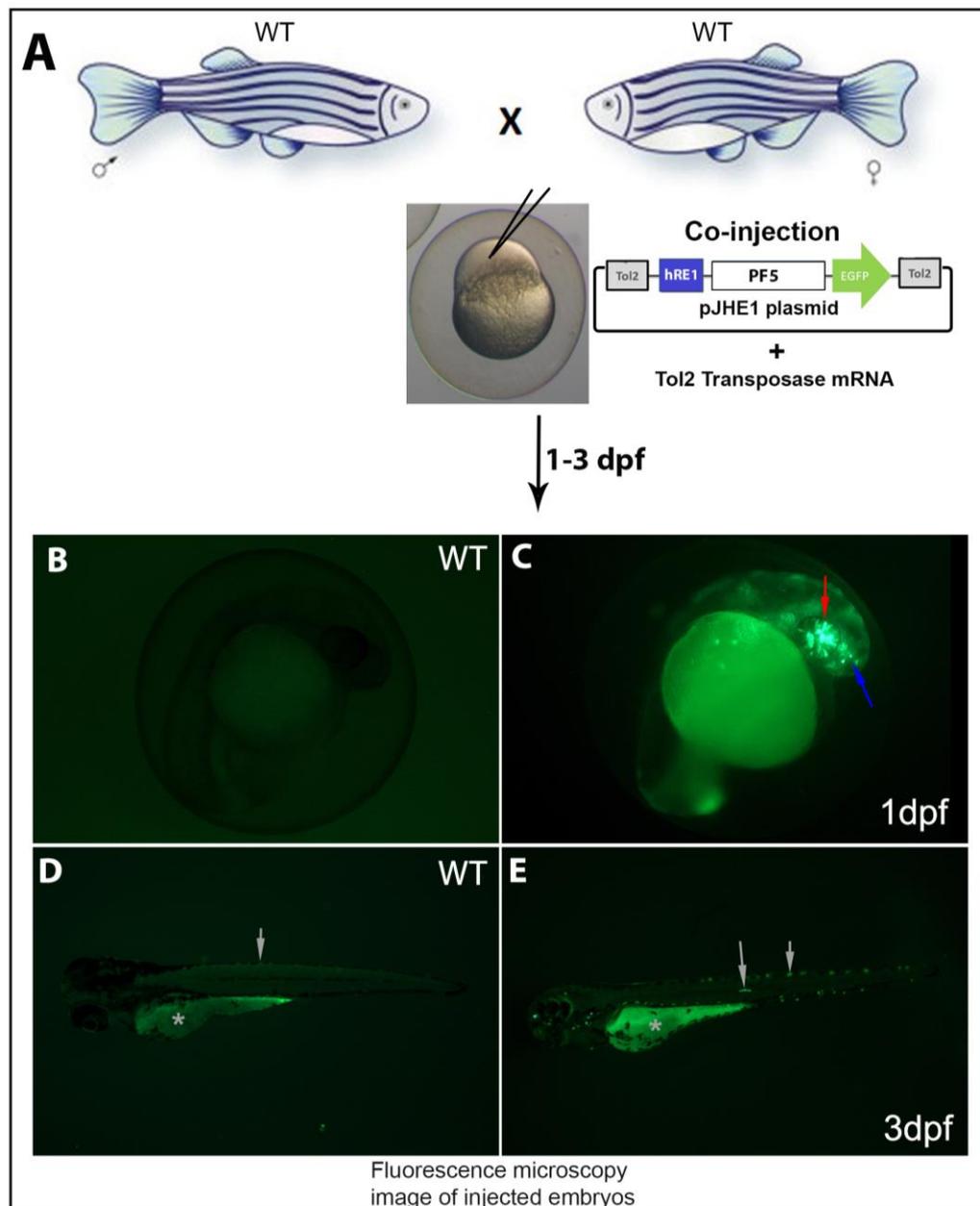


Figure 3.30 Establishment of the stable human regulatory element one (hRE1) zebrafish transgenic line by injecting the Tol2 pJHE1-egfp

(A) Wild-type (wt) zebrafish were incrossed and one-cell stage zebrafish embryos were co-injected with the pJHE1 DNA construct and the *Tol2* transposase mRNA. At 1-3 days post fertilisation (dpf), normal embryos were then analysed under a fluorescent microscope for GFP fluorescent protein. (B and D) Show WT control uninjected embryo at 1-3 dpf. These are siblings of the injected embryo in C and D. (C) Depict the expression of transient GFP in olfactory placodes (OP) (blue arrow), the retina (Re) (Red arrow). (D and E) At 3 dpf, auto-fluorescence was observed in the pigment cells (grey arrowheads) and in the yolk cell (grey asterisk) in the uninjected controls (D) and in injected embryos (E).

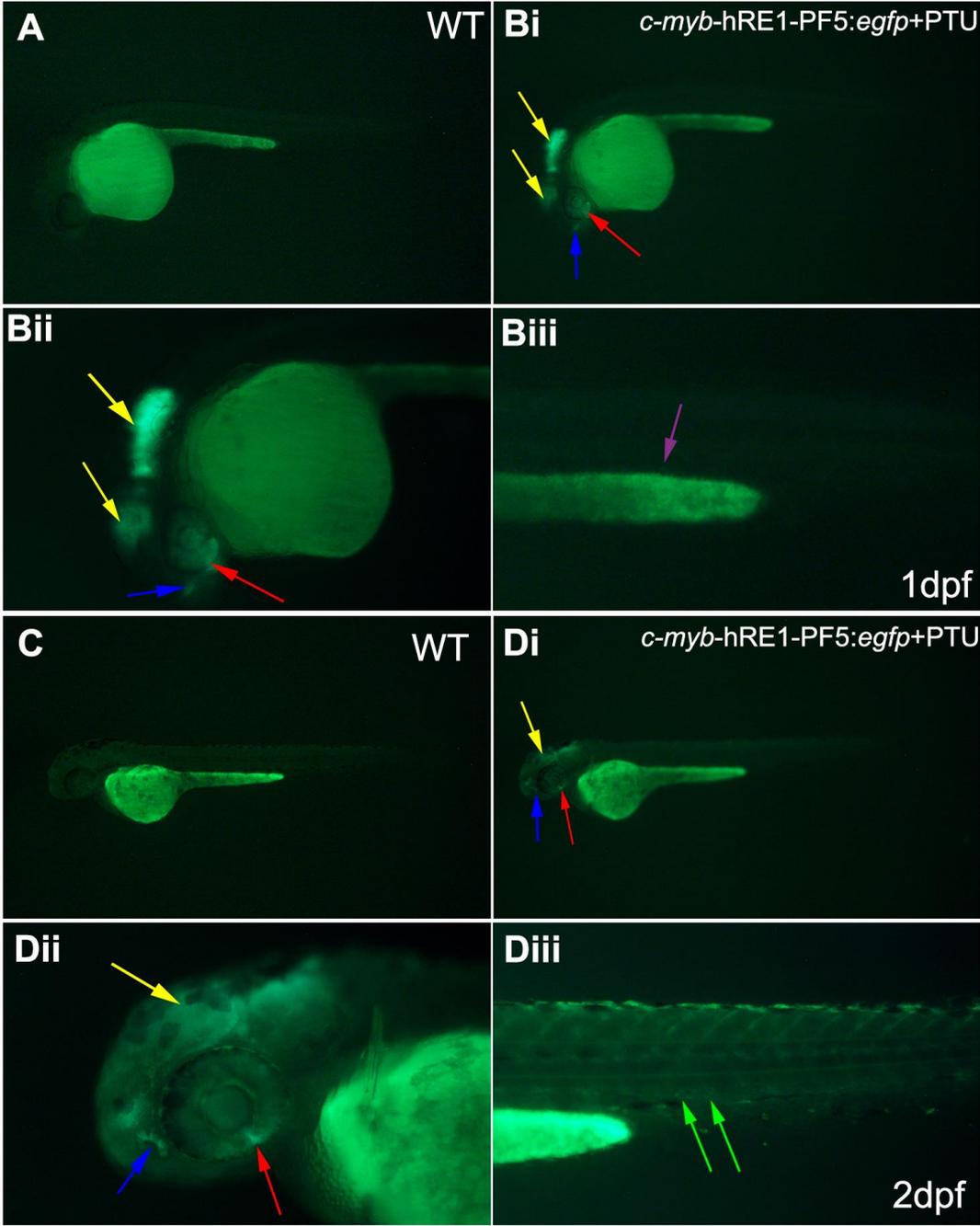
3.7.2 Zebrafish embryos that carry the Tg(c-myb-hRE1-PF5:egfp)qmc195 transgene display GFP reporter gene expression in cells of the caudal haematopoietic tissue

To verify germline transmission of the *c-myb*-hRE1-PF5:*egfp* transgene, adult fish that had successfully been injected with the transgene construct were crossed with wild-type fish. Embryos derived from these crosses (non-transgenic and transgenic) were examined for GFP expression under the fluorescent microscope in 1-4 days post fertilisation (Figure 3.31A-Hiii). Out of 24 potential founders successfully crossed with wt fish, 4 transgenic founders were identified. These were founders of the *qmc191*, *qmc195*, *qmc196* and *qmc197* lines (Table 3.2). On days 1 to 4, transgenic embryos were identified based on the PF5-driven GFP expression pattern, i.e., the expression in the mid/hindbrain boundary (MHB), the retina (Re), the olfactory placodes (OP), the branchial arches (BA) and the pronephric ducts (PND). Interestingly, in one of these lines, the *qmc195* line, weak GFP fluorescence was also observed in cells of the CHT of these embryos by 2 dpf (Figure 3.31Diii), suggesting that hRE1 is able to enhance PF5 activity in definitive zebrafish blood cells. At 3-4 dpf, GFP expression was still obvious in the CHT, albeit not at the same level (Figure 3.31Fiii and Hiii). Noticeably, the level of the GFP expression in the CHT was found to be low and extremely variable between embryos, suggesting that the embryos carried a variable number of transgene copies.

Table 3.2 Identified Tg(*c-myb*-hRE1-PF5:*egfp*) transgenic founders

<i>c-myb</i> -hRE1-PF5: <i>egfp</i> Identified founders									
Founder Name	GFP ⁺ embryos	Level of expression	Retina GFP ⁺	Olfactory GFP ⁺	Brain GFP ⁺	Pronephric ducts GFP ⁺	CHT GFP ⁺	Branchial arches GFP ⁺	Line established
qmc191*	13	Low	Yes	Yes	Yes	No	No	Yes	No
qmc195	28	Strong	Yes	Yes	Yes	Yes	Yes (weak)	Yes	Yes
qmc196*	60	Low	Yes	Yes	Yes	No	No	Yes	No
qmc197*	55	Extremely low	Yes	Yes	Yes	No	No	Yes	No

*(see appendix 5.20)



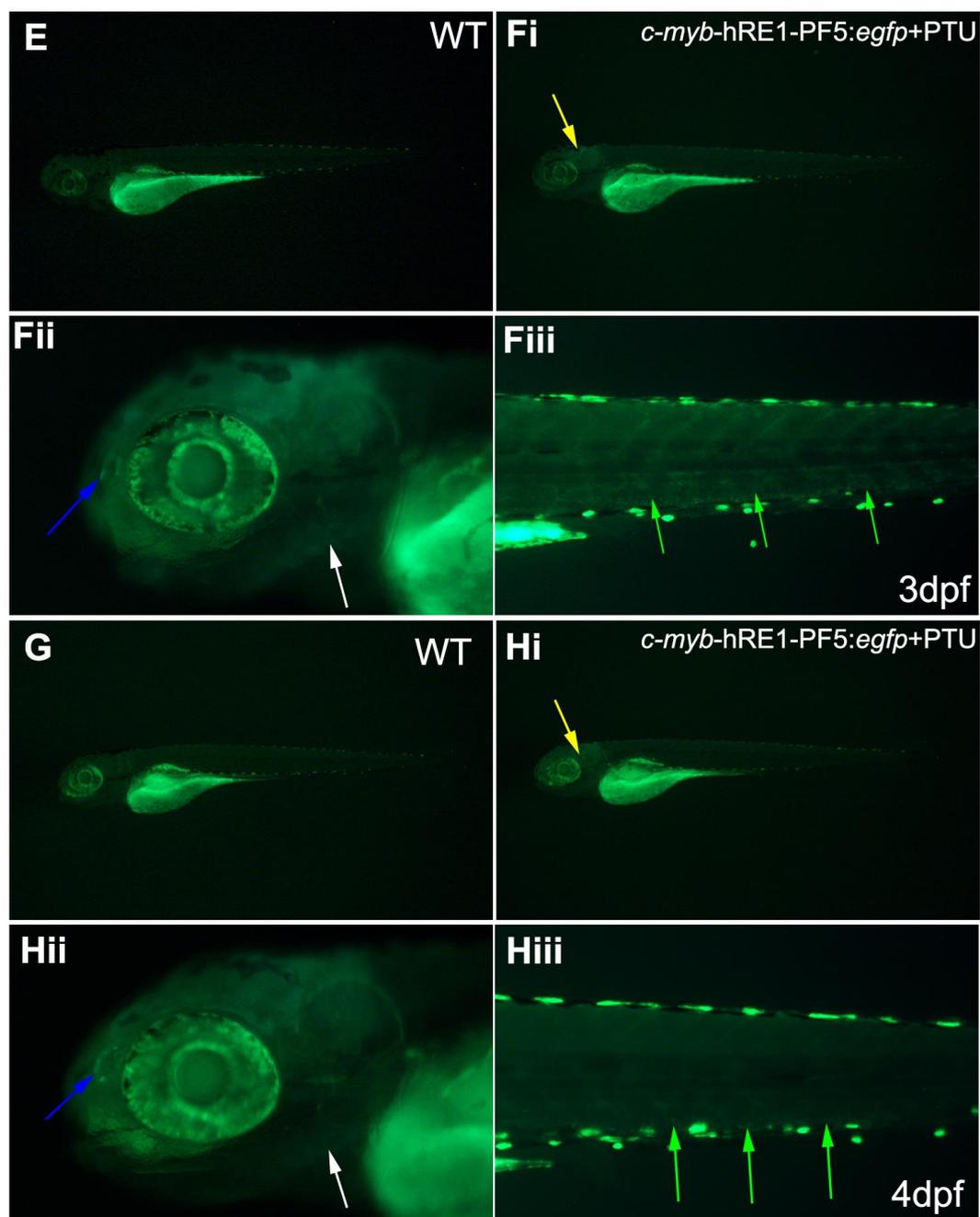


Figure 3.31 Expression of EGFP reporter gene in in the caudal haematopoietic tissue (CHT) of 2-day old embryos of *Tg(c-myb-hRE1-PF5:egfp)^{qmc195}* transgene

(A) Lateral view of non-transgenic sibling of the *Tg(c-myb-hRE1-PF5:egfp)^{qmc195}* embryo shown in Bi-iii. Anterior is to the left, posterior to the right, dorsal is up. **(Bi-iii)** Embryos were characterised as transgenic based on the GFP expression in the retina (Re) (red arrow), the olfactory placode (OP) (blue arrow) and the mid/hindbrain boundary (MHB) (yellow arrow). This GFP expression is driven by the PF5 promoter fragment. Weak expression in the pronephric ducts (PND) (purple arrow) was observed at 1 dpf. **(C)** Control non-transgenic embryo at 2 dpf. **(Di-iii)** A transgenic embryo shows reporter expression in the OP, Re and MHB at 2 dpf. The

expression of the reporter gene can be slightly observed in caudal haematopoietic tissue (CHT) (green arrow). **(E)** Control non-transgenic embryo at 3 dpf. **(Fi-iii)** A transgenic embryo displays GFP expression in OP, BR, MHB and the branchial arches (BA) (white arrow) at 3 dpf. The expression of the reporter gene can be seen in the CHT region at a low level. **(G)** The image depicts the control non-transgenic embryo at 4 dpf. **(Hi-iii)** Embryo displaying expression in the OP, MHB and CHT at 4 dpf.

In order to confirm that GFP expression was observed in the CHT of the $Tg(c-myb-hRE1-PF5:egfp)^{qmc195}$ in live transgenic embryos, WISH experiments with an *egfp* mRNA probe were performed on wt (Figure 3.32Ai-iii), *qmc85* (Figure 3.32Bi-iii) and *qmc195* embryos (Figure 3.32Ci-iii), at 3 dpf. GFP mRNA was detected in the retina and in the olfactory placode of transgenic embryos of the *qmc85* and *qmc195* lines but not in any of the wt embryos. In addition, it was noticed that GFP mRNA was also present in cells of the caudal haematopoietic tissue of *qmc85* (Figure 3.32Biii) and *qmc195* transgenics (Figure 3.32Ciii) at 3 dpf. This suggests that the green fluorescence observed in the cells of the CHT of $Tg(c-myb-hRE1-PF5:egfp)^{qmc195}$ transgenic embryos was indeed due to the presence of GFP protein.

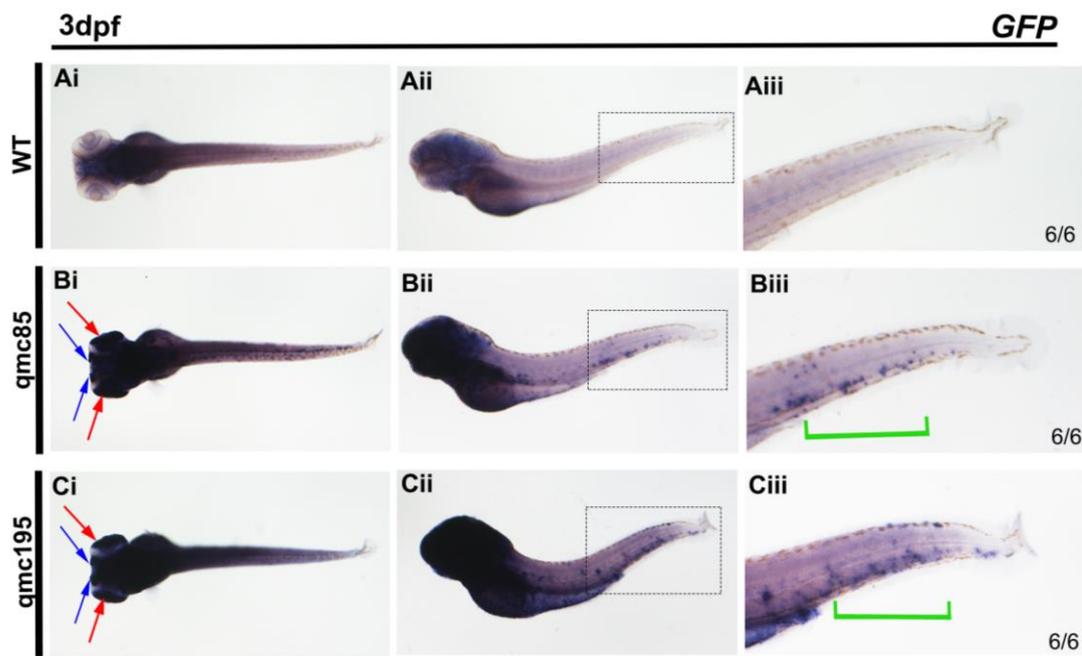


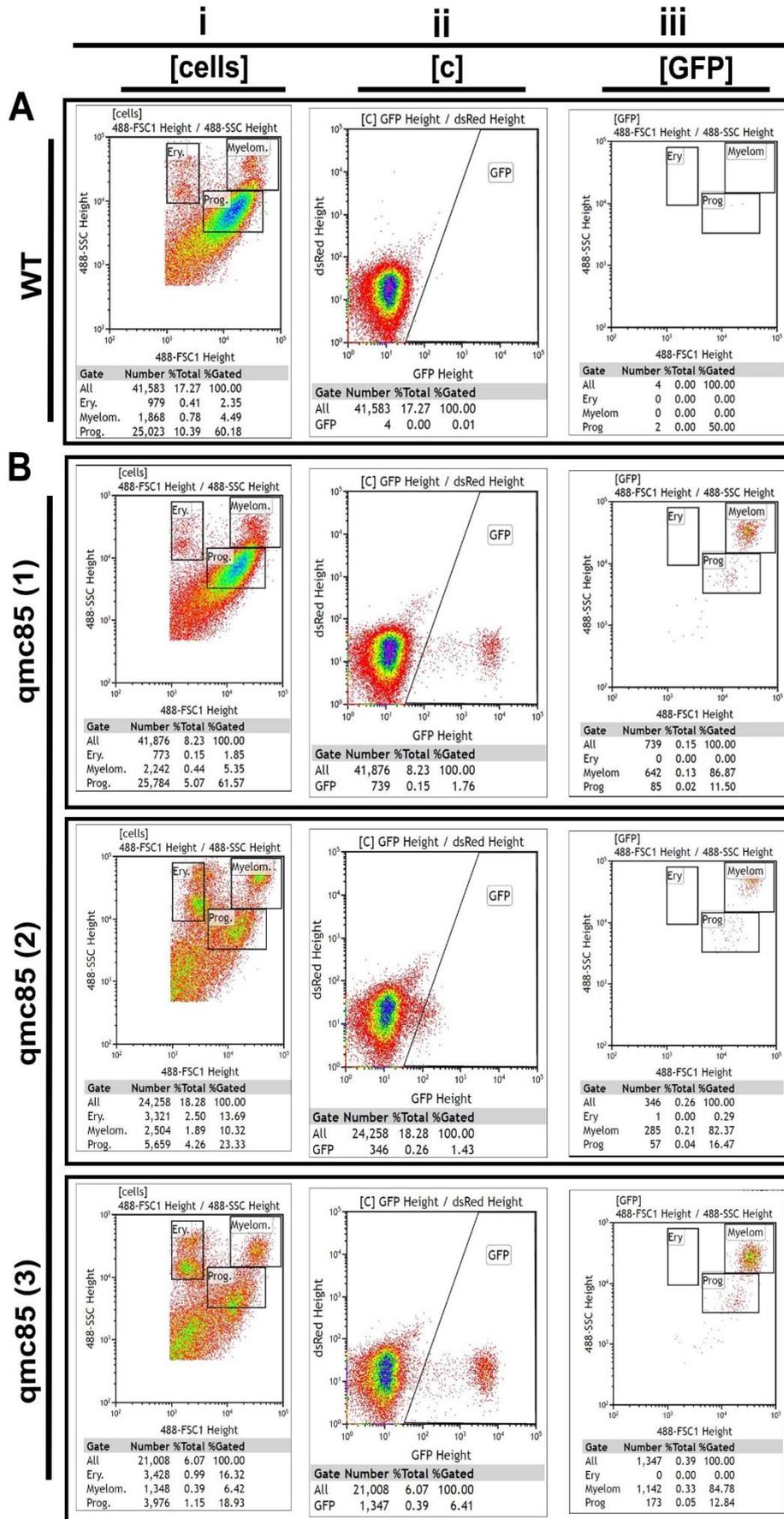
Figure 3.32 Expression of *GFP* mRNA in CHT of *Tg(c-myb-hRE1-PF5:egfp)^{qmc195}* transgenic zebrafish embryos. (A-D)

Whole mount in situ hybridisation on zebrafish embryos at 3 dpf. Overstaining in the head regions of these embryos was needed to get the nice staining in the CHT region. **(Ai-Ci)** Embryos are present in dorsal views and **(Aii-Cii)** in lateral views with anterior to the left and dorsal up. **(A)** Show non-transgenic embryo as a negative control. **(Bi and Ci)** *GFP* mRNA expression in non-haematopoietic tissues was observed in embryos carrying the *Tg(c-myb-zRE1-PF5:egfp)^{qmc85}* and the *Tg(c-myb-hRE1:PF5:egfp)^{qmc195}* transgenes in olfactory placodes (OP) (blue arrow) and retina (Re) (red arrow). **(Bii)** *GFP* expression is observed in cells of the caudal haematopoietic tissue (CHT) (green bracket) in *Tg(c-myb-zRE1-PF5:egfp)^{qmc85}* embryos. **(Cii)** *GFP* expression is detected in cells of the CHT in *Tg(c-myb-hRE1-PF5:egfp)^{qmc195}* embryos. **(iii)** A close-up of the CHT of the embryos in **(ii)**.

3.7.3 The *c-myb*-hRE1-PF5:*egfp* expression is retained in adult haematopoietic cells in the *qmc195* kidney

The presence of *qmc195:GFP*⁺ cells in the CHT, the larval site of haematopoiesis, suggested that there may also be *qmc195:GFP*⁺ haematopoietic cells in the adult site of haematopoiesis, the kidney marrow. In order to examine this, the kidneys of wild-type fish, *qmc85* and *qmc195* transgenic fish, were dissected. Blood cells were flushed out and then analysed on a Beckman Coulter Astrios EQ fluorescence-activated cell sorter. Sytox red was used to exclude any dead cells. Non-transgenic WT cell samples (Figure 3.33Ai-iii) were used to set the gates that were then applied to the WKM replicates of the *qmc85* (1, 2 and 3) and *qmc195* (1, 2 and 3) transgenic fish. The flow cytometric data on two of the *qmc85* replicates (2 and 3) were generated by Martin Gering in a previous experiment and were re-analysed in the context of this study. The analysis of the green versus red fluorescence of *qmc85* replicates 1, 2 and 3 revealed that the percentage of GFP⁺ cells gated was 1.76%, 1.43% and 6.41%, respectively, with an average 3.2% (Figure 3.33Bii). To identify the type of cells that expressed GFP, forward and side scatter characteristics of the GFP⁺ cells were analysed. These indicate the cells' size and granularity, respectively. The FSC and SSC analysis of GFP⁺ cells gated for *qmc85* replicates show that an average of 13.6% of these cells fell inside the progenitor gate, while an average

of 84.6% of the cells was found in the myelomonocyte gate (Figure 3.33Biii). The investigation of *qmc195* replicates 1, 2 and 3 show that the percentage of GFP+ cells gated was 0.13%, 0.45% and 0.07%, respectively, with an average of 0.22% (Figure 3.33Cii). FSC and SSC characteristics of GFP+ cells gated for *qmc195* replicates show that an average of 69.6% fell into the progenitor gates. Also, the analysis of *qmc195* replicates shows that the GFP+ cells gated found in the myelomonocytic gate were an average 21.4% (Figure 3.33Ciii). Thus, the flow cytometry revealed that there was a difference in the percentage of progenitors and myelomonocytes among the GFP+ cells of the zebrafish whole kidney marrow in *qmc195* compared to *qmc85* control (Figure 3.33D & E). These results suggest that the hRE1 could be sufficient to direct transgene expression in the definitive haematopoietic progenitor cells of the zebrafish embryos and adult kidney marrow. Unlike the zebrafish orthologue, it does not appear to drive convincing expression in more mature myelomonocytes.



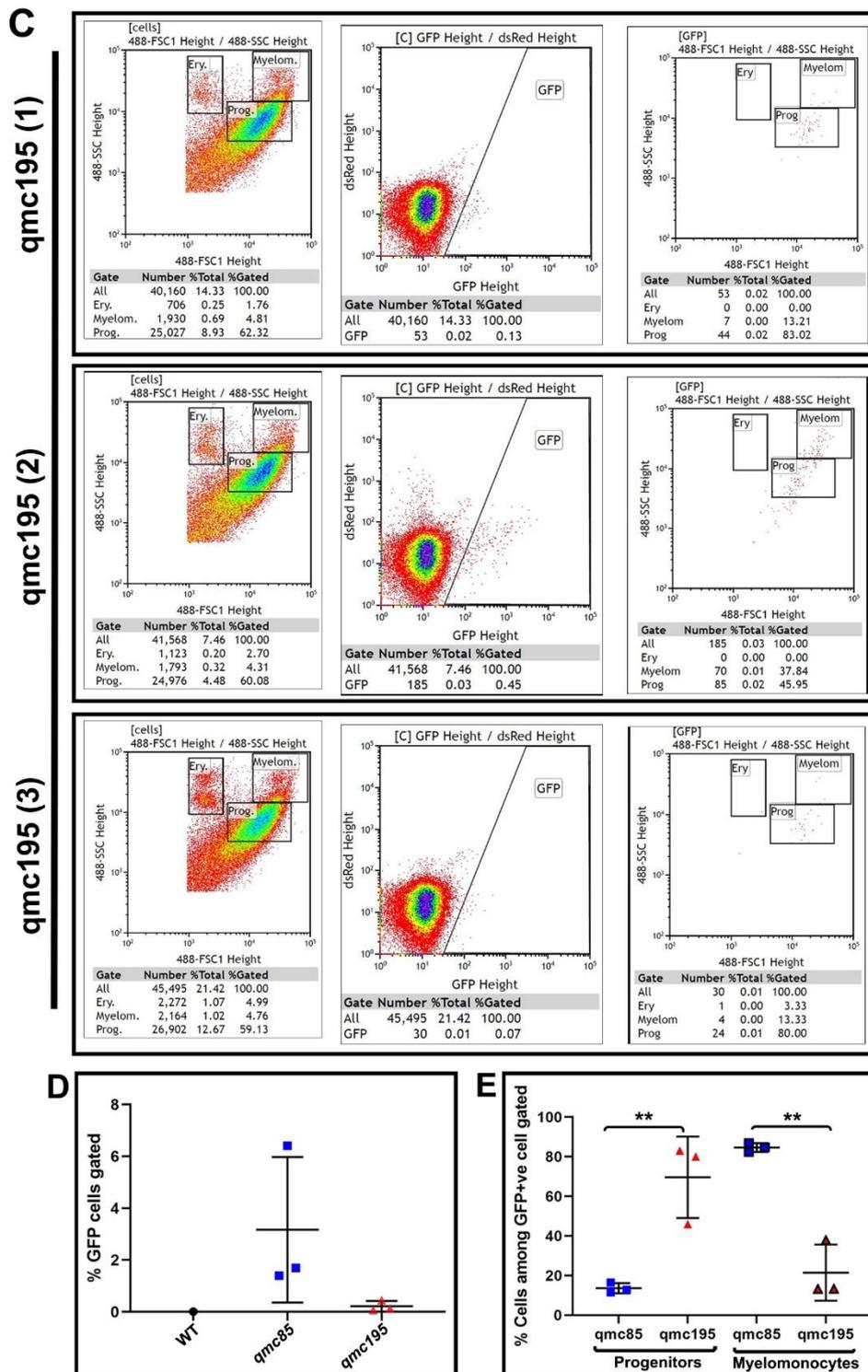


Figure 3.33 Flow cytometric analysis show that $Tg(c-myb-hRE1:PF5:egfp)^{qmc195}$ transgenic fish carry GFP+ haematopoietic progenitor and myelomonocytic cells in kidney marrow

Flow cytometric analysis of haematopoietic cells isolated from the kidney marrow of wt, $Tg(c-myb-zRE1-PF5:egfp)^{qmc85}$ $Tg(c-myb-hRE1-PF5:egfp)^{qmc195}$ transgenic fish. n = 1 (wt), 3 (qmc85) and 3 (qmc195). (Ai-

iii) Flow cytometric analysis of KM adult wt fish (non-transgenic). **(Ai)** The panel shows the light scatter characteristics of all cells for the wt control fish. **(Aii)** The panel shows a density plot that displays green versus red fluorescence in the kidney marrow cells of wt fish. **(Aiii)** The panel shows the light scatter characteristics of all cells for the GFP+ cells identified in the wt fish. **(Bi-iii and Ci-iii)** Flow cytometric analysis of adult kidneys of *qmc85* and *qmc195* transgenic fish replicates. **(Bi and Ci)** Panels show the light scatter characteristics of all cells in the transgenic fish of *qmc85* and *qmc195* replicates. **(Bii and Cii)** Panels show density plots that display green versus red fluorescence as detected in the kidney marrow cells of each of the transgenic fish replicates analysed. **(Biii and Ciii)** Panels show the light scatter characteristics of all cells for the GFP+ cells identified in the transgenic fish of *qmc85* and *qmc195* replicates. **(D)** The dot plot graph showing the percentage of GFP+ cells gated in wt, *qmc85* and *qmc195* replicates. **(E)** The dot plot graph showing the percentage of progenitors and myelomonocyte cells within the GFP+ cells gated in wt, *qmc85* and *qmc195* replicates. The mean values proved to be statistically different. (*t*-test: ***p*<0.01). **Shapes:** black circle represents wt; blue squares (with and without black border) represent *qmc85* replicates and red triangles (with and without black border) represents *qmc195* replicates. The overall number of cells and those in the particular gates and the corresponding percentages are shown under each panel. Gates from left to right: [Cells]= time AND live AND cells; [c]= time AND live AND cells; [GFP]= time AND live AND cells AND GFP. The cell sorting was performed on a Beckman Coulter Astrios EQ operated by Nicola Croxall.

4 Discussion

4.1 Summary

Previous analysis has shown that the regulatory element one CNE1 downstream of the *c-myb* gene is conserved in vertebrates (Hsu, 2010). Zebrafish regulatory element one (zRE1), which contain zebrafish conserved-non coding element one (zCNE1) tested in Tg(*c-myb*-zRE1-PF5:*egfp*)^{qmc85} line, demonstrate that zCNE1 is sufficient to direct *c-myb* promoter activity to haematopoietic cells. This *qmc85* line displays GFP expression in blood cells of the CHT embryo and adult zebrafish kidney marrow. In this study, we carried out a Single-cell RNA Sequencing (scRNA-Seq) to learn more about the exact nature of the GFP+ haematopoietic cells of zebrafish in the adult kidney marrow of the Tg(*c-myb*-zRE1-PF5:*egfp*)^{qmc85} zebrafish line. The flow cytometric analysis had revealed that the majority of GFP+ cells were in the myelomonocytic gate and that most of the myelomonocytes were GFP-positive. The scRNA-Seq analysis showed that the GFP+ cells of *qmc85* were neutrophils at different maturation stages. We used the CRISPR/Cas9 editing system to delete the zRE1 containing zCNE1 alone (*qmc193*) and in combination with the three potential elements (PEs) (*qmc194*) from the zebrafish *c-myb* locus to investigate whether the zCNE1 and other elements are essential for the of *c-myb* expression during zebrafish haematopoiesis. The homozygous embryos showed normal

c-myb expression in blood cells of the caudal haematopoietic tissue at four days post fertilisation, and the adults were viable and fertile. Our findings revealed that though zCNE1 is sufficient to drive the activity of the *c-myb* promoter in neutrophil cells, it is not necessary for the expression of the *c-myb* gene. The human regulatory element one (hRE1) containing human conserved-non coding element one (hCNE1) was tested in our newly generated stable Tg(*c-myb*-hRE1-PF5:*egfp*)^{qmc195} line to investigate whether the hCNE1 is sufficient to direct transgene expression to haematopoietic cells in the embryo and adult kidney marrow. The result suggests that the hCNE1 can direct the expression of *c-myb* gene to definitive haematopoietic cells in the zebrafish embryos and adult kidney marrow. However, in the *qmc195* line, the expression of GFP was low, and the number of GFP+ myelomonocyte cells was small compared to the *qmc85* line.

4.2 Single-cell RNA Sequencing (scRNA-Seq) of whole kidney marrow (WKM) of adult zebrafish Tg(*c-myb*-zRE1:PF5:*egfp*)^{qmc85} line

In the *qmc85* zebrafish transgenic line, GFP-positive kidney marrow cells are neutrophils. The transcriptome analysis shows that all the GFP+ cells express the neutrophil granulocyte gene lysozyme (*lyz*). This was consistent with the GFP-positive cells found within the myelomonocyte gate of the wt whole KM light scatter profile as defined by Traver and colleagues (Traver et al., 2003a). These

myelomonocytic cells displayed typical neutrophil morphology (segmented and non-segmented nucleus) when these cells were isolated, cytopun and stained with Giemsa, as found in the previous microscopic study in our lab (unpublished). To explore further whether the myelomonocytic cells were neutrophils, *qmc85* was crossed to the *Lyz:dsRed* line (generated in Crosier's lab (Hall et al., 2007)). In adult zebrafish of double transgenic *lyz:dsRed:qmc85:GFP*, flow cytometric analysis of kidney marrow revealed that most of the myelomonocytic cells were positive for both fluorescent transgenes. Our findings show that all these cells express *Lyz*, a neutrophil-specific gene, which encodes zebrafish lysozyme enzyme (Yang et al., 2012). This could imply that the *qmc85* GFP+ cells belong to the neutrophil cell lineage. Neutrophils are the first responders to migrate towards the inflammation, tissue damage and to kill microbes (Lieschke et al., 2001, Jenne et al., 2018). In zebrafish, it has been found that lysozyme, which is a key component of neutrophil granules, is expressed in the primitive and definitive neutrophils (Athanasiadis et al., 2017, Hall et al., 2007, Liu and Wen, 2002). Lysozyme has an antibacterial activity that helps to break specific linkages in the bacterial cell wall (Gordon et al., 1974). In the study of Jin and colleagues, homozygous carriers of *c-myb* mutant allele, called *hkz3*, display a specific defect in primitive neutrophil development which fail to initiate the expression of the

neutrophil marker *lyz* (Jin et al., 2016). A number of neutrophil markers, like *npsn*, *srgn*, *lect2l* expressed in our cell cohort were selected in comparison with *lyz* based on the highest total number of UMI and r^2 values. They were widely expressed and are known as neutrophil-specific genes. Neutrophil marker, *srgn* (serglycin), is an intracellular proteoglycan of hematopoietic cells that help to localise and packing of granule proteins in neutrophils (Niemann et al., 2007). The zebrafish marker *npsn* (nephrosin) is a neutrophil-specific granzyme used for host defence against microbial infections (Di et al., 2017). Similar to mammals, *lect2l* (leukocyte cell-derived chemotaxin 2) plays a significant role against microorganism infection in neutrophils (Huo et al., 2019). Some of the neutrophil specific genes, such as *cxcr4b*, *rac2* and *wasb* were also expressed in our cells. These genes were also confirmed to be specific for neutrophils in the studies of Furze and Rankin, Rosowski and colleagues and Kumar and colleagues, respectively (Furze and Rankin, 2008, Kumar et al., 2012, Rosowski et al., 2016). Athanasiadis and colleagues' study further confirm that these genes (*cxcr4b*, *rac2* and *wasb*) are specific to neutrophils (Athanasiadis et al., 2017). Our GFP+ *qmc85* cells also expresses, *cebpb*, *cfl1*, *illr4*, *mpx* and *ncf1* genes. This was in agreement with the study of Athanasiadis and colleagues on the upregulation of these genes in the neutrophil cells branch (Athanasiadis et al., 2017). In their study,

they used FACS index sorting with single-cell RNA Seq to show the gene expression profile of a number of blood cells using different zebrafish transgenic reporter lines. The blood lineage tree of the zebrafish generated was organised along their differentiation path due to their transcriptional differences. The morphology of sorted cells was also found to match the cell type branches (Athanasiadis et al., 2017). Our scRNA-Seq data also reveal that the markers which are characteristic for other blood cell lineages such as macrophage, red blood and T cells were absent across all the *qmc85:GFP+* (2246) cells. This probably suggests that the CNE1 under the control of the *c-myb* promoter (PF5) could direct haematopoietic stem cell to myelomonocytes (neutrophil granulocytes) transition. In our adult zebrafish, the flow cytometric analysis of the kidney marrow cells in *qmc85* transgenic line suggests that zRE1 drives GFP expression into myelomonocytes and progenitors. In most of the haematopoietic system models, myeloid cell development begins with the common myeloid progenitor (CMP), which separate early from the lymphoid lineage and gives rise to the granulocyte-monocyte progenitor (GMP). These are bipotent progenitors which differentiate to monocyte-macrophages and neutrophil granulocytes (Iwasaki and Akashi, 2007, Orkin and Zon, 2008a, Orkin and Zon, 2008b, Stachura and Traver, 2011). However, even though the monocyte/macrophage and granulocyte generated arose from a

common cell branch GMP, none of the macrophage markers was expressed across all our *qmc85:GFP+* cells.

The GFP-positive neutrophils in the kidney marrow of *qmc85* transgenic include cells at different maturation stages. Further analysis of the scRNA-Seq data of the neutrophils cells shows that the cells were grouped into different cell clusters of gene expression profile, implying different maturation stages. The gene expression profiles of our zebrafish scRNA-Seq analysis were compared with the most recent study on humans (Dinh et al., 2020). We took advantage of this study because they used different methods that confirmed that the cells were at different maturation stages. In their study, they used FACS, microscopy, in-vitro and in-vivo progenitor proliferation analysis and bulk RNA-Seq (Dinh et al., 2020). Their analysis gave us more information and a clear understanding of the different maturation stages of the neutrophil cells. We chose the study of Dinh and colleagues in our comparison in place of Athanasiadis and colleagues' study because, in the latter study, they used a small number (~337) of neutrophil cells as compared to our 2246 cells. The analysis of the maturation or differentiation stages of the neutrophils was not included in Athanasiadis and colleagues' study. To know what each cell cluster represents, we compared the gene expression profiles of our results with that of Dinh and colleagues by re-analysing their RNA-Seq data obtained from NCBI

database. The comparison shows that the cells were at different stages of maturation. This was in agreement with the study of Grassi and colleagues' on human neutrophils differentiation (Grassi et al., 2018). They found out that there was a gradual maturation from P/Ms to SNs in the human bone marrow. The findings of Dinh and colleagues' study as well as those of Grassi's and colleagues and our Dinh-based analysis, were in agreement with our findings that there could be a gradual maturation across the different clusters from extreme cluster one to extreme cluster two. Cluster one could imply immature neutrophils, and cluster two mature neutrophils. During neutrophil differentiation, the various maturation stages encompass the transition of the immature (progenitors) cells to mature cells. Most of this development takes place in mammalian bone marrow or in zebrafish kidney, where the precursors of these cells reside (Bainton et al., 1971, Lieschke et al., 2001). The most significant differentially expressed genes in cluster one were found to be highly expressed in the earliest neutrophil progenitor (ePreN) of Dinh and colleagues' study and the P/Ms (immature neutrophils) of Grassi and colleagues' study. This cluster probably represents an immature stage of neutrophil differentiation. Some of the cells in cluster one also expresses *pes* (Pescadilo) and *fbf* (Fibrillarin) genes, which are involved in the generation of ribosomes and are important for stem cell survival (Figure 4.1). This was in agreement with the findings of

Athanasiadis and colleagues' study, where they identified these genes as stem cells (HSCs) markers (Athanasiadis et al., 2017). The presence of ribosomal genes in cluster one may be due to the protein synthesis which immature neutrophil cells needed during the proliferation and differentiation processes. The conservation analysis between zebrafish and humans suggested that there was a strong similarity between the ribosomal gene expression in human HSCs and the zebrafish progenitor population, i.e., GMP (Athanasiadis et al., 2017). They also found out that there was a gradual downregulation of the ribosomal genes during lineage-specific differentiation. The expression of *c-myb* gene was found in a subset of cluster one, which further confirms that the cells in cluster one may represent immature neutrophils (Figure 4.1). The level of *c-myb* expression has been found to be high in immature haematopoietic cells and gradually is repressed as soon as they differentiate during their hematopoietic maturation (Gonda and Metcalf, 1984, Shen et al., 2013). Cluster two could probably represent the mature stage of neutrophil differentiation. The most significant differentially expressed genes in cluster two were found to be highly expressed in the mature neutrophils (Neuts) of Dinh and colleagues' study and the SNs (mature) of Grassi and colleagues' study. For example, CXCR1, CXCR4, which are expressed on the surface of mature neutrophils, direct their migration to infected or damaged tissues

(Holmes et al., 1991, Link, 2005). The CXCL8-CXCR1 and CXCL12-CXCR4 ligand-receptor pairs are conserved between zebrafish and human (Oehlers et al., 2010, Tulotta et al., 2016). The functionality of mature neutrophils also relates to the functions of the immune system function and the processes of vesicle transport (Grassi et al., 2018).

Our scRNA-Seq analysis has shed light on the different maturation stages of neutrophils in zebrafish. There is also a need to carry out an in-depth analysis of the rest of the clusters (from 3 to 9) and compare the results with Dinh and colleagues' study. Evolutionary studies could be carried out to determine if there is any conservation of the differentially expressed gene in our zebrafish neutrophil cells with orthologues in other vertebrate species.

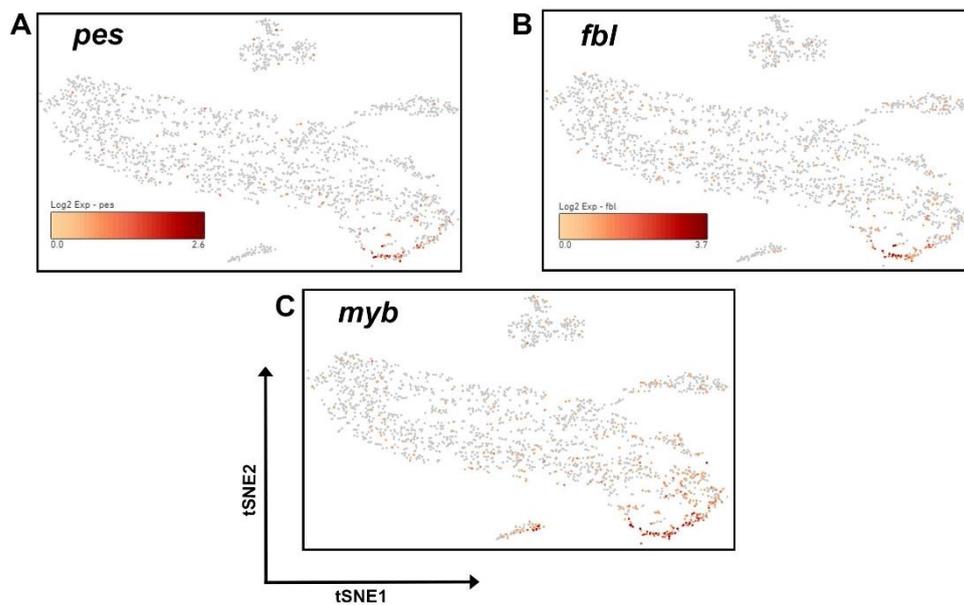


Figure 4.1 Single-cell RNA Sequencing of $Tg(c-myb-zRE1:PF5:egfp)^{qmc85}$ of zebrafish kidney marrow reveals that cluster one cells express stem or early progenitor genes

The single-cell RNA-Seq data for 2246 cells were analysed by a clustering algorithm, the t-stochastic neighbour embedding (t-SNE) algorithm, that clusters cells based on their gene expression profile and provides a two-dimensional representation of the multi-dimensional data set. A heat map representation of log₂ values of the number of transcripts (UMI) counts for individual genes is projected for every single cell depicted in the t-SNE plot. The panels **(A-C)** show the expression of *pes*, *fbl* and *c-myb* genes in a subset of cells in cluster one.

4.3 Deletion of zebrafish conserved non-coding element zCNE1 and potential enhancers (PEs) by CRISPR/Cas9 system

The zCNE1 is sufficient but not essential to direct the expression of *c-myb* gene to the haematopoietic blood lineage of neutrophils in both embryo and adult zebrafish. In this study, we carried out three deletions involving zRE1 alone, three PEs (MP40036-, MP40036 and MP40037+ (table 1.1)), and a combination of zRE1 with the three PEs downstream of the *c-myb* gene using the CRISPR/Cas9 system. The deletion of zRE1 alone and in combination with the three PEs were successfully transmitted to the next generation of zebrafish founders. This was comparable with a previous study where the high efficiency of zebrafish founder embryo was observed (Hwang et al., 2013a). The mutant fish *qmc193* (from zRE1 deletion) and *qmc194* (from combine zRE1 and PEs deletion) did not show any phenotypic abnormalities. Regarding the third deletion (three PEs alone), no mutant was established since the mutant allele was not transmitted to the next generation. Possible reasons for this could be that the targets sites (TS7 and TS16) having a low average score could have affected the efficiency for Cas9 to generate a double-strand break (DSB). The sgRNAs target sites with a high score are more efficient in generating Cas9 DSB (Moreno-Mateos et al., 2015). Another possibility could be that the mutagenesis rate might have been affected by the injection time of the Cas9-sgRNA complex. The

earlier Cas9-sgRNA introduces the DSB and non-homologous end joining (NHEJ) repairs the damage during embryonic development, the higher the chances that mutagenesis will occur in the germ cells for transmission to the next generation. The time for assembly of Cas9 and sgRNA could be essential with its rapid cell division and development (Zhang et al., 2018). In zebrafish, during the blastula stage (1k-cell), there are only approximately four germ cells (Raz, 2003). The occurrence of Cas9-induced mutagenesis around the blastula stage may result in a high probability of targeting the somatic cells than the germ cells (Jao et al., 2013). These drawbacks could be overcome by designing another sgRNA target site with a high score value in an attempt to delete the three PEs downstream of the *c-myb* gene. Also, a large number of embryos could be injected with Cas9-sgRNA complex.

The lack of any phenotypic abnormalities in our mutants *qmc193* and *qmc194* could imply that other elements could be playing a redundant role in driving the expression of *c-myb* in neutrophils. The absence of any phenotypic abnormalities after the deletion of regulatory elements could be attributed to enhancer redundancy (Antosova et al., 2016, Osterwalder et al., 2018, Sagai et al., 2017). In the same manner, as gene redundancy has proven to be responsible for the lack of phenotypes associated with gene deletion, enhancer redundancy could also offer a similar explanation for the

lack of phenotypic abnormalities with the deletion of enhancer (Ahituv et al., 2007). In the study of Osterwalder and his colleagues, no phenotypic abnormalities were observed when they used CRISPR/Cas9 system to individually delete ten mouse embryonic enhancers with each controlling robust limb activity in transgenic reporter assays (Osterwalder et al., 2018). Ahituv and his colleague carried out an experiment by deleting four CNE (which act as enhancers) to address whether they were necessary to regulate genes expression (Ahituv et al., 2007). They found out that the mice lacking these conserved sequences were viable, fertile, and had no noticeable morphological abnormalities (Ahituv et al., 2007).

Though zCNE1 is sufficient to drive *c-myb* promoter activity to neutrophil cells, we found out that the level of expression of both *lyz* and *c-myb* mRNA was normal in the developing embryos carrying *qmc193/qmc194* and *c-myb*^{t25127} loss-of-function alleles in the WISH experiment. The loss of *c-myb* gene expression in mutant alleles *t25127* (Soza-Ried et al., 2010) and *hkz3* (Zhang et al., 2011) has been previously described. Also, The loss of the *lyz* and *c-myb* genes expression in *c-myb*^{t25127} mutant has been reported previously (Hess et al., 2013). In our lab, it has been proven that the zRE1 in the *qmc85* transgenic line drives the expression of *c-myb* gene to the blood cells in the CHT at 3 dpf (Hsu, 2010). In our study, there were no obvious changes in the expression level of *c-*

myb gene in all our mutant embryos at 4 dpf compared with the wide type where there was a normal expression at the same developmental time. The study of the expression of *c-myb* gene at a different time in the developmental stages of wt zebrafish embryos revealed that the expression of *c-myb* gene in CHT was strong between 4-6 dpf (Murayama et al., 2006).

Since the deletion of zRE1 alone and in combination with the three PEs did not affect the *c-myb* expression in CHT, future studies could investigate the probable role of PEs in directing the expression of *c-myb* in other haematopoietic cells, such as prRBCs. For this, the three PEs could be cloned separately upstream of the *c-myb* promoter in the GFP reporter transgene construct. The constructs could then be used to generate stable zebrafish transgenic lines, and the expression of GFP could be tested. There could be a need for further analysis to study the neutrophil kidney marrow cells in *qmc193/194* homozygotes. This study could involve the cross between the double transgenic line *qmc85;Lyz:dsRed* with *qmc193/194* homozygous to generate *qmc193/194 qmc85;Lyz:dsRed* lines. The neutrophil kidney marrow cells labelled with *qmc85;Lyz:dsRed* could then be characterised using flow cytometry to find out whether the number of neutrophils is normal. If the number of cells is abnormal, the *GFP+:dsRed+* cells could be FACS sorted for further analysis. Cyto centrifugation of the sorted

cells followed by Giemsa staining could then be used to show any changes in morphology and the number of neutrophil cells. Since the deletion of the zCNE1 did not affect the expression of the *c-myb* gene, the presence of another enhancer with similar tissue specificity (enhancer redundancy) could be the reason for the absence of phenotypic abnormalities observed. It will be interesting to search for other potential regulatory elements (enhancers) that could be driving the expression of the *c-myb* gene in the haematopoietic blood cells. These potential enhancers could be located in the introns within the *c-myb* gene or far away from *c-myb* locus. A previous alignment study carried out in our lab (Hsu, 2010) shows several non-coding elements in the *c-myb* locus, which are conserved in zebrafish and other species. Also, a number of intronic regions of the *c-myb* locus have been found to be conserved between other fish species (medaka, stickleback, tilapia and fugu) (Hsu, 2010). Thus, this could represent a potential enhancer in the introns of the *c-myb* gene in zebrafish. Further study could, however, be needed to determine if potential enhancers are embedded in these regions. The bioinformatics tool, PEGASUS (Predicting Enhancer Gene Association Using Synteny) (Clément et al., 2018), could be used to search for other potential intron or remote enhancers. PEGASUS work by first defining regulatory elements based on sequence conservation then links them with

targeting genes using a synteny conservation score. The analysis result could also be compared with Hsu (Hsu, 2010) (zCNE1) and (Aday et al., 2011) (PEs) findings in the vicinity of *c-myb*. Alternative enhancers that regulate the expression of *c-myb* in granulocytes could be located a distance away from the *c-myb* gene vicinity, thus enabling promoter-enhancer- *c-myb* interaction. An enhancer that regulates *Shh* expression in mouse limbs was found in approximately 1 Mb of the *Shh* targeting gene (Lettice et al., 2003). The deletion of this enhancer leads to loss of limb *Shh* expression (Sagai et al., 2005). The interaction between promoters and distal enhancers by chromatin looping is essential for transcriptional regulation of gene expression (Amano et al., 2009). This leads to a three-dimensional (3D) chromatin folding in different cell types, which could further be organised into topologically associating domains (TADs) (Dixon et al., 2012). In mammals, the C-Technologies, 3C (Dekker et al., 2002), 4C, 5C, and Hi-C (Lieberman-Aiden et al., 2009) have been used to study the 3D genome organisation, which confirm the long-range interaction between genomic loci within TADs. Just as in mammals, Hi-C has also been used to study 3D genome organisation in whole zebrafish (Gómez-Marín et al., 2015, Kaaij et al., 2018). The different number of cells ranging from 1-3 million (Franke et al., 2020) to approximately 10 million (Gómez-Marín et al., 2015) from whole

zebrafish embryo could be used as input cells for the Hi-C experiment. For specific cell types such as granulocytes as in our study, many biological replicates may be needed to achieve the high number of cells for this assay.

4.4 Incomplete functional conservation of human regulatory one (hRE1) in zebrafish transgenic line

Due to the important role of *c-myb* in haematopoiesis, several studies have been carried out previously to analyse the regulatory elements that control *c-myb* gene expression during haematopoiesis (Hsu, 2010, Mukai et al., 2006, Stadhouders et al., 2012, Zhang et al., 2016, Wahlberg et al., 2009). Previous studies in our lab have shown that both regulatory elements, zRE1 and mRE1, are able to direct the expression of the *c-myb* gene in haematopoietic cells in zebrafish transgenic lines (Hsu, 2010, Mohamed, 2015). In this study, hRE1 has been shown to direct the expression of *c-myb* gene to progenitors and myelomonocytes of adult kidney marrow in zebrafish transgenic line. In this study, we generated stable transgenic zebrafish to explore whether hRE1, including the hCNE1, can direct the expression of the *c-myb* gene in haematopoietic cells in the embryo and adult zebrafish. The hRE1 with an opposite orientation was placed at the upstream position of the promoter of the *c-myb* gene. Despite the orientation of the hRE1 in the construct, it was still capable of driving the expression of GFP to hematopoietic

cells, which also suggests a degree of positional independence. Enhancers can activate transcription remotely by using distant promoters, therefore acting in an orientation-independent manner (Ong and Corces, 2011). In the *qmc195* line, GFP reporter gene expression was found in cells of the CHT between 2 and 4 dpf. There was a weak expression of GFP in both the embryonic and adult cells of our line. The *GFP* mRNA expression was detected in the CHT of 3 dpf embryos in *in-situ* hybridisation experiments. We carried out the WISH experiment to confirm that the weak expression of the GFP observed in CHT was due to the presence of GFP protein and not caused by auto-fluorescence. Though there was a low GFP expression in our *qmc195*, hCNE1 may be responsible for directing the expression of *c-myb* gene in hematopoietic progenitor and myelomonocytes. Flow cytometric analysis was used to confirm the presence of the GFP expression in the adult WKM of zebrafish. The analysis also shows that the hCNE1 can also drive the expression of GFP in progenitors and in a few myelomonocytes. We compare the GFP expression pattern with that of *qmc85*, and it shows that they both have two populations of progenitors and myelomonocytes. However, in *qmc195*, the number of progenitors cells were more than those of myelomonocytes in *qmc85* and vice versa.

We studied the functional conservation of orthologues CNE1 by considering data from previous studies in our lab whose regulatory

element activity has been tested in zebrafish. To have a better understanding of the functional conservation of zCNE1, we did sequence alignment of similar vertebrates (dog, mouse, chicken, opossum, human and frog) (Figure 4.1A). Multiple sequence alignment has previously demonstrated that the CNE1 sequence is largely conserved among eight species, including mice and humans (Hsu, 2010). The *qmc85* line has shown that the combination of zebrafish *c-myb* PF5 and zRE1 line were able to drive transgene expression in the CHT in zebrafish embryos and blood cells (progenitors and myelomonocytes) of the adult kidney (Hsu, 2010). The mouse orthologue mRE1 (*qmc156*) study from our lab has been shown to be functionally conserved and able to direct the expression of zebrafish *c-myb* promoter to haematopoietic cells (Mohamed, 2015). The alignment revealed that the zebrafish CNE1 sequence has a 72.9%, 70%, 68.6%, 68.6, 62.9 and 57.1% sequence match similarity with those of frog, human, opossum, chicken, mouse, and dog, respectively (Figure 4.1B). We carried out an *in-silico* analysis of the conserved sequences of the aligned vertebrate (7) species to determine if CNE1 may contain putative TF binding site motifs that could play a role in regulating the expression of the *c-myb* gene during haematopoietic activities. Conserved TF binding site motifs among the species were selected based on the statistical significance ($P < 0.05$) (Figure 4.1C). Conserved motif sequences on all species

were then further analysed with two databases, JASPAR (Sandelin et al., 2004a) and TRANSFAC (Matys et al., 2003) (Appendix table 6.9). These databases could have some limitation in that they predict TF binding motifs rather than doing actual analysis of the motifs. They also openly obtain TF binding motifs information from both internal and external sources. These sources could have some biases depending on the nature of the data and the form of the analysis. From the analysis, the following putative TF binding sites; Pu1, Gabpa, Hsf1, Cebpa, Zfx, Tp53, Meis1, Cebpa, Runx1, Atf4, Cad, Pou6f1, Tbp, Irf1, Irf2 and Stat1 that could be involved in the regulation of haematopoietic activities were selected. These putative TF binding sites are conserved in most of the aligned species and are likely to be involved in regulating the *c-myb* expression in haematopoietic cells (Azcoitia et al., 2005, Cvejic et al., 2011, Huang et al., 2010, Liu and Patient, 2008, Pillay et al., 2010, Song et al., 2011, Wei et al., 2016). For example, previous studies have shown that some of these TF factors, pu.1 (Jin et al., 2012), cebpa (Jin et al., 2016) and runx1 (Huang et al., 2021) can be involved in the regulation of the neutrophils. In zebrafish, it has been reported that cebpa and runx1 probably synergise with pu.1 in the transcriptional regulation of *lyz* during zebrafish myelopoiesis by binding to the -2.4 kb *lyz* promoter (Kitaguchi et al., 2009). This may suggest the role of zCNE1 in directing the expression of *c-myb* gene in neutrophil

cells. A previous study has also identified a number of putative TF binding sites, such as, ets, gata, runx1, and c-myb on CNE1 of the conserved sequences of eight species (Hsu, 2010).

We analysed the sequence of zCNE1, mCNE1 and hCNE1 to determine whether a loss of TF binding motifs is responsible for the difference in the expression of the GFP in the zebrafish *qmc85*, *qmc156* and *qmc195* transgenic lines (Figure 4.2). One AML1 (Runx1) putative TF binding site as predicted by the databases (JASPAR and TRANSFAC) had TCTGTGTTC sequence for zCNE1 and AAGCACAGT sequence for mCNE1, while none in hCNE1. In vertebrate, the consensus sequence for AML1 (Runx1) is TGT/CGGT or ACCA/GCA. The ACCA was found in the analysis of our mCNE1 sequence while none of the consensus sequences were found in the analysis of our zCNE1 sequence. Runx1 is an essential factor for the emergence of HSC and is expressed in HSCs (North et al., 1999). In zebrafish, runx1 can regulate embryonic myeloid fate decisions by promoting granulocytic over macrophage lineage during neutrophils development (Jin et al., 2012). The presence of runx1 TF binding site in zCNE1 of *qmc85* line could explain the nature of the GFP+ granulocytes (neutrophils) at different maturation stages of our scRNA-Seq results. This could imply that the runx1 TF binding site activates the transgene during the differentiation of the neutrophils. GFI (growth factor independent 1) putative TF binding site was

detected in mCNE1 and hCNE1 but not in zCNE1. In mammals, there are two forms of these transcription repressors, Gfi1 and Gfi1b. They play essential roles in HSCs (Möröy et al., 2015). A Gfi1 is also required during the differentiation of the granulocytes (Karsunky et al., 2002). In zebrafish, there are three *gfi1* paralogues. The paralogous *gfi1aa* and *gfi1ab* are important during primitive erythropoiesis, while *gfi1b* plays a crucial role in definitive haematopoiesis (Moore et al., 2018, Thambyrajah et al., 2016). The presence of GFI TF binding site in the hCNE1 of *qmc195* line could explain the possible reason for getting the low level of GFP expression. It could suggest that the GFI is likely to switch the transgene in *qms195* off, and no GFP expression could be detected when the cells differentiate to granulocytes. It could be argued, therefore, that the deletion of a particular GFI motif from the hCNE1 may lead to the detection of a higher level of GFP expression in the granulocyte cells. The TF binding site IRF1 (Interferon regulatory factor 1), was observed in mCNE1 and hCNE1 but not in zCNE1. IRF1 is primarily involved during host immune response (Huang et al., 2010). IRF1 is activated during the control of normal haematopoiesis (Choo et al., 2006). Even though IRF1 TF binding site was detected in hCNE1, no population of GFP+ lymphoid cells was observed in our *qmc195* line. It has been reported that IRF1 is required during the maturation of lymphoid (T-cells) (Ohteki et al.,

1998). The GATA-1 TF putative binding site was seen in the three elements (zCNE1, mCNE1 and hCNE1). Six TF putative binding sites of GATA-1 were observed in the hCNE1 sequence, while three in zCNE1 and one in mCNE1. GATA-1 is essential for the commitment, differentiation, and survival of erythroid cells (Fujiwara et al., 1996). The presence of several GATA-1 TF binding sites in hCNE1 could be an indication of the important role of the *c-MYB* gene in human erythropoiesis (Vegiopoulos et al., 2006). However, since the expression of GFP is recapitulated endogenous *c-myb* in some tissues, we have not detected the expression of GFP in erythroid cells of the embryo and adult of *qmc195*. This could suggest there might be other factors that play a role in initiating the expression of *c-myb* in erythroid blood cells in transgenic lines. In human, it has been identified open chromatin structure of the *HBS1L-MYB* intergenic region in human erythroid cells expressing *c-MYB* that could be essential in haematopoiesis by controlling *c-MYB* gene expression (Wahlberg et al., 2009). In a previous study of zebrafish transgenic lines, none of the regulatory elements that were identified in the intergenic region of *hbs1l-myb* was able to drive the expression of *c-myb* in blood cells (Hsu, 2010). In the same manner, GFP expression was also not observed in the primitive red blood cells in ICM in *qmc85* line (Hsu, 2010) and in the adult kidney marrow of *qmc85* (Savage, 2012).

The evolutionary distance between living teleosts and mammals from their common ancestors is about 430 million years (Powers, 1991). Despite the massive evolutionary distance between zebrafish and humans, the development and function of the blood system are curiously conserved (Robertson et al., 2016). Even though there was incomplete functional conservation of hCNE1 in relation to zCNE1 in the transgenic line, it was able to drive the expression of *c-myb* in hematopoietic cells of the zebrafish transgenic line. This suggests that the CNE1 was existent in the common ancestor, and it was preserved by selective forces which maintain its function. The different TF binding sites observed in the two elements (hCNE1 and zCNE1) could be responsible for the different expression of GFP+ cells and their distribution among progenitors and myelomonocytes populations.

Further analysis of the TF binding sites motifs could be carried out *in vivo* to better understand the mechanisms underlying *c-myb* expression in hematopoietic cells. This could be done by deleting single or multiple TF binding sites and then testing the *c-myb* gene expression in the context of a transgenic reporter construct in live zebrafish. In addition, confocal analysis of the zebrafish embryos double transgenic lines *qmc195;Lyz:dsRed* could be carried out to check for any visible overlap in expression pattern between the two transgenic lines in the CHT of the embryos. To further characterise,

it will be interesting to carry out the flow cytometric analysis of the kidney marrow cells of those double transgenic line *qmc195:Lyz:dsRed* to better understand what cells population are being labelled with double transgenics. Since the DNase1 HS region relates to transcriptional activity, it might be interesting to extend the sequence of hRE1 in a new transgenic line that might include the expression of other blood lineages during haematopoiesis (Figure 4.2).

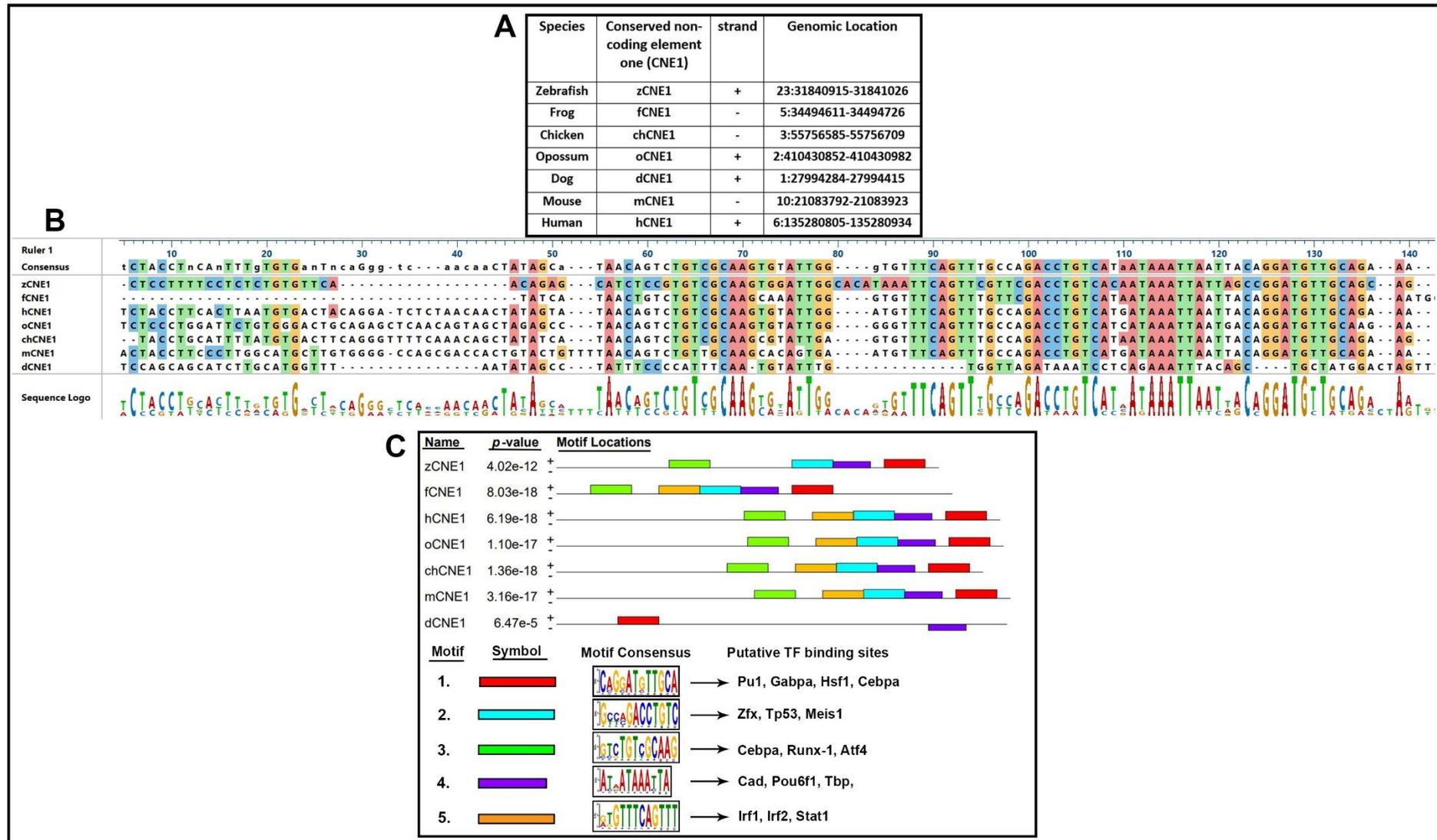


Figure 4.2 Putative binding sites of transcription factors on conserved non-coding element one (CNE1)

(A) A table showing the features of the species with a list of *C-MYB* downstream conserved non-coding element one (CNE1), direction (strand) of the gene expression, the genomic location from Ensembl genome browser (v 102) and the length (bp) of their sequences. **(B)** Shows multiple sequences alignment of CNE1 *C-MYB* genes of frog, human, opossum, chicken, mouse, and dog with reference to the zebrafish sequence. Coloured nucleotides represent the region of conservation matched to the zebrafish reference sequence using MegAlign Pro (v 17.2.1). The sequences of the species are arranged in the order in which they match the reference sequence from highest to lowest conservation. **(C)** The graphic representation of the transcription factor binding motif that was identified by MEME (Multiple Expectation maximisations for Motif Elicitation) (Bailey et al., 2009). STAMP online tool (Mahony and Benos, 2007) was used to characterise the identified motifs. Each of the motifs were further screened against JASPAR (v 2010) and TRANSFAC (v 11.3) database to mark binding sites for important transcription factors.

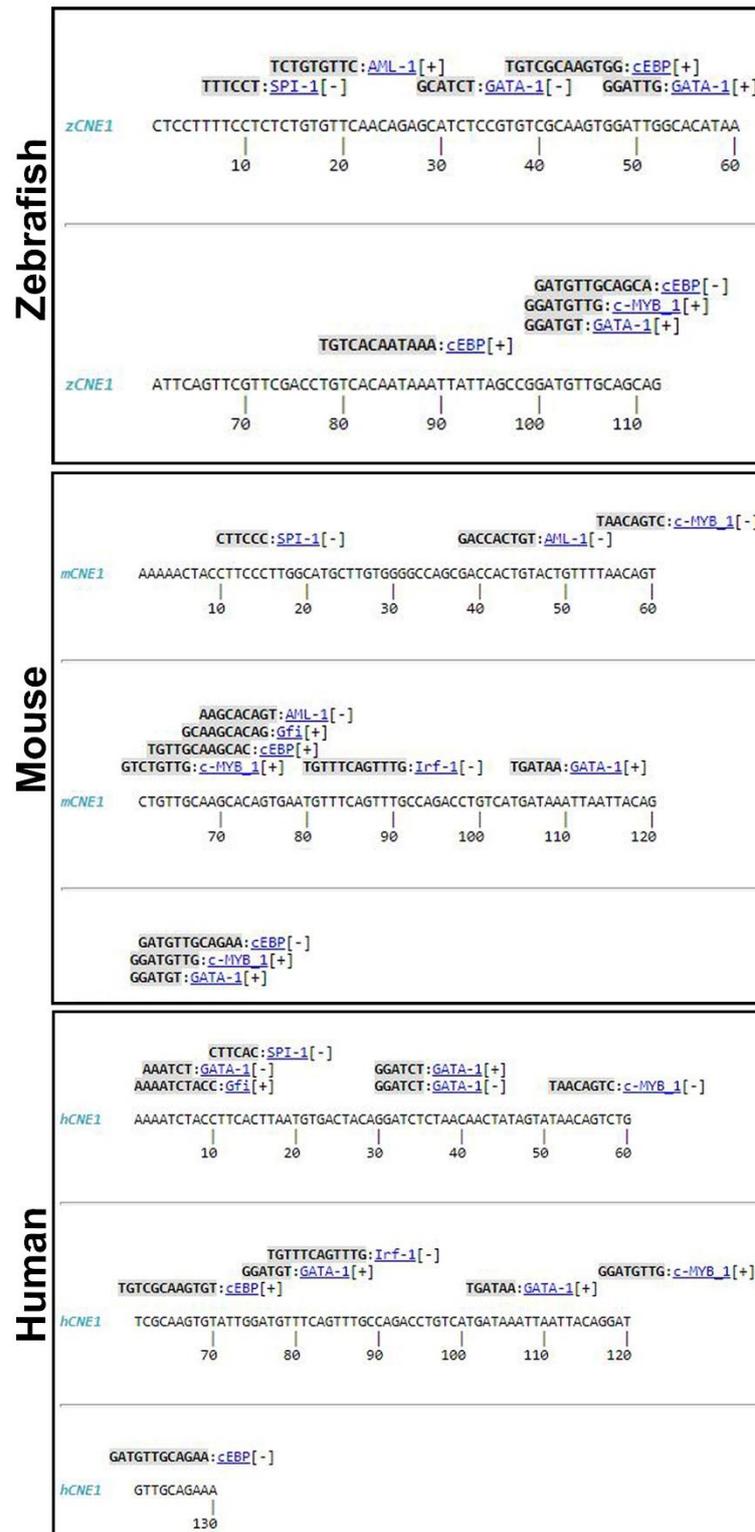


Figure 4.3 Putative binding sites of transcription factors on individual sequences of zebrafish, mouse and human conserved non-coding element one (CNE1)

Sequence view showing the putative TF binding sites among zCNE1, mCNE1 and hCNE1. ConSite database (<http://consite.genereg.net/cgi->

[bin/consite](#)) was used for this analysis, which uses the JASPAR datasets (TF score cutoff 80%) (Sandelin et al., 2004b).

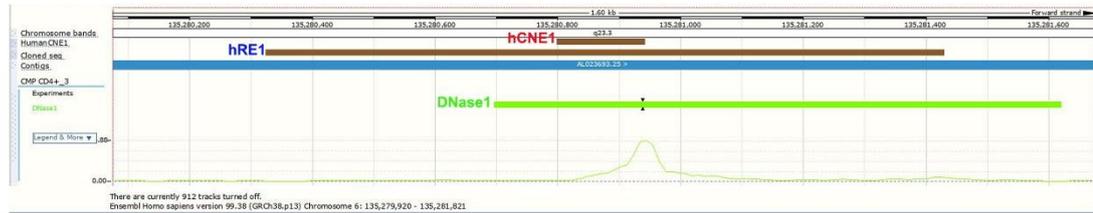


Figure 4.4 Localisation of the sites of DNase1 HS within the conserved non-coding element one downstream of human *c- MYB* gene

The diagram shows the human conserved non-coding element one (hCNE1) within human regulatory element one (hRE1) on chromosome 6 in human, as shown on the ENSEMBL database with custom tracks. The diagram also illustrates a DNaseI hypersensitive site (green box) within the human genome. Ensembl Homo sapiens version 99.38 (GRCh 38.p13) Chromosome 6: 135,208,074-135,281,675.

5 Conclusion

The previous study has shown that the zebrafish regulatory element one (zRE1) containing zebrafish conserved-non coding element one (zCNE1) in *qmc85* stable transgenic line is sufficient to direct *c-myb* promoter activity to haematopoietic cells (Hsu, 2010). My work has contributed to our understanding of the regulation of the *c-myb* gene in three aspects. Firstly, the scRNA-Seq analysis of the GFP+ kidney marrow cells in the *qmc85* line shows that all these cells expressed lysozyme and several neutrophil granulocyte markers. Additional analysis reveals the presence of GFP+ neutrophils at different maturation stages. Secondly, the deletion of the zCNE1 alone and in combination with all three potential *cis*-regulatory elements (PEs) using the CRISPR/Cas9 system reveals that though the zCNE1 is sufficient to direct the expression of *c-myb* gene, they are, however, not essential. Thirdly, CNE1 was found to be conserved among different vertebrates. The human conserved-non-coding element one (zCNE1), tested in a stable zebrafish transgenic line *qmc195*, is able to direct *c-myb* gene expression to haematopoietic cells though there was weak GFP expression compared to *qmc85* line. A comparison of putative transcription factor binding sites revealed differences between the zebrafish and the human CNE1 sequences that could explain the differences in the observed GFP expression patterns in the stable *qmc85* and *qmc195* transgenic lines. It will be interesting to search for other essential regulatory elements that

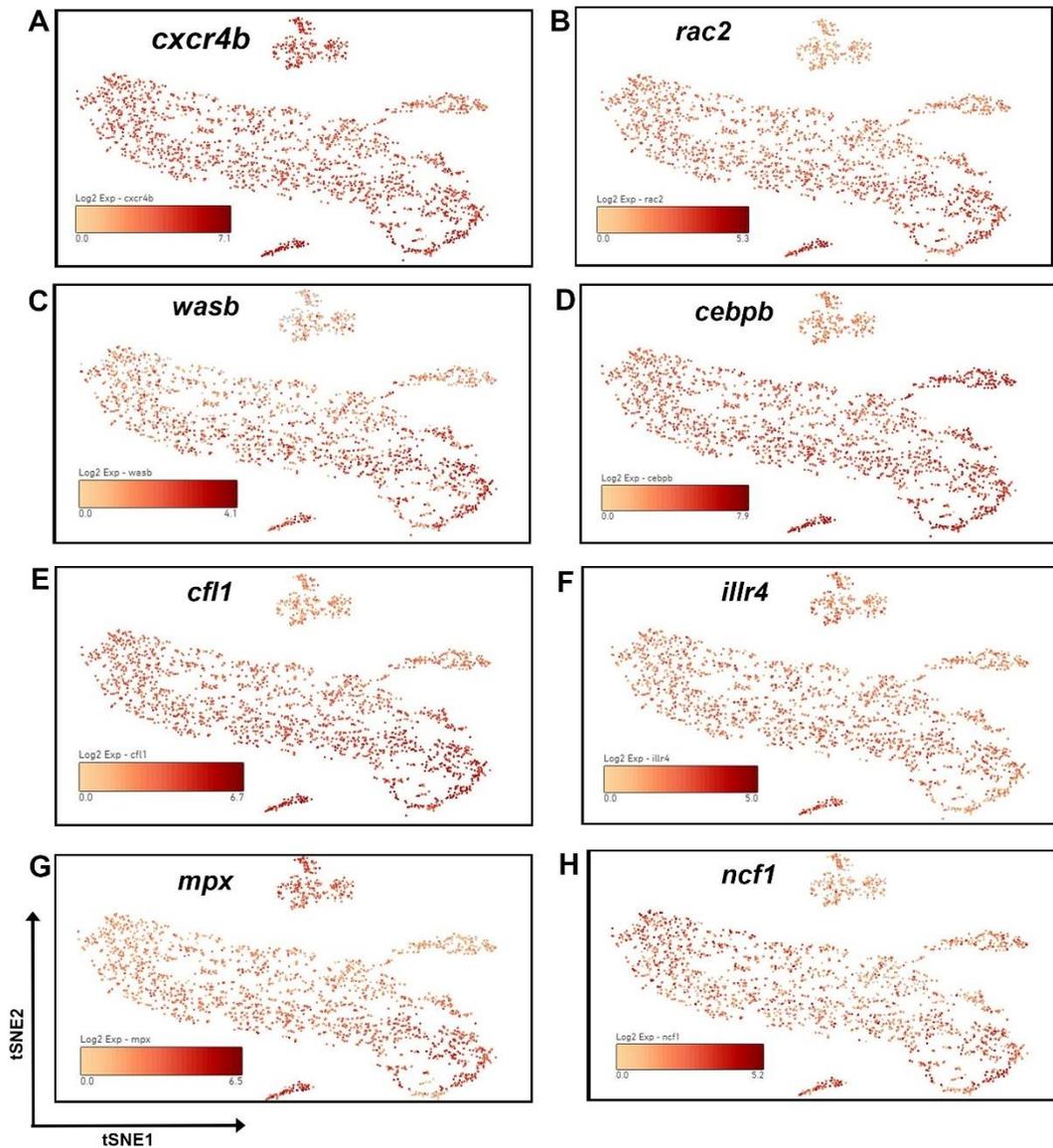
might be responsible for directing *c-myb* gene expression to haematopoietic cells.

6 Appendices

6.1 Appendix figures

Appendix figure 6. 1 A sequence of the human *C-MYB* locus. Orange bars represent the PCR product. Grey boxes present the polymorphisms.



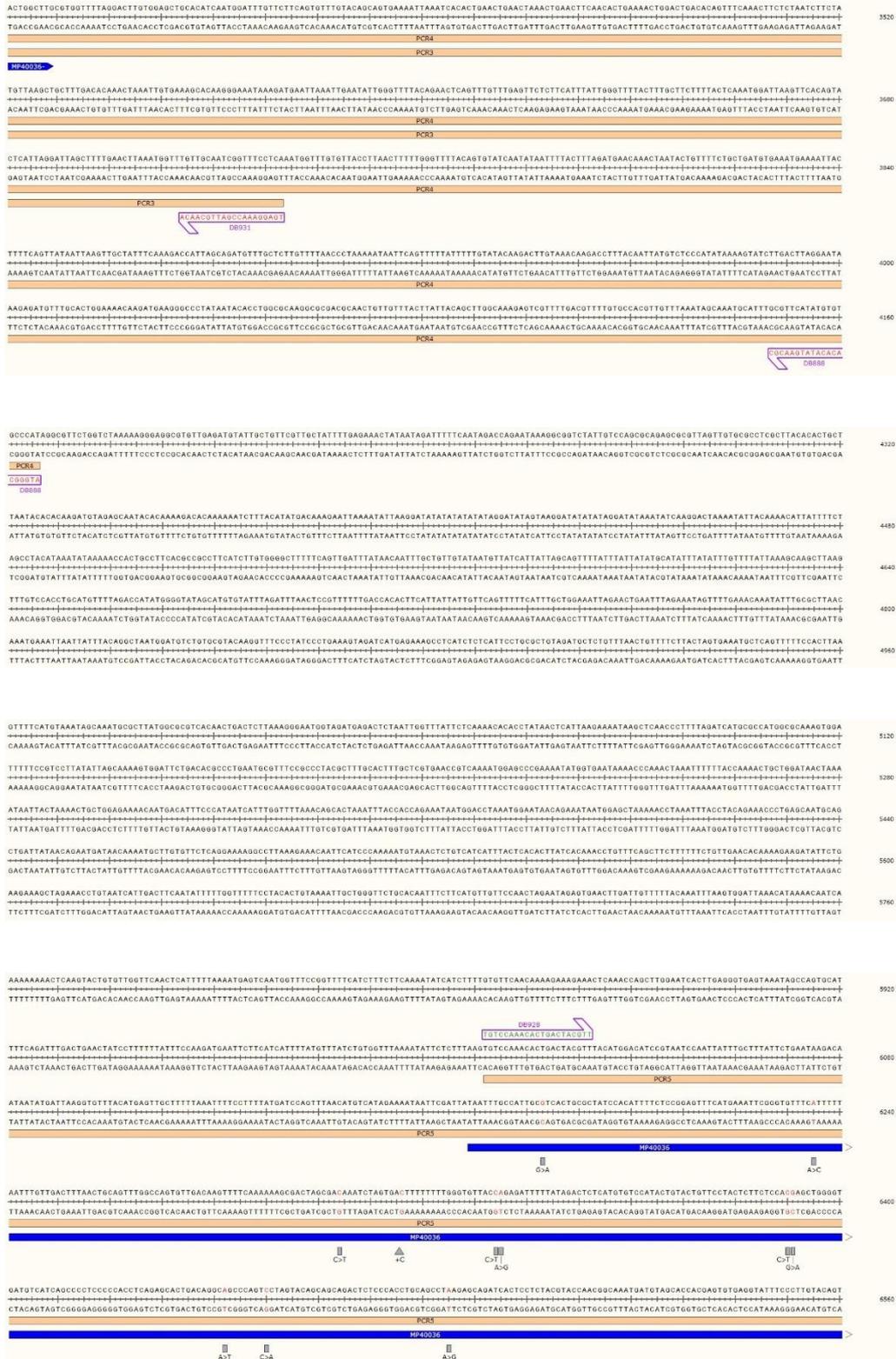


Appendix figure 6.2 Single-cell RNA Sequencing of $Tg(c-myb-zRE1:PF5:egfp)^{qmc85}$ of zebrafish kidney marrow express neutrophil genes

The single-cell RNA Seq data for 2246 cells were analysed by a clustering algorithm, the t-stochastic neighbour embedding (t-SNE) algorithm, that clusters cells based on their gene expression profile and provides a two-dimensional representation of the multi-dimensional data set. A heat map representation of log2 values of the number of transcripts (UMI) counts for individual genes is projected for every single cell depicted in the t-SNE plot. The panels **(A-H)** display the expression of a number of neutrophil genes.

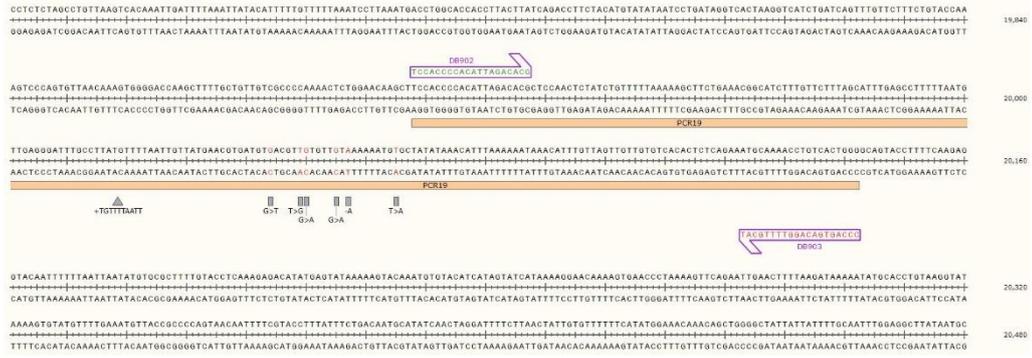
Appendix figure 6.3 A sequence of the zebrafish *c-myb* locus. Orange bars represent the PCR products. Grey boxes present the polymorphisms.

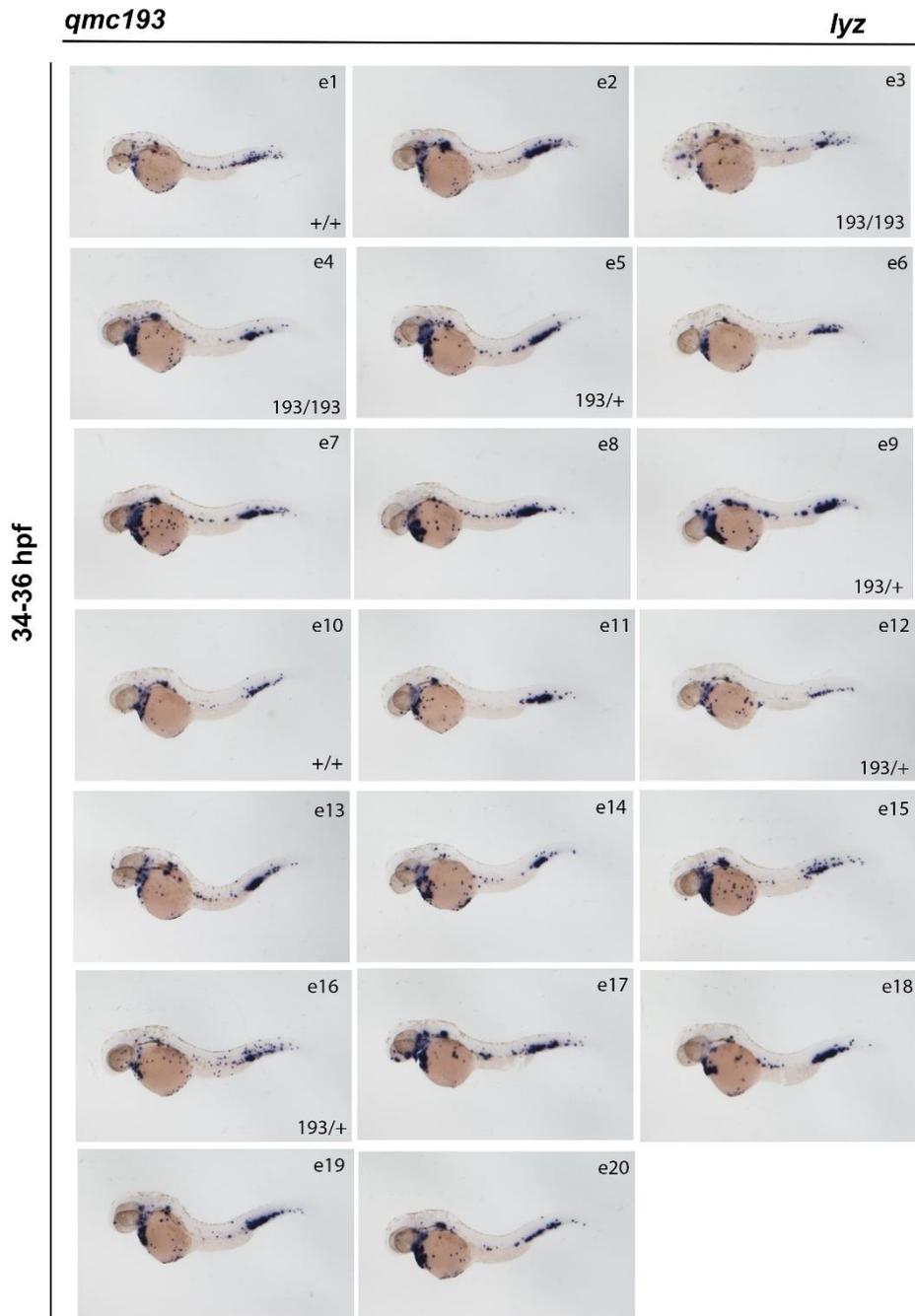






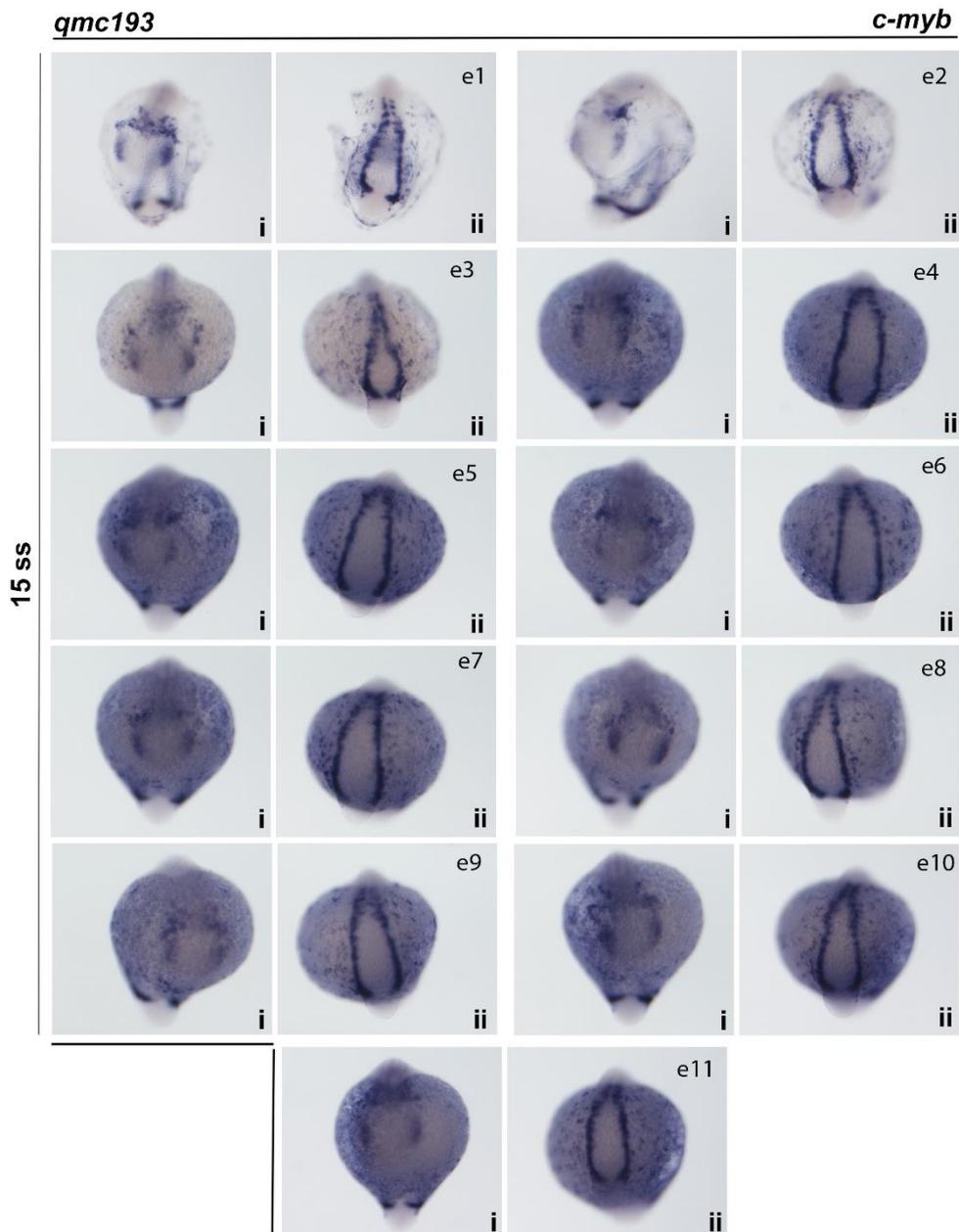






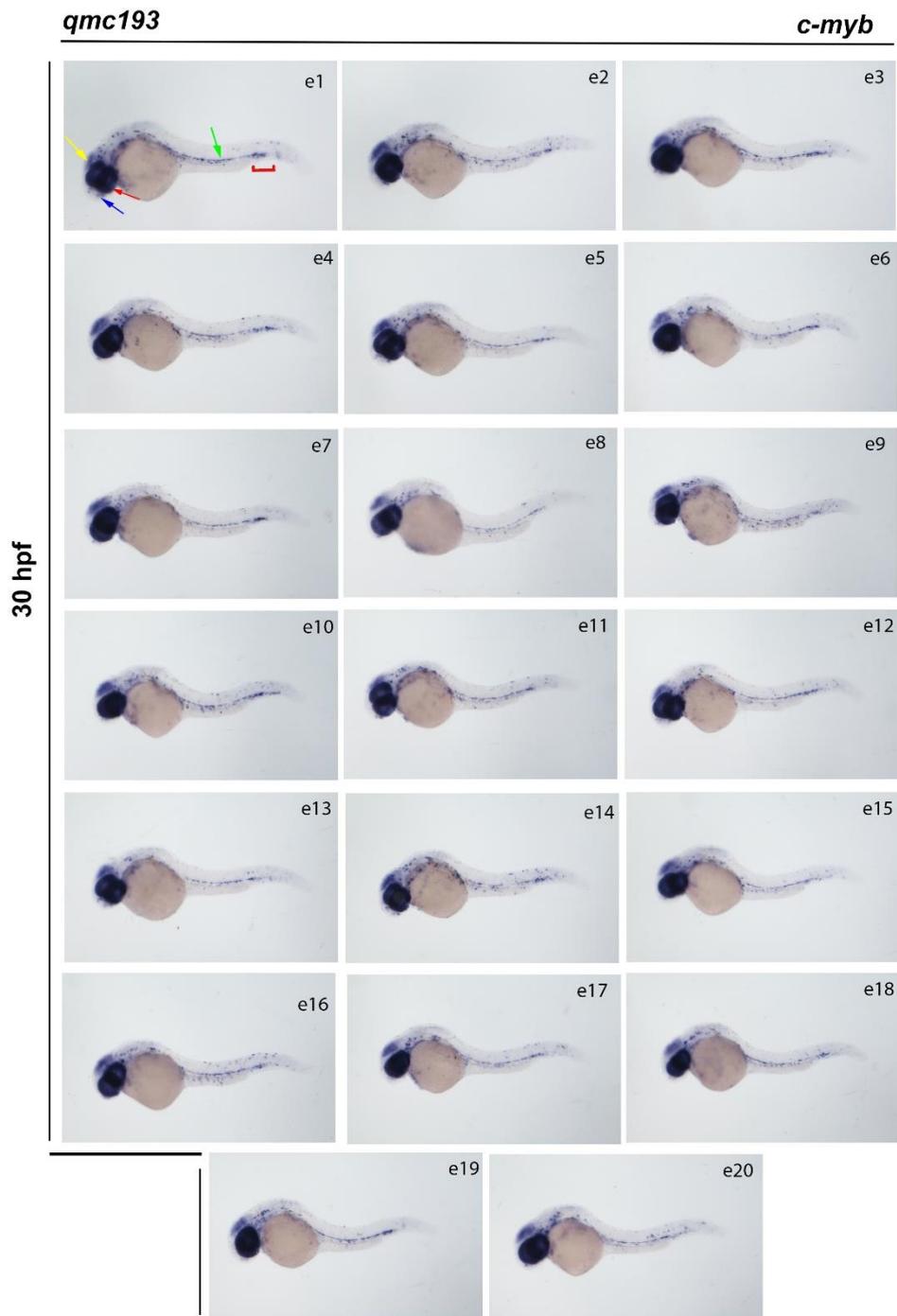
Appendix figure 6. 4 *lyz* expression was maintained in 34-36 hpf *qmc193* embryos

WT, *qmc193* heterozygous and homozygous embryos were showing *lyz* mRNA expression detected by WISH. Embryos were derived from an incross of *qmc193* heterozygous fish. All embryos are presented in a lateral view with anterior facing the left and dorsal up.



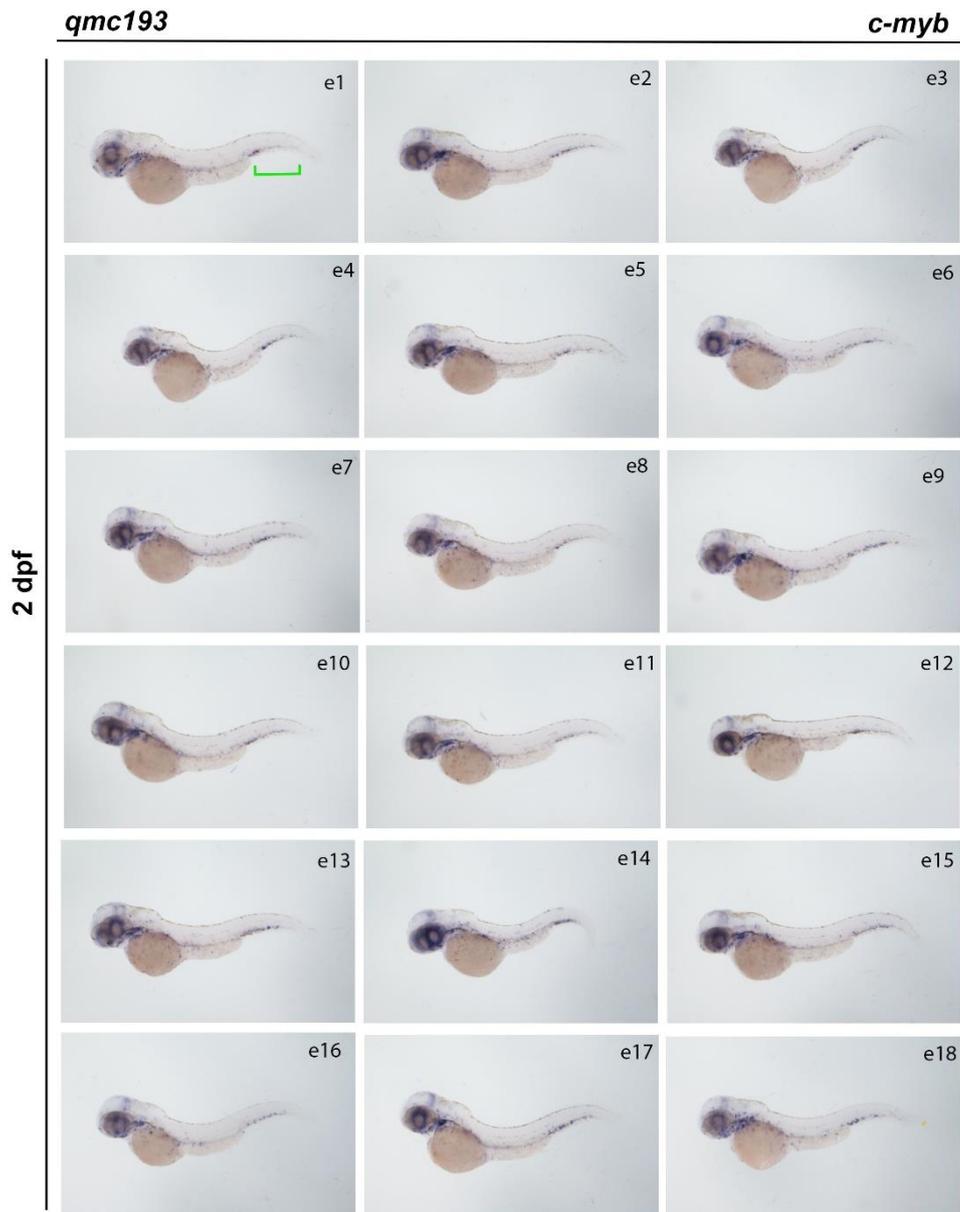
Appendix figure 6.5 *c-myb* expression was maintained in anterior primitive myeloid and posterior primitive erythroid progenitors of all 15 ss *qmc193* embryos

WT, *qmc193* heterozygous and homozygous embryos were derived from an incross of *qmc193* heterozygous fish and stained by WISH using a *c-myb* probe. All (i) embryos are presented in anterior view. All (ii) embryos are presented in a dorsoposterior view.



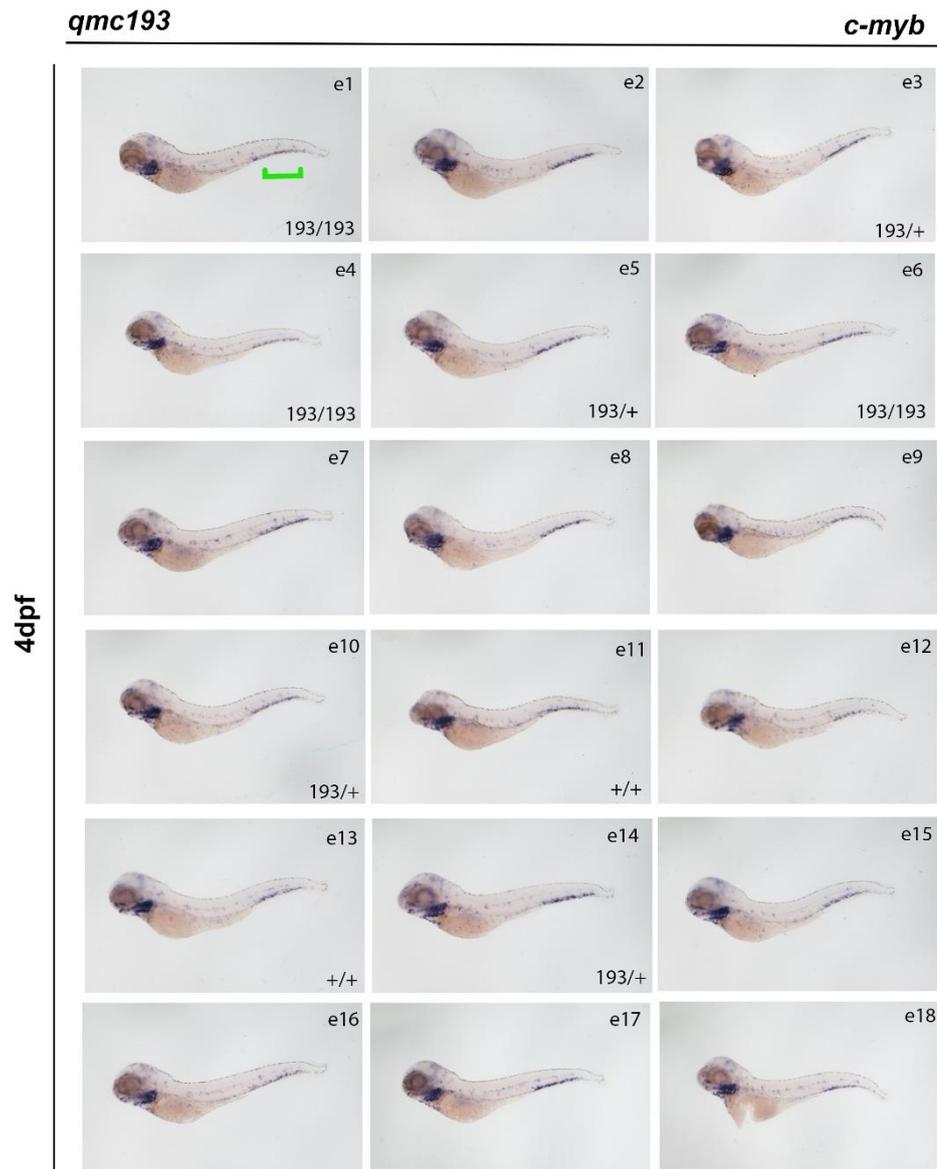
Appendix figure 6. 6 *c-myb* expression was maintained in 30 hpf *qmc193* embryos

WT, *qmc193* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. Embryos were derived from an incross of *qmc193* heterozygous fish. All embryos are presented in a lateral view with anterior facing the left and dorsal up. The coloured symbols represent expression in specific tissues: mid/hindbrain (yellow arrow), olfactory placodes (blue arrow), retina (red arrow), the ventral wall of the dorsal aorta (green arrow) and posterior blood island (red bracket).



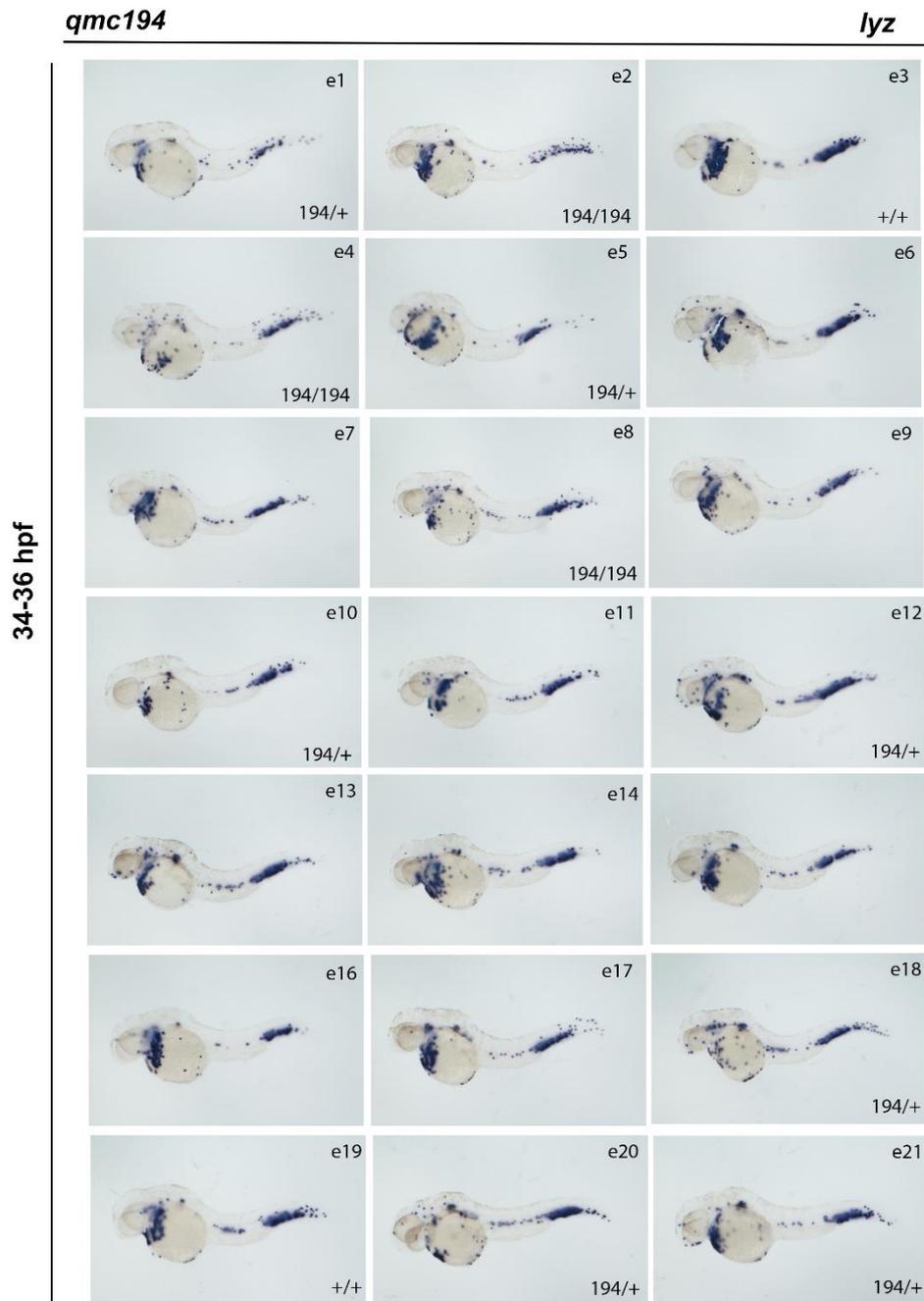
Appendix figure 6. 7 *c-myb* expression was maintained in 2 dpf *qmc193* embryos

WT, *qmc193* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. Embryos were derived from an incross of *qmc193* heterozygous fish. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT).



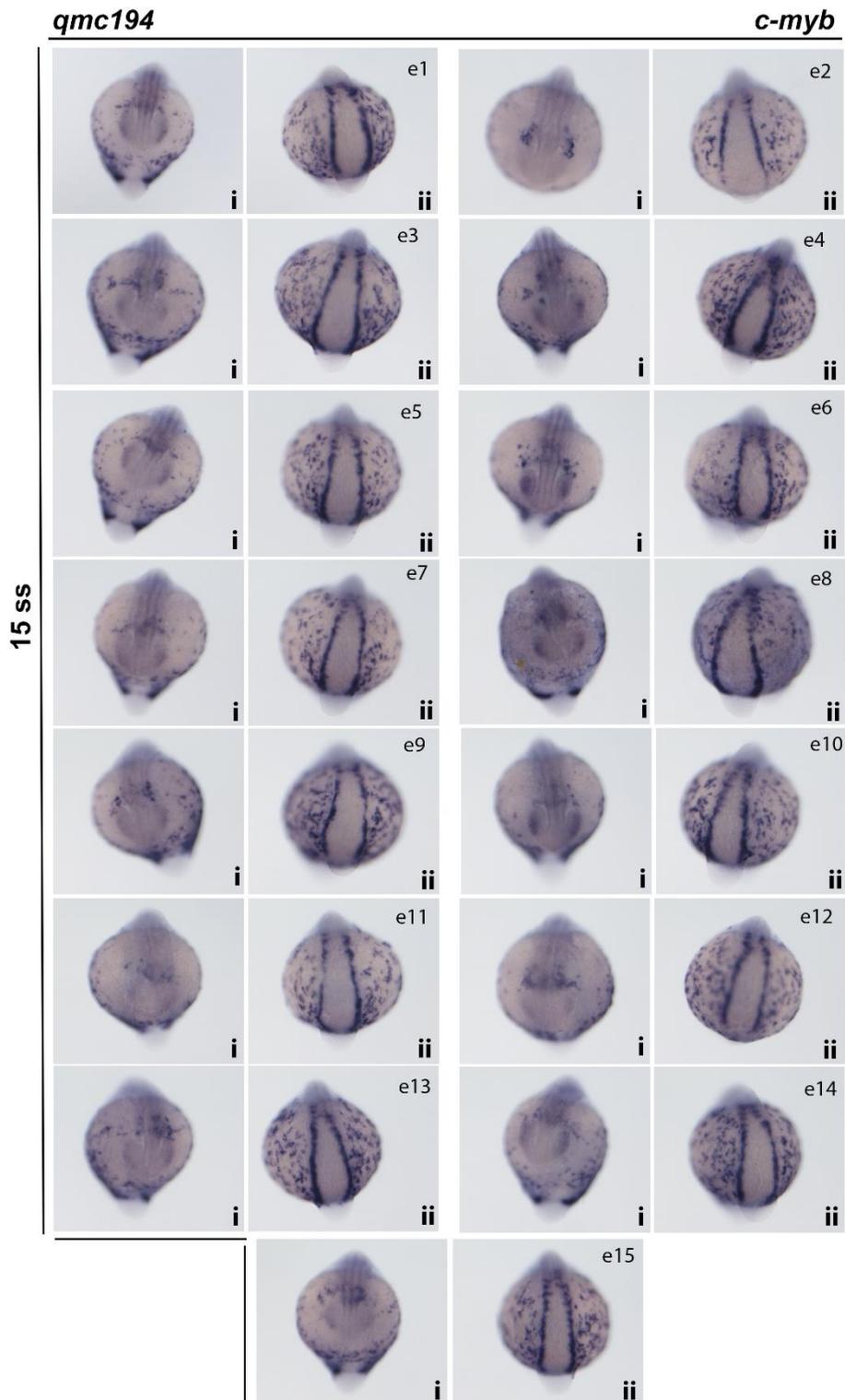
Appendix figure 6.8 *c-myb* expression was not altered in 4 dpf *qmc193* embryos

WT, *qmc193* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. Embryos were derived from an incross of *qmc193* heterozygous fish. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT).

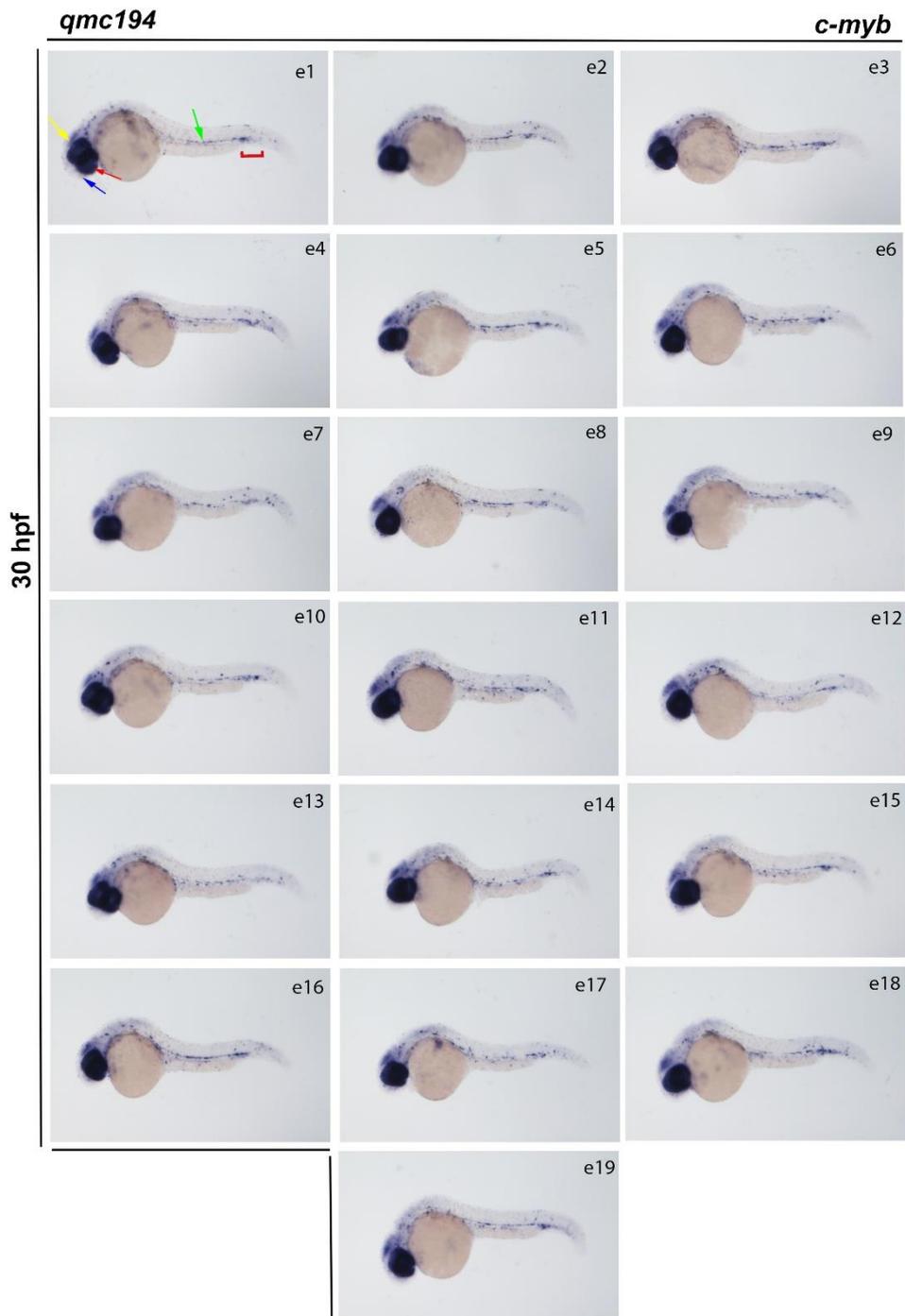


Appendix figure 6.9 *lyz* expression is unaffected in 34-36 hpf *qmc194* embryos

WT, *qmc194* heterozygous and homozygous embryos showing *lyz* mRNA expression detected by WISH. Embryos were derived from an incross of *qmc194* heterozygous fish. All embryos are presented in a lateral view with anterior facing the left and dorsal up.

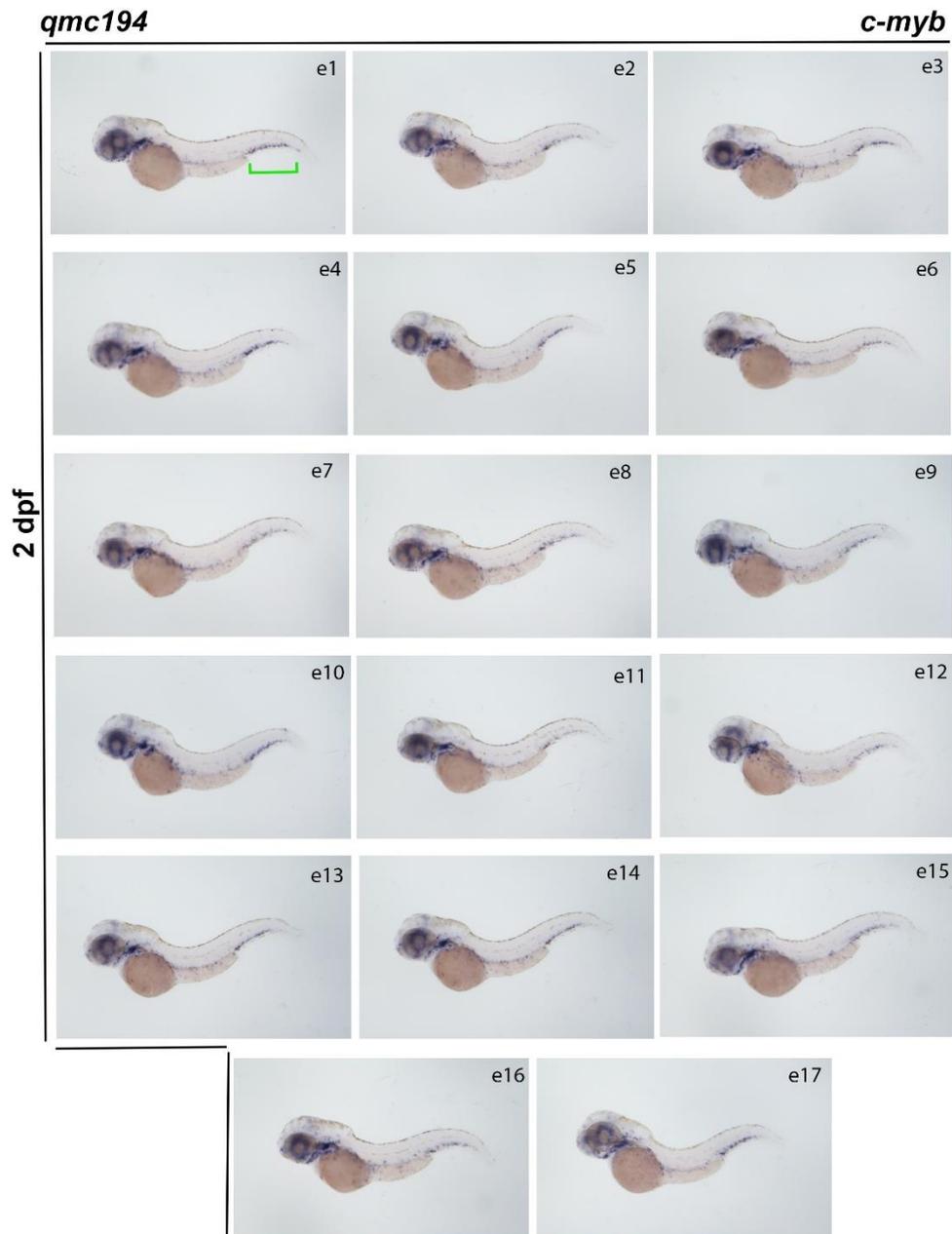


Appendix figure 6.10 *c-myb* expression was similar in all 15 ss embryos derived from an incross of *qmc194* heterozygous carriers WT, *qmc194* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. All (i) embryos are presented in anterior view. All (ii) embryos are presented in a dorsoposterior view.



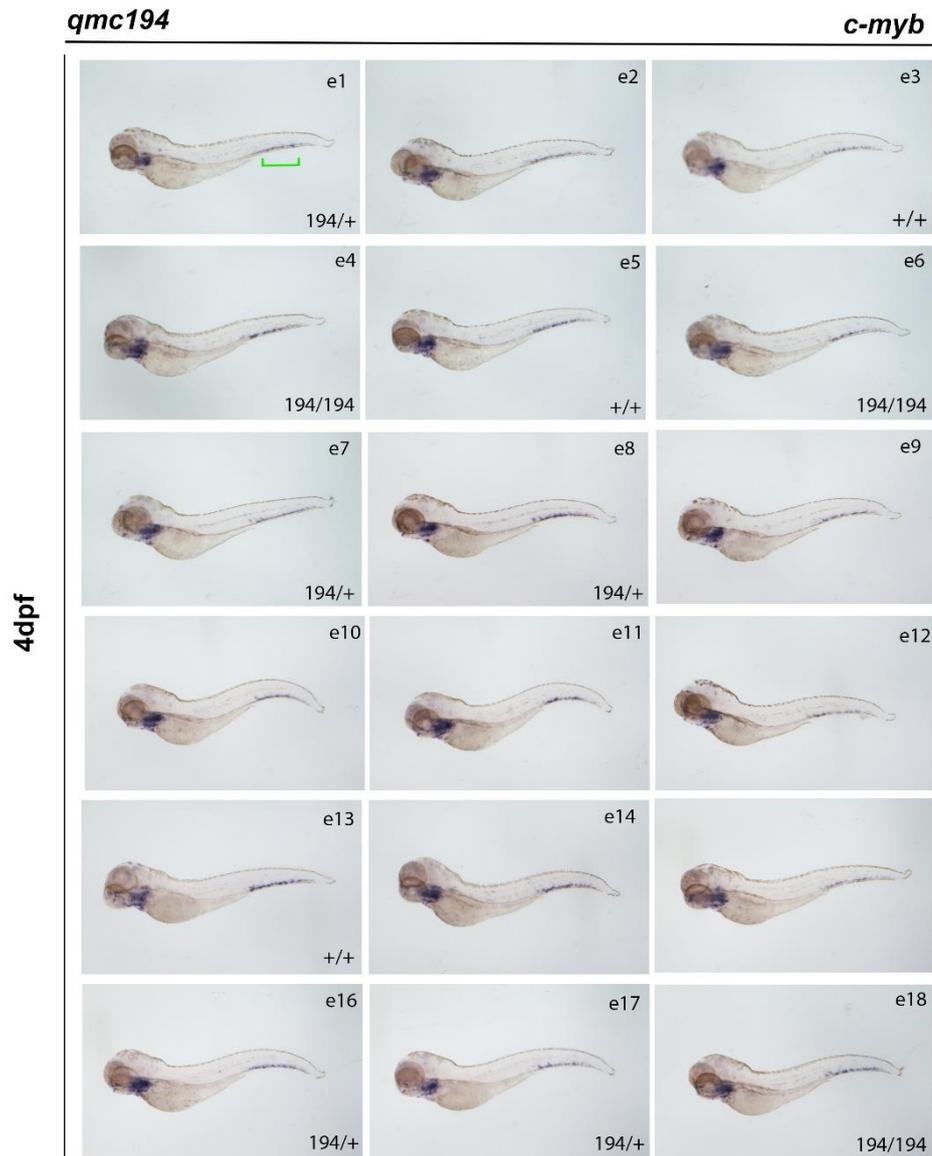
Appendix figure 6.11 The level of *c-myb* expression was comparable in all 30 hpf embryos derived from an incross of *qmc194* heterozygous carriers

WT, *qmc194* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. All embryos are presented in a lateral view with anterior facing the left and dorsal up. The coloured symbols represent expression in specific tissues: mid/hindbrain (yellow arrow), olfactory placodes (blue arrow), retina (red arrow), the ventral wall of the dorsal aorta (green arrow) and posterior blood island (red bracket).



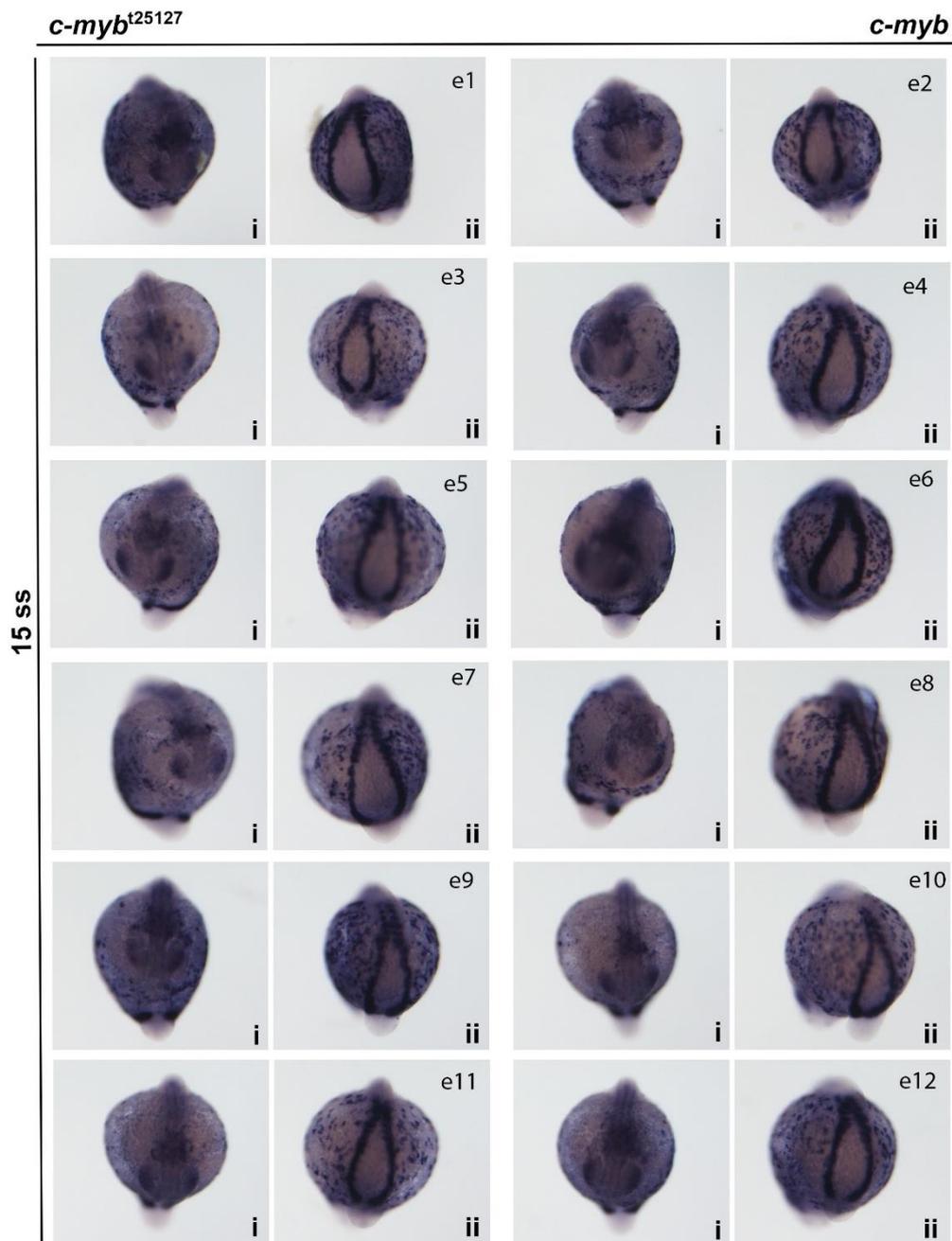
Appendix figure 6.12 The level of *c-myb* mRNA was similar in all 2 dpf embryos derived from an incross of *qmc194* heterozygous carriers

WT, *qmc194* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT).



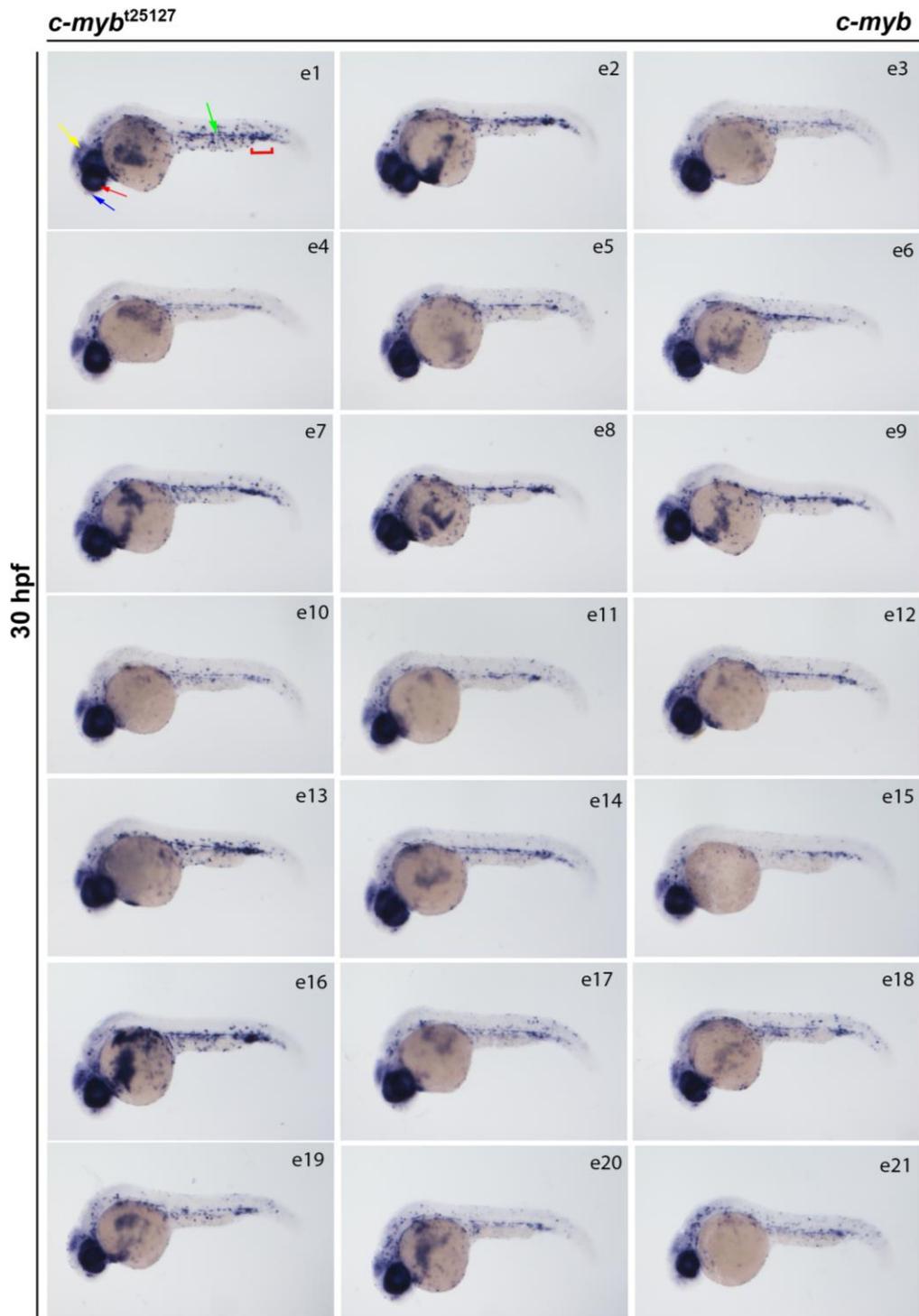
Appendix figure 6.13 *c-myb* expression is unaffected in 4 dpf *qmc194* embryos

WT, *qmc194* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. Embryos were derived from an incross of *qmc193* heterozygous fish. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT).



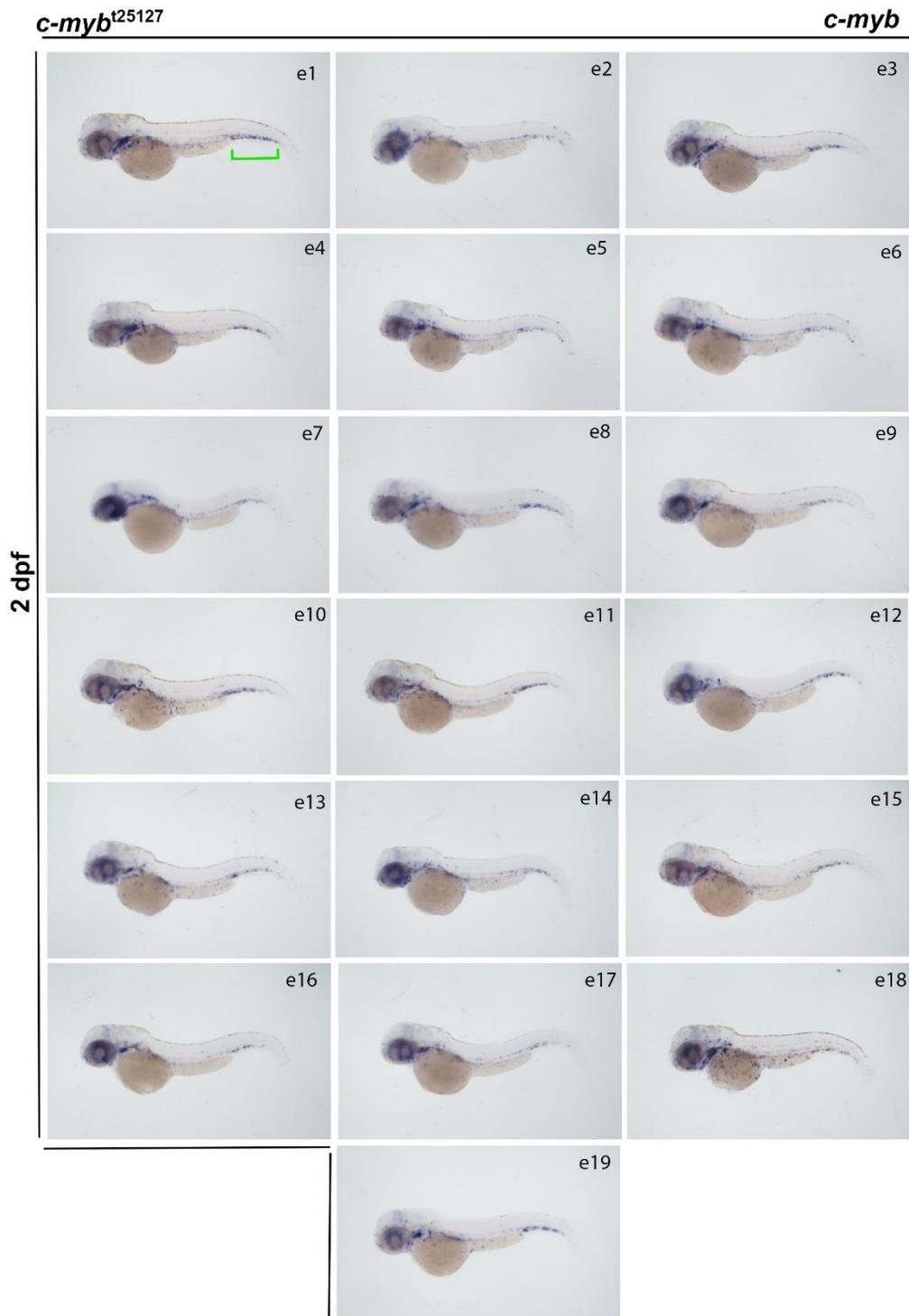
Appendix figure 6.14 *c-myb* expression was unaffected in 15 ss *t25127* embryos

WT, *t25127* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. All (i) embryos are presented in anterior view. All (ii) embryos are presented in a dorsoposterior view.



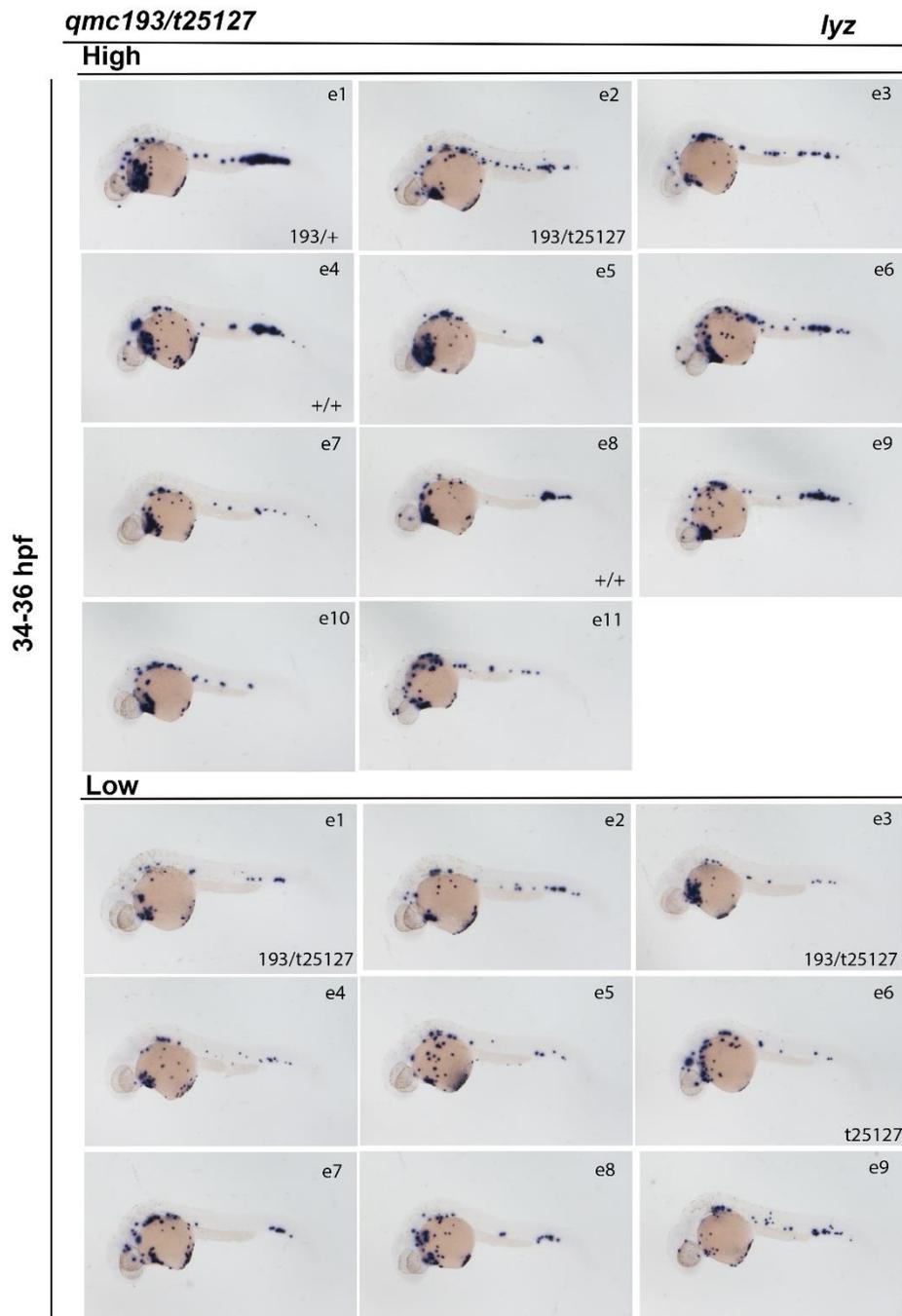
Appendix figure 6.15 *c-myb* expression was maintained in 30 hpf *t25127* embryos

WT, *t25127* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. All embryos are presented in a lateral view with anterior facing the left and dorsal up. The coloured symbols represent expression in specific tissues: mid/hindbrain (yellow arrow), olfactory placodes (blue arrow), retina (red arrow), the ventral wall of the dorsal aorta (green arrow) and *posterior blood island* (red bracket).



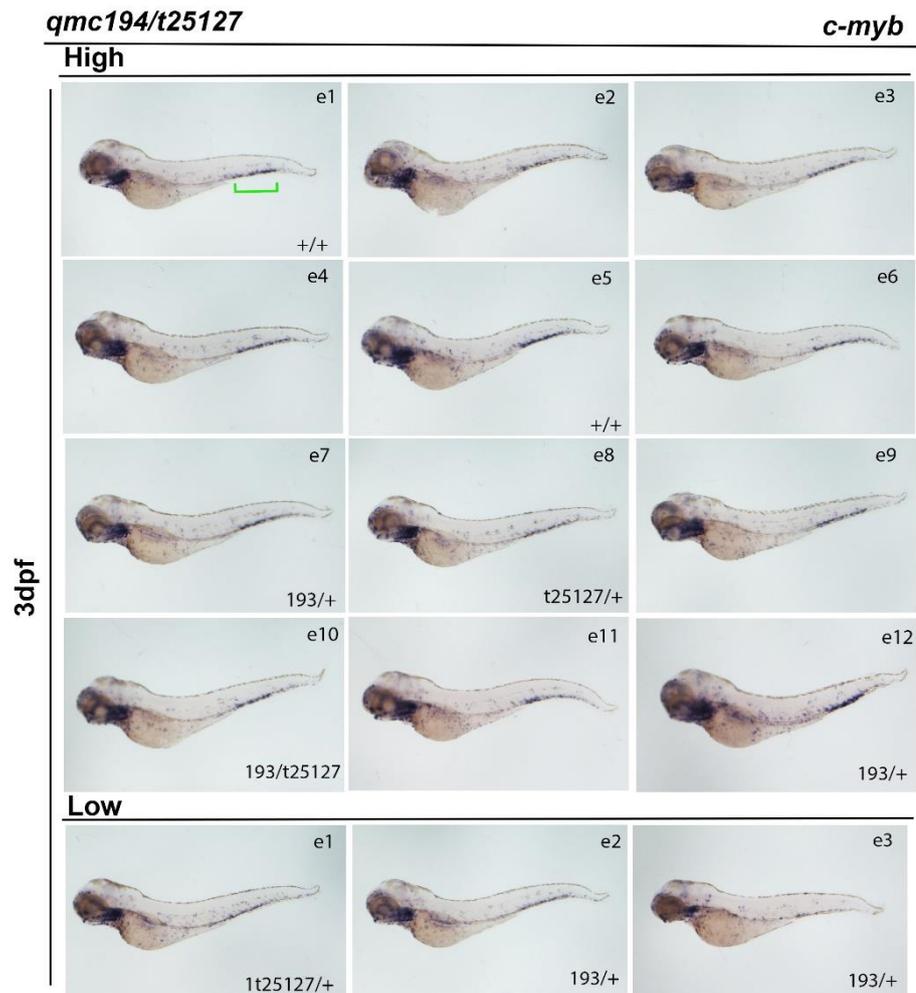
Appendix figure 6.16 *c-myb* expression was maintained in 2 dpf *t25127* embryos

WT, *t25127* heterozygous and homozygous embryos showing *c-myb* mRNA expression detected by WISH. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudle haematopoietic tissue (CHT).



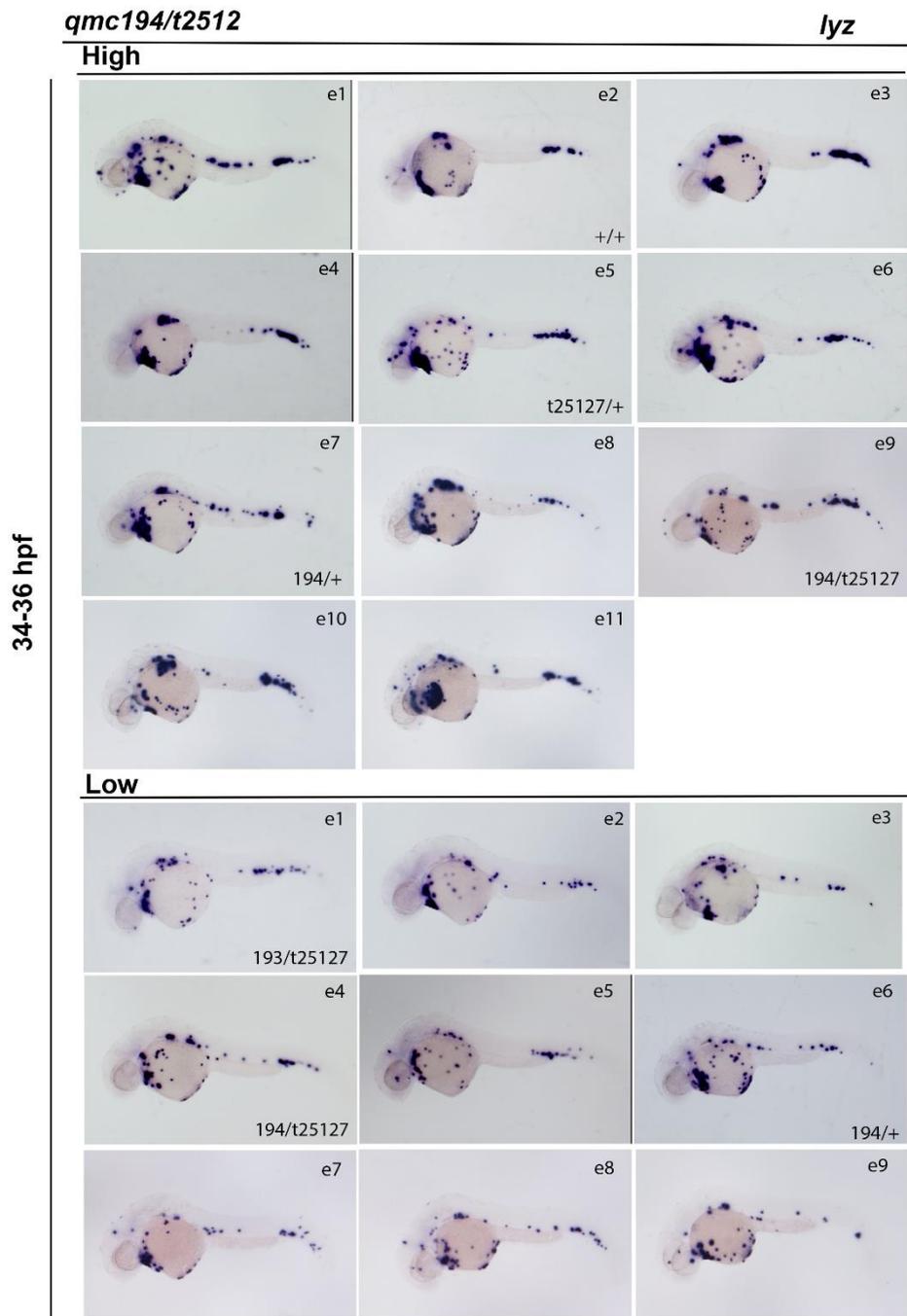
Appendix figure 6. 17 *Lyz* expression was maintained in 34-36 hpf in *c-myb*^{t25127/qmc193} embryos

Lateral view of embryos obtained from crossing a *qmc193* heterozygous carrier *t25127* heterozygous zebrafish stained in a WISH using *lyz* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up.



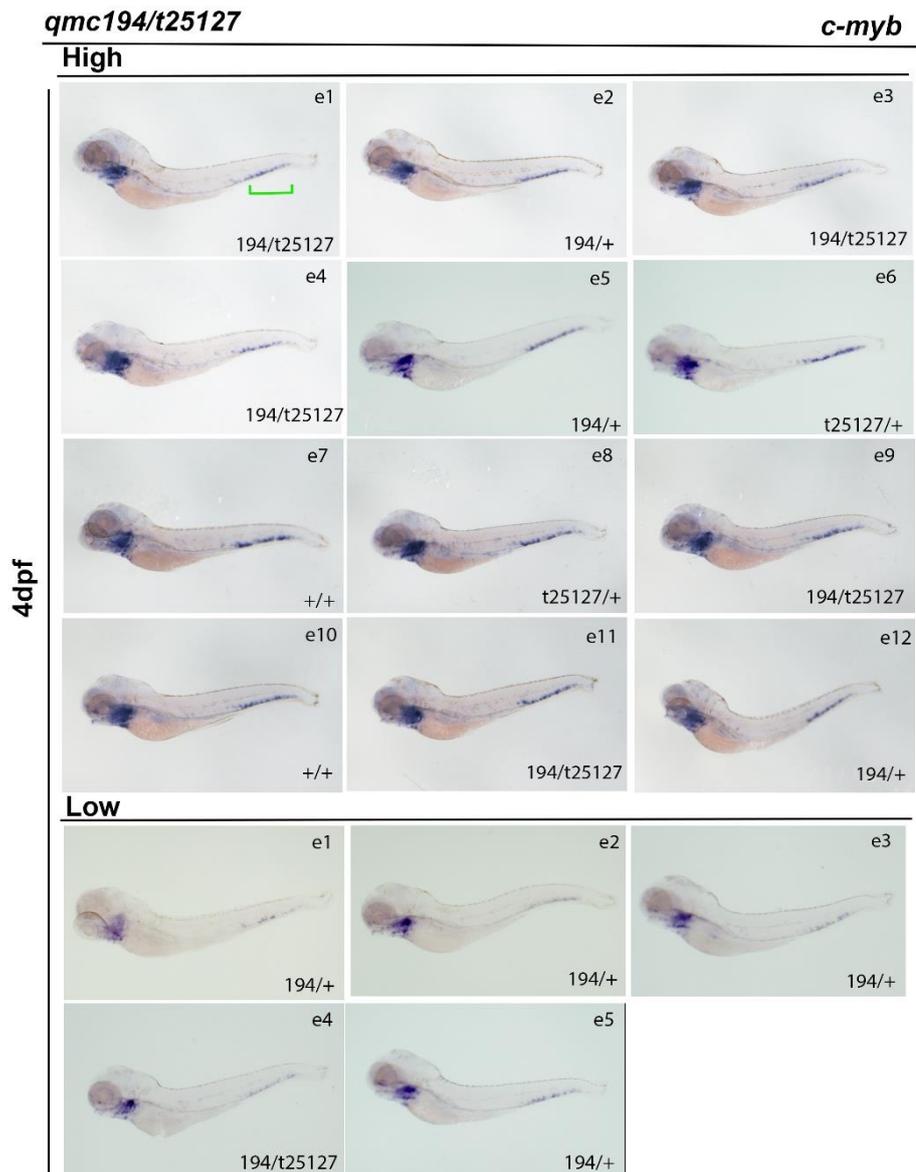
Appendix figure 6.18 *c-myb* expression was maintained in 3 dpf in *c-myb*^{t25127/qmc193} embryos

Lateral view of embryos obtained from crossing a *qmc193* heterozygous carrier *t25127* heterozygous zebrafish stained in a WISH using *c-myb* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT).



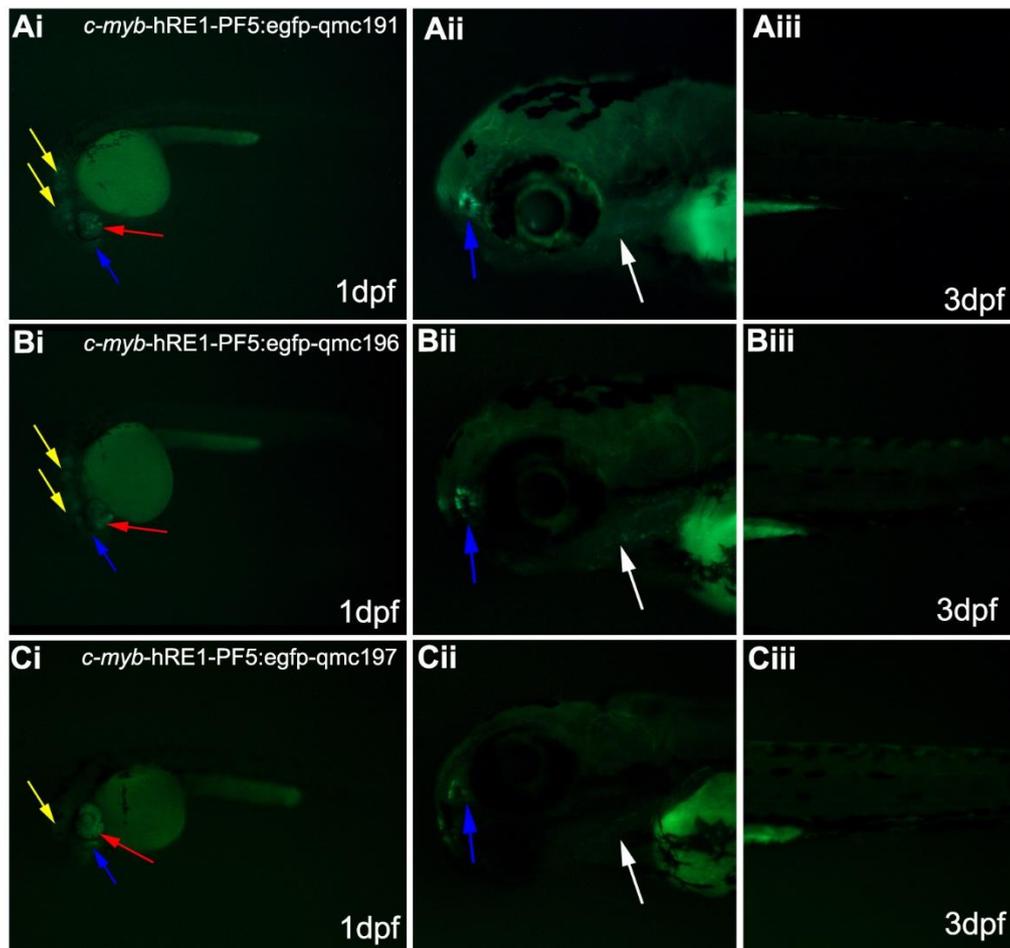
Appendix figure 6.19 The expression level of *lyz* in 34-36 hpf was not changed in the compound heterozygote of *c-myb*^{t25127/qmc194} embryos

Lateral view of embryos obtained from crossing a *qmc194* heterozygous carrier *t25127* heterozygous zebrafish stained in a WISH using *lyz* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up.



Appendix figure 6.20 The expression level of *c-myb* in 4 dpf was not altered in the compound heterozygote of *c-myb*^{t25127/qmc194} embryos

Lateral view of embryos obtained from crossing a *qmc194* heterozygous carrier *t25127* heterozygous zebrafish stained in a WISH using *c-myb* probe. All embryos are presented in a lateral view with anterior facing the left and dorsal up. Green bracket: caudal haematopoietic tissue (CHT).



Appendix figure 6.21 Identification of Tg(*c-myb*-hRE1-PF5:*egfp*) transgenic founders

(Ai, Bi and Ci) Lateral view of the *qmc191*, *qmc196* and *qmc197* embryos at 1 dpf. Anterior is to the left, posterior to the right, dorsal is up. Transgenic embryos show reporter expression in the OP, Re and BR at 1 dpf. **(Aii, Bii and Cii)** Transgenic embryos display eGFP expression in OP, and BA at 3 dpf. **(Aiii, Biii and Ciii)** The images depict embryos at 3 dpf and the expression of the reporter gene cannot be seen in the CHT region. The coloured symbols represent expression in specific tissues: mid/hindbrain (BR) (yellow arrow), olfactory placodes (OP) (blue arrow), retina (Re) (red arrow) and branchial arches (BA) (white arrow).

6.2 Appendix Tables

Appendix table 6.1 List of the differently expressed ($p < 0.05$) genes that are more highly expressed in cluster one than cluster two.

No	Genes ID	Gene Name	Cluster 1 Average	Cluster 1 Log2 Fold Change	Cluster 1 P-Value
1-	ENSDARG00000039579	cfb	1.519419439	4.355198228	3.93E-07
2-	ENSDARG00000023290	fabp3	1.988860141	4.257949094	2.99E-12
3-	ENSDARG00000062519	abcc13	2.202341295	3.337854584	4.66E-10
4-	ENSDARG00000075664	si:ch1073-429i10.1	11.23499033	3.153248219	6.31E-10
5-	ENSDARG00000080010	adh5	1.109884164	3.002004112	3.15E-08
6-	ENSDARG00000057789	lyz	509.7439827	2.939715046	1.83E-09
7-	ENSDARG00000021339	cpa5	17.87686829	2.817644836	2.61E-08
8-	ENSDARG00000056200	abcb9	1.41703562	2.735254576	2.08E-07
9-	ENSDARG00000058734	prdx1	2.326508905	2.590974571	5.71E-07
10-	ENSDARG00000010244	rpl22l1	4.828377329	2.58954865	6.53E-07
11-	ENSDARG00000094845	MFAP4 (1 of many)	1.071762529	2.549532146	1
12-	ENSDARG00000008155	sms	4.745598923	2.380894123	2.13E-06
13-	ENSDARG00000014165	ssr3	2.494244098	2.376854499	4.01E-06
14-	ENSDARG00000099572	hmgn2	27.42688237	2.34367546	2.51E-05
15-	ENSDARG00000036966	hsd3b7	1.033640895	2.183227345	3.93E-05
16-	ENSDARG000000102640	pdia3	2.122830457	2.086052524	4.63E-05
17-	ENSDARG00000053990	hmgb2b	1.659924893	2.014559694	0.00028377
18-	ENSDARG00000069100	aldh9a1a.1	1.556451884	1.939253025	0.00015181
19-	ENSDARG00000026369	dbi	2.616233329	1.926685458	0.00013644
20-	ENSDARG00000019444	ssr4	2.162041281	1.875837742	0.00023169
21-	ENSDARG00000007697	fabp7a	1.280886925	1.767976689	0.00100909
22-	ENSDARG00000005821	ncf2	1.504170785	1.755694402	0.0006279
23-	ENSDARG00000010423	npsn	69.41949675	1.74622968	0.00025589
24-	ENSDARG00000003795	idh2	1.604376225	1.711169787	0.00079062
25-	ENSDARG00000011201	rplp2l	29.86884537	1.707970154	0.00034173
26-	ENSDARG00000005926	ak2	3.768595885	1.677161061	0.0008713
27-	ENSDARG00000036875	rps12	31.00051332	1.611101438	0.00078171
28-	ENSDARG00000005230	ssr2	2.303635924	1.595342969	0.00209182
29-	ENSDARG00000028335	hmga1a	2.082530443	1.577075404	0.00348088
30-	ENSDARG000000105116	p4hb	1.032551705	1.575986858	0.00906009
31-	ENSDARG00000019230	rpl7a	25.74517368	1.569038893	0.00103203
32-	ENSDARG00000003032	eif4a1b	1.553184316	1.565854763	0.00246127
33-	ENSDARG00000003599	rpl3	21.26206944	1.556066186	0.00122777
34-	ENSDARG00000038728	ch25hl2	4.875212481	1.496238974	0.00363358
35-	ENSDARG00000007320	rpl7	22.26085627	1.494701881	0.00188917
36-	ENSDARG00000056119	eef1g	11.24697142	1.478482939	0.00329034
37-	ENSDARG000000116491	h2afvb	2.988736159	1.462808972	0.00719418
38-	ENSDARG00000076532	si:ch211-222i21.1	5.824985779	1.461966826	0.00438204
39-	ENSDARG00000056600	papss2b	2.102135855	1.435517298	0.00619765

40-	ENSDARG00000102291	eef1da	3.892763495	1.433857781	0.00515629
41-	ENSDARG0000014690	rps4x	20.75232873	1.394375957	0.0040869
42-	ENSDARG0000035692	rps3a	19.19369847	1.3839303	0.00452839
43-	ENSDARG0000042708	tuba8l	3.903655391	1.379589018	0.02255185
44-	ENSDARG0000070426	chac1	3.646606654	1.359613332	0.00930759
45-	ENSDARG0000044521	eef1b2	11.26222007	1.346962661	0.00842578
46-	ENSDARG0000076568	sec61b	2.500779235	1.343916119	0.0103632
47-	ENSDARG0000009285	rpl15	23.4262891	1.319676296	0.00692521
48-	ENSDARG0000029533	rpl18	24.61023816	1.309398862	0.00756047
49-	ENSDARG0000058451	rpl6	18.30491978	1.304885748	0.00797929
50-	ENSDARG0000054155	pcna	1.189395002	1.294373439	0.03675172
51-	ENSDARG0000051783	rplp0	34.50770371	1.277080684	0.00926127
52-	ENSDARG0000030408	rps26l	16.0699028	1.275904613	0.00985129
53-	ENSDARG0000043848	sod1	1.500903217	1.267132416	0.01855365
54-	ENSDARG0000055996	rps8a	21.73804528	1.266867403	0.01008126
55-	ENSDARG0000037350	rpl9	18.35937926	1.251512554	0.01101999
56-	ENSDARG0000015862	rpl5b	8.427059643	1.228924873	0.01949359
57-	ENSDARG0000020197	rpl5a	6.688713102	1.214623052	0.02086569
58-	ENSDARG0000042566	rps7	25.73428179	1.200470218	0.01585113
59-	ENSDARG0000011405	rps9	21.63348309	1.197226231	0.01690495
60-	ENSDARG00000104011	rps17	11.02804432	1.184144428	0.01787099
61-	ENSDARG0000090697	eif3ea	2.170754798	1.17654245	0.03079487
62-	ENSDARG00000101844	mibp2	7.01764835	1.142316889	0.03280915
63-	ENSDARG00000100371	eef2b	7.66353776	1.122044646	0.03718342
64-	ENSDARG00000092807	RPL41	72.72954383	1.112617731	0.02796234
65-	ENSDARG0000025073	rpl18a	26.81149026	1.111202234	0.02872562
66-	ENSDARG0000042905	rpl10a	20.58459354	1.109712224	0.02897586
67-	ENSDARG0000041619	rack1	16.99244636	1.097890245	0.03076302
68-	ENSDARG0000014867	rpl8	23.78572166	1.091993176	0.03184113
69-	ENSDARG0000036629	rps14	19.13270385	1.081905065	0.03366779
70-	ENSDARG0000034897	rps10	16.56221648	1.072642497	0.03566248
71-	ENSDARG0000057556	rpl17	20.85362336	1.071889636	0.03569184
72-	ENSDARG0000019778	rps6	22.78475645	1.068253018	0.03675172
73-	ENSDARG0000043453	rps5	22.82069971	1.067899127	0.03641769
74-	ENSDARG0000033170	sult2st1	5.612593814	1.05637	0.0560572
75-	ENSDARG0000046157	zgc:114188	12.0627744	1.054167638	0.04018721
76-	ENSDARG0000099380	rpl13	19.33093635	1.053938774	0.03996739
77-	ENSDARG0000053457	rpl23	28.79272608	1.046302671	0.04205813
78-	ENSDARG0000041182	rpl4	10.14035482	1.045529693	0.04302683
79-	ENSDARG0000021864	rplp1	37.88419135	1.039625671	0.04299335
80-	ENSDARG0000029500	rpl34	22.89367541	1.03877136	0.04286911
81-	ENSDARG0000099022	faua	16.19842717	1.034265059	0.0455082
82-	ENSDARG0000034291	rpl37	32.81728151	1.026411425	0.04669993
83-	ENSDARG0000043509	rpl11	25.59813309	1.026086542	0.04718108
84-	ENSDARG0000054818	rpl32	25.78329532	1.022342759	0.04781362

Appendix table 6.2 List of the differently expressed ($p < 0.05$) genes that are more highly expressed in cluster two than cluster one.

No	Gene ID	Gene Name	Cluster 2 Average	Cluster 2 Log2 Fold Change	Cluster 2 P-Value
1-	ENSDARG00000017299	fabp11a	1.53635652	7.13919071	0.00927853
2-	ENSDARG000000091996	si:ch211-117m20.5	9.44685239	6.3770975	2.69E-59
3-	ENSDARG000000052088	cxcr1	1.98881103	5.30889087	1.38E-44
4-	ENSDARG000000026500	xkr9	1.07892998	5.09519773	5.95E-36
5-	ENSDARG00000102435	plekhf1	3.12740534	4.94504361	1.40E-42
6-	ENSDARG000000090730	cfbl	3.13237737	4.66631163	1.40E-42
7-	ENSDARG000000013598	tnfb	2.15288794	4.5715062	3.37E-27
8-	ENSDARG000000055723	hsp70l	3.81354515	4.50317036	2.47E-30
9-	ENSDARG000000025903	lgals9l1	1.75512573	4.44314659	1.31E-31
10-	ENSDARG000000026726	anxa1a	1.43194394	4.42918275	1.38E-26
11-	ENSDARG000000025428	socs3a	1.16842648	4.36719773	7.55E-21
12-	ENSDARG000000099411	zgc:158343	1.32753136	4.32034626	1.76E-17
13-	ENSDARG000000074851	s1pr4	6.86139805	4.31728399	2.11E-41
14-	ENSDARG000000056499	ca6	1.99378306	4.26619049	6.76E-34
15-	ENSDARG000000056783	raraa	3.30639834	4.20682076	2.00E-35
16-	ENSDARG000000092362	hsp70.2	5.8421324	3.90085097	2.02E-22
17-	ENSDARG000000024746	hsp90aa1.2	13.5139709	3.83064247	7.33E-31
18-	ENSDARG000000045549	bik	1.03915376	3.7699569	5.46E-20
19-	ENSDARG000000029688	hsp70.1	2.24238444	3.56222987	3.58E-17
20-	ENSDARG000000087623	BX649485.1	1.43691597	3.54472453	3.87E-21
21-	ENSDARG000000089706	ANPEP (1 of many)	2.95338438	3.5069225	2.65E-25
22-	ENSDARG000000021924	hsp70.3	5.6084471	3.50245693	2.38E-19
23-	ENSDARG000000098761	rgs2	2.2175243	3.47342559	3.29E-21
24-	ENSDARG000000111188	cdaa	4.96208352	3.3549194	2.65E-25
25-	ENSDARG000000068784	vsir	7.29396445	3.29523972	2.48E-25
26-	ENSDARG000000075261	timp2b	5.37973383	3.26863245	1.92E-24
27-	ENSDARG000000074262	nck1a	1.86451034	3.19268879	3.65E-20
28-	ENSDARG000000079903	si:ch73-343l4.2	1.20323067	3.19094367	3.31E-18
29-	ENSDARG000000025254	s100a10b	17.9440475	3.13395497	2.95E-19
30-	ENSDARG000000039034	marcks1a	1.16345445	3.08423143	1.73E-16
31-	ENSDARG000000005481	nfkbiaa	4.18147519	2.93112727	1.04E-15
32-	ENSDARG000000056874	lygl1	1.44686002	2.90721341	4.03E-14
33-	ENSDARG00000104474	il6r	9.40707617	2.89902559	8.29E-20
34-	ENSDARG00000101482	hk2	1.01429362	2.89869022	8.74E-15
35-	ENSDARG000000093124	scpp8	11.3063907	2.88646499	9.14E-20
36-	ENSDARG00000114451	mmp13a	39.7165563	2.86657392	3.59E-18
37-	ENSDARG000000042816	mmp9	20.54939	2.84682567	2.27E-19
38-	ENSDARG000000077499	plekho1b	1.03418174	2.82626434	2.31E-13
39-	ENSDARG000000044039	stx11a	1.75015371	2.70015154	1.23E-14

40-	ENSDARG00000103720	zgc:162730	2.17774808	2.69628499	6.89E-14
41-	ENSDARG00000041394	dnajb1b	3.82846123	2.67754937	4.59E-15
42-	ENSDARG00000011934	gyg1a	2.77439139	2.66784269	9.77E-15
43-	ENSDARG00000054543	samsn1a	5.1758807	2.60465445	1.39E-15
44-	ENSDARG00000079387	si:ch211-102c2.4	2.85394383	2.6042197	7.93E-15
45-	ENSDARG00000078824	si:ch211-66e2.3	3.90304165	2.59149161	3.03E-15
46-	ENSDARG00000059049	zgc:174904	2.14791591	2.57253021	1.84E-13
47-	ENSDARG00000093193	chia.6	3.43567105	2.5589267	2.34E-14
48-	ENSDARG00000090552	si:dkey-7j14.6	1.24797892	2.51841814	5.05E-12
49-	ENSDARG00000041060	lgals9l3	1.11373418	2.49892971	4.10E-11
50-	ENSDARG00000044852	wbp2nl	5.30515342	2.47750601	3.12E-14
51-	ENSDARG00000075666	tsc22d3	3.72902068	2.41123401	2.70E-12
52-	ENSDARG00000055186	ccr9a	2.92355221	2.40697865	2.28E-09
53-	ENSDARG00000031745	si:busm1-266f07.2	1.46674813	2.40009433	1.19E-10
54-	ENSDARG00000035858	cnn2	2.12305577	2.34149889	1.74E-11
55-	ENSDARG00000068434	h3f3b.1	3.65941229	2.31605252	5.18E-12
56-	ENSDARG00000018569	tnfrsf1a	1.73523762	2.27865351	1.43E-10
57-	ENSDARG00000104077	fcer1gl	23.3337254	2.27503359	2.15E-12
58-	ENSDARG00000098787	cd7al	9.39713211	2.26660413	6.66E-12
59-	ENSDARG00000035557	gabarapa	2.39154526	2.25867995	6.48E-11
60-	ENSDARG00000090767	si:ch211-136m16.8	2.09819564	2.24773016	1.64E-10
61-	ENSDARG00000030440	rsrp1	5.24548909	2.23928456	2.12E-11
62-	ENSDARG00000086337	si:dkey-102g19.3	5.42945411	2.19559276	6.48E-11
63-	ENSDARG00000014348	stk17b	2.99813263	2.18022516	1.57E-10
64-	ENSDARG00000067916	lyn	1.09384607	2.17199706	9.55E-09
65-	ENSDARG00000098774	si:zfos-741a10.3	1.62088099	2.15443694	2.84E-09
66-	ENSDARG00000037091	asah1a	1.14356634	2.14192568	9.82E-09
67-	ENSDARG00000033735	ncf1	11.5897963	2.13633228	3.66E-10
68-	ENSDARG00000041505	itm2bb	6.65257289	2.13210376	1.34E-10
69-	ENSDARG00000056615	cybb	1.60099288	2.11950532	5.03E-09
70-	ENSDARG00000010317	gpr183a	1.01429362	2.109587	3.42E-08
71-	ENSDARG00000039269	arg2	3.89806962	2.09937056	6.90E-09
72-	ENSDARG00000098722	si:ch73-170d6.3	1.81976209	2.08992081	4.18E-09
73-	ENSDARG00000094485	si:dkey-27h10.2	1.71534951	2.08824117	7.32E-09
74-	ENSDARG00000089362	grn1	3.75885285	2.08736579	0.22052235
75-	ENSDARG00000057698	ctsd	2.49595784	2.0825934	2.66E-05
76-	ENSDARG00000055365	si:dkey-25e12.3	3.16718156	2.06153785	2.99E-09
77-	ENSDARG00000074782	adgrg3	1.41702786	2.03652671	3.33E-08
78-	ENSDARG00000088040	si:dkeyp-27c8.2	1.4965803	2.03538186	2.61E-08
79-	ENSDARG00000030938	fermt3b	1.07892998	2.01066422	1.17E-07
80-	ENSDARG00000044125	txn	3.37600672	2.0019283	1.06E-08

81-	ENSDARG00000036967	smox	1.91423062	1.99937431	5.41E-08
82-	ENSDARG00000058348	scinlb	5.04163596	1.99858128	3.56E-09
83-	ENSDARG00000089368	hopx	2.86885991	1.98488065	1.25E-08
84-	ENSDARG00000008363	mcl1b	4.75823039	1.98296774	7.75E-09
85-	ENSDARG00000034187	calm1b	2.74455922	1.97871746	1.35E-08
86-	ENSDARG00000077975	si:ch73-34314.8	4.5444332	1.97822822	6.77E-09
87-	ENSDARG00000056532	serinc2	1.10379012	1.97466705	1.73E-07
88-	ENSDARG00000018146	gpx1a	5.40459397	1.96753557	8.93E-09
89-	ENSDARG00000043555	tmem30ab	1.18334256	1.95143489	1.83E-07
90-	ENSDARG00000104118	antxr2a	3.01802074	1.95112248	2.07E-08
91-	ENSDARG00000044694	fybb	2.38160121	1.93960016	3.56E-08
92-	ENSDARG00000067797	spi1a	1.25295095	1.9343683	3.35E-07
93-	ENSDARG00000087152	sowahd	1.24797892	1.93345564	1.88E-07
94-	ENSDARG00000063905	mt-co1	25.238012	1.90744622	1.31E-08
95-	ENSDARG00000008075	illr4	6.08078972	1.8810943	3.56E-08
96-	ENSDARG00000117781	CABZ01073834.1	2.17774808	1.86503945	2.24E-07
97-	ENSDARG00000021059	alas1	1.82970615	1.85449041	3.76E-07
98-	ENSDARG00000007823	atf3	2.38160121	1.85348026	7.37E-07
99-	ENSDARG00000099839	si:ch211-9d9.1	14.8514464	1.85344189	5.56E-08
100-	ENSDARG00000088641	grn2	2.23741241	1.85223091	0.25431069
101-	ENSDARG00000012513	sdcbp2	4.69856606	1.83467658	1.30E-07
102-	ENSDARG00000019033	tmem59	1.25792298	1.82938318	9.83E-07
103-	ENSDARG00000078619	pnp5a	2.78433544	1.82299591	4.43E-07
104-	ENSDARG00000010384	lpar5a	1.97389495	1.81880401	6.56E-07
105-	ENSDARG00000098443	zfand6	2.36668512	1.81839726	3.71E-07
106-	ENSDARG00000078547	si:ch211-264f5.2	2.93349627	1.81080814	3.44E-07
107-	ENSDARG00000063914	mt-nd3	5.29520937	1.80523405	1.99E-07
108-	ENSDARG00000071251	ppp1r18	1.37725164	1.80010974	1.35E-06
109-	ENSDARG00000032631	ltb4r	2.51584595	1.78929339	4.81E-07
110-	ENSDARG00000088283	si:ch73-248e21.5	3.90801367	1.78423429	2.63E-07
111-	ENSDARG00000113899	zgc:77650	11.8682298	1.77494972	2.14E-07
112-	ENSDARG00000015887	b2ml	3.99253814	1.76324356	4.80E-07
113-	ENSDARG00000007682	ppdpfa	9.18830696	1.75849378	3.88E-07
114-	ENSDARG00000104636	si:dkey-112a7.4	2.13797186	1.75166312	1.66E-06
115-	ENSDARG00000086996	si:ch73-234b20.5	1.71534951	1.74925791	1.61E-06
116-	ENSDARG00000007098	itm2ba	1.2976992	1.74674219	4.08E-06
117-	ENSDARG00000068233	zgc:64051	4.56929334	1.71081233	1.09E-06
118-	ENSDARG00000114387	CU499336.2	4.67370592	1.70659334	1.24E-06
119-	ENSDARG00000031317	ppdpfb	2.44623757	1.70627195	3.30E-06
120-	ENSDARG00000109289	mrps15	2.6500907	1.70459372	2.11E-06

121-	ENSDARG00000038068	ddx5	4.23616749	1.69739466	1.30E-06
122-	ENSDARG00000008735	atp6ap2	2.26724457	1.69686421	3.20E-06
123-	ENSDARG00000018283	cyba	5.08638421	1.6861452	1.67E-06
124-	ENSDARG00000105530	si:dkey-19b23.14	1.57613274	1.66498734	1.37E-05
125-	ENSDARG00000105068	SMIM22	3.33125847	1.66227402	2.45E-06
126-	ENSDARG00000098293	si:dkey-27i16.2	6.86637008	1.65319543	4.43E-06
127-	ENSDARG00000099970	CR383676.1	192.168866	1.64580308	1.94E-06
128-	ENSDARG00000053836	si:ch211-284o19.8	1.95897886	1.63827574	1.33E-05
129-	ENSDARG00000092870	BX936337.1	1.00434957	1.6360112	3.96E-05
130-	ENSDARG00000073718	si:ch73-233k15.2	1.16345445	1.6310588	2.59E-05
131-	ENSDARG00000016939	itgb2	2.00372711	1.62161876	9.99E-06
132-	ENSDARG00000099175	hmgb1a	2.7594753	1.62094089	7.56E-06
133-	ENSDARG00000045490	ndufb2	1.48663624	1.60645817	2.80E-05
134-	ENSDARG00000053136	b2m	17.5114811	1.60536535	5.29E-06
135-	ENSDARG00000036588	mhc1zba	7.18955187	1.60169457	6.22E-06
136-	ENSDARG00000063917	mt-nd4	5.44437019	1.59390208	8.02E-06
137-	ENSDARG00000006266	stat1a	1.71534951	1.59171663	2.49E-05
138-	ENSDARG00000063895	mt-nd1	5.16096462	1.58758884	1.00E-05
139-	ENSDARG00000030012	lrrfip1a	1.90925859	1.58141622	1.98E-05
140-	ENSDARG00000018145	mid1ip1l	2.15785997	1.57917859	2.70E-05
141-	ENSDARG00000093313	ubb	3.90304165	1.57456121	1.30E-05
142-	ENSDARG00000036282	rnaset2	4.92727933	1.57313667	1.24E-05
143-	ENSDARG00000105096	si:ch1073-67j19.1	1.19328662	1.56650228	0.04599819
144-	ENSDARG00000042927	mapre1a	6.55810437	1.56334209	1.07E-05
145-	ENSDARG00000004311	ldlrap1a	1.55624463	1.55777421	4.11E-05
146-	ENSDARG00000013655	tpd52l2b	1.6954614	1.54756448	3.10E-05
147-	ENSDARG00000012874	snap23.1	1.30267122	1.54443329	5.98E-05
148-	ENSDARG00000057853	atp6v0ca	1.45183205	1.5437009	7.06E-05
149-	ENSDARG00000001452	adam8a	3.65444027	1.54364943	2.05E-05
150-	ENSDARG00000100729	tln1	1.49160827	1.54209713	5.12E-05
151-	ENSDARG00000091902	b3gnt2b	1.18831459	1.52964257	8.75E-05
152-	ENSDARG00000074656	ctss2.1	5.89185267	1.52728885	2.94E-05
153-	ENSDARG00000104181	tmsb1	19.4207397	1.52578517	1.76E-05
154-	ENSDARG00000063912	mt-co3	41.0788918	1.50635517	2.16E-05
155-	ENSDARG00000033614	rasgef1ba	2.40148932	1.50580514	9.75E-05
156-	ENSDARG00000062049	hmha1b	1.7302656	1.50022067	6.58E-05
157-	ENSDARG00000070792	lrrc15	1.91423062	1.49900198	7.36E-05
158-	ENSDARG00000042824	nfe2l2a	1.54132855	1.49422162	9.02E-05
159-	ENSDARG00000015657	zgc:77112	1.4965803	1.4867612	0.00014079
160-	ENSDARG00000007693	nfkbiab	6.50341207	1.48328355	9.53E-05
161-	ENSDARG00000067751	si:rp71-68n21.9	1.05406985	1.48091217	0.00018759

162-	ENSDARG00000055504	si:ch211-212k18.7	4.30577588	1.45302949	6.09E-05
163-	ENSDARG00000015551	fth1a	9.63578944	1.45100805	0.00011365
164-	ENSDARG00000069461	rnaseka	2.48601379	1.4411342	0.00010512
165-	ENSDARG00000007377	odc1	9.50154469	1.44065313	9.04E-05
166-	ENSDARG00000069175	vps4b	1.2976992	1.44034027	0.00019477
167-	ENSDARG00000039490	pitpnaa	1.23803487	1.4296795	0.000257
168-	ENSDARG00000017320	f11r.1	2.71472706	1.4239659	0.00013072
169-	ENSDARG00000063295	myh9a	1.67060126	1.41920785	0.00021759
170-	ENSDARG00000000906	skap2	3.52019552	1.41862203	0.00010144
171-	ENSDARG00000094673	rhoab	4.60409753	1.41206331	0.00013332
172-	ENSDARG000000103483	litaf	2.62523056	1.40715957	0.00015051
173-	ENSDARG00000063908	mt-co2	50.7644015	1.39545038	0.00011052
174-	ENSDARG00000006468	grap2a	1.15848242	1.39358102	0.00049778
175-	ENSDARG00000006029	lta4h	5.50403452	1.39334956	0.00016618
176-	ENSDARG00000063899	mt-nd2	4.66376186	1.39020486	0.00018329
177-	ENSDARG00000055120	ctsba	1.34741947	1.38067354	0.00155121
178-	ENSDARG00000053979	chmp2a	1.84959426	1.37047422	0.00034876
179-	ENSDARG000000101785	laptm5	4.94219541	1.36711409	0.00023088
180-	ENSDARG00000096331	si:ch73-248e21.7	5.08638421	1.36561361	0.00024306
181-	ENSDARG00000021113	ptmaa	7.62211827	1.36346777	0.00025982
182-	ENSDARG00000035018	thy1	4.8328108	1.35640289	0.00031534
183-	ENSDARG00000009753	sf3b6	1.41702786	1.35618105	0.00048021
184-	ENSDARG00000057867	lasp1	4.72839822	1.34256482	0.00031361
185-	ENSDARG00000063924	mt-cyb	13.2603975	1.34223414	0.00025982
186-	ENSDARG00000027249	btg1	4.72342619	1.34020367	0.00034775
187-	ENSDARG000000103118	cib1	1.25792298	1.3344753	0.00073641
188-	ENSDARG00000052082	gabarapb	3.34617456	1.32893375	0.00033034
189-	ENSDARG00000000767	spi1b	8.52702729	1.31190168	0.00042642
190-	ENSDARG00000086332	bri3	1.07892998	1.30932254	0.00116704
191-	ENSDARG00000089838	si:dkey-262k9.4	2.31199282	1.30602915	0.00059489
192-	ENSDARG000000101849	adipor2	1.441888	1.30378055	0.0009487
193-	ENSDARG00000094396	CR848841.1	1.00434957	1.29958353	0.00153348
194-	ENSDARG00000021647	gnai1	2.58048231	1.29597023	0.00061711
195-	ENSDARG000000103718	chmp5b	1.38719569	1.29190961	0.00110098
196-	ENSDARG00000041878	rab11ba	2.69483894	1.27262331	0.00077381
197-	ENSDARG00000068214	ccni	4.5444332	1.27191535	0.00079754
198-	ENSDARG00000007347	degs1	1.27283906	1.26877801	0.00152447
199-	ENSDARG00000039830	gng5	3.02299276	1.26286531	0.00077693
200-	ENSDARG000000116058	rhoac	1.32255933	1.2598484	0.00143935
201-	ENSDARG000000102482	zgc:165573	2.43132148	1.25899498	0.00097496
202-	ENSDARG00000009313	zgc:63831	2.72467111	1.25361101	0.00096088
203-	ENSDARG00000033466	tagln2	4.91236324	1.24607786	0.00108436

204-	ENSDARG00000100823	selenok	1.08390201	1.24210504	0.00253803
205-	ENSDARG00000112087	ube2l3a	1.05904187	1.22949386	0.0025956
206-	ENSDARG00000039914	gapdhs	11.7837053	1.2294836	0.00111226
207-	ENSDARG00000075963	mhc1uba	6.33933516	1.22919421	0.00140251
208-	ENSDARG00000058366	si:dkey-222f8.3	1.4070838	1.22621707	0.00204363
209-	ENSDARG00000019062	arpc5b	1.58110477	1.21469464	0.00239486
210-	ENSDARG00000018174	gnai2a	1.67557329	1.20866331	0.00250226
211-	ENSDARG00000025522	sgk1	1.74518168	1.20563417	0.00289029
212-	ENSDARG00000058128	msna	4.53946117	1.20406113	0.0018836
213-	ENSDARG00000071203	sptssa	1.03915376	1.20227239	0.00344227
214-	ENSDARG00000017602	ccng2	1.84959426	1.20054922	0.0040168
215-	ENSDARG00000068367	nfkbie	1.38719569	1.19793346	0.003209
216-	ENSDARG00000022456	eno1a	2.66500678	1.19442822	0.00225268
217-	ENSDARG00000088439	gm2a	1.09384607	1.19153725	0.00408352
218-	ENSDARG00000033144	psme2	1.51646841	1.18971681	0.00340747
219-	ENSDARG00000054063	arpc4	3.43567105	1.18603719	0.00191461
220-	ENSDARG00000099624	chmp1b	1.53635652	1.1849944	0.00311006
221-	ENSDARG00000004301	rhogb	1.74020965	1.18294074	0.00318185
222-	ENSDARG00000103735	elovl1b	3.09260115	1.18037038	0.00212183
223-	ENSDARG00000012972	cfl1l	6.10564986	1.17738431	0.00253686
224-	ENSDARG00000101127	map1lc3b	1.15848242	1.17271607	0.00477544
225-	ENSDARG00000041959	cxc4b	30.5879136	1.1726207	0.00221808
226-	ENSDARG00000026865	fam107b	1.4070838	1.1698346	0.00422947
227-	ENSDARG00000020645	slc7a3a	1.43691597	1.1676868	0.00461117
228-	ENSDARG00000097478	qkia	1.21814676	1.16750698	0.00502756
229-	ENSDARG00000013968	psap	1.441888	1.16285543	0.00439119
230-	ENSDARG00000000069	dap	1.72529357	1.16237149	0.00330895
231-	ENSDARG00000054578	arl6ip1	3.06276899	1.16064943	0.00274136
232-	ENSDARG00000026712	rab5c	1.15848242	1.16054136	0.00465858
233-	ENSDARG00000089043	ptpn6	1.59104882	1.16048471	0.00430914
234-	ENSDARG00000058593	sri	6.17525825	1.15058828	0.00329506
235-	ENSDARG00000052438	actr2a	1.11373418	1.15015265	0.00628106
236-	ENSDARG00000039980	ppt1	3.32628645	1.1488454	0.00414379
237-	ENSDARG00000026829	cotl1	14.3293835	1.14258815	0.00302068
238-	ENSDARG00000019881	bsg	1.33250339	1.13751569	0.0052145
239-	ENSDARG00000070822	cnp	2.40148932	1.12944305	0.00409555
240-	ENSDARG00000013979	ndfip1	1.18831459	1.12898899	0.00635697
241-	ENSDARG00000012729	hcls1	1.30267122	1.12851637	0.00597069
242-	ENSDARG00000101169	grap2b	1.59602085	1.127894	0.00524879
243-	ENSDARG00000070161	vamp3	1.02920971	1.12721091	0.00726589
244-	ENSDARG00000012688	EIF1B	3.8582934	1.11788021	0.00390439
245-	ENSDARG00000029353	serpine2	1.31261528	1.11072164	0.0069804

246-	ENSDARG00000063911	mt-atp6	12.3157123	1.10680805	0.00464922
247-	ENSDARG00000075758	ywhabb	2.03355928	1.10458632	0.00552051
248-	ENSDARG00000022315	atp6v1g1	1.9291467	1.10387213	0.00638974
249-	ENSDARG00000102320	mob1a	2.5059019	1.08165465	0.0061621
250-	ENSDARG00000044254	anxa3b	2.25730052	1.0780792	0.00751513
251-	ENSDARG00000101626	arf6a	1.96892292	1.07315576	0.00780143
252-	ENSDARG00000103340	clic1	2.42137743	1.07270446	0.00736577
253-	ENSDARG00000004034	arhgdig	6.36419529	1.06821828	0.00721699
254-	ENSDARG00000055064	prdx5	7.66686652	1.06743294	0.00722665
255-	ENSDARG00000102741	sumo2b	2.74455922	1.06731138	0.00786382
256-	ENSDARG00000077777	tmsb4x	155.828316	1.06029807	0.0068771
257-	ENSDARG00000044573	cdc42	2.35176904	1.05971718	0.00773191
258-	ENSDARG00000020893	slc25a55a	1.27781109	1.05667031	0.01249793
259-	ENSDARG00000026845	rhoaa	1.43691597	1.05656446	0.01081984
260-	ENSDARG00000090454	gnb1a	1.72529357	1.05239865	0.00963738
261-	ENSDARG00000038010	rac2	7.61217421	1.024728	0.01070707
262-	ENSDARG00000099672	capgb	5.73274779	1.01868553	0.0119111
263-	ENSDARG00000068436	h3f3b.1	2.25232849	1.01436341	0.01289305
264-	ENSDARG00000015343	pgd	2.05344739	1.01120262	0.01369742
265-	ENSDARG00000104372	gnb1b	2.35176904	0.99558684	0.0146374
266-	ENSDARG00000040158	cdc42l	2.30204877	0.99448852	0.01490678
267-	ENSDARG00000102632	ubc	8.6911042	0.99369036	0.01468794
268-	ENSDARG00000020929	fam49ba	1.62585302	0.99179962	0.01871876
269-	ENSDARG00000010279	scamp2	1.39216772	0.98850384	0.01928791
270-	ENSDARG00000071384	slc7a11	1.00434957	0.98614417	0.02695107
271-	ENSDARG00000029036	rab32a	1.21814676	0.97776179	0.0232693
272-	ENSDARG00000104782	oaz1b	2.23244038	0.97502124	0.01774407
273-	ENSDARG00000068708	ifrd1	1.04412579	0.96064975	0.0284844
274-	ENSDARG00000043608	EIF4EBP1	1.6407691	0.95576897	0.02696043
275-	ENSDARG00000099690	vaspa	1.14853837	0.95021826	0.02910939
276-	ENSDARG00000022303	higd1a	3.29148225	0.94960464	0.02091581
277-	ENSDARG00000058225	arpc4l	2.70975503	0.93567688	0.02369118
278-	ENSDARG00000036456	anxa4	3.99253814	0.9291555	0.0232693
279-	ENSDARG00000000690	sypl2b	1.01926565	0.92898042	0.03697662
280-	ENSDARG00000045248	h3f3d	6.63765681	0.9275501	0.02691138
281-	ENSDARG00000075768	sdhb	1.06898593	0.91474234	0.03917282
282-	ENSDARG00000069846	zgc:162944	2.51584595	0.91462765	0.03015787
283-	ENSDARG00000002369	si:ch211-202a12.4	2.38160121	0.90082578	0.03343797
284-	ENSDARG00000009505	prelid3b	2.55065015	0.89944825	0.03278284
285-	ENSDARG00000007323	chmp4bb	1.34741947	0.89741781	0.04163467
286-	ENSDARG00000052674	csnk1a1	1.04412579	0.89486634	0.04571591

287-	ENSDARG00000095451	si:ch211-196l7.4	1.84959426	0.87783938	0.04394859
288-	ENSDARG00000104068	gstp1	2.17277605	0.87497307	0.0418767
289-	ENSDARG00000012987	gpia	2.03355928	0.87437054	0.040575
290-	ENSDARG00000101479	BX908782.2	62.6376034	0.87403201	0.04031942
291-	ENSDARG00000043154	ucp2	2.66997881	0.86843948	0.04192688
292-	ENSDARG00000099766	myl12.1	8.36295038	0.86598379	0.04249427
293-	ENSDARG00000005162	tpm3	3.24176198	0.86350824	0.03870404
294-	ENSDARG00000054755	alox5ap	6.50838409	0.8551587	0.04953143
295-	ENSDARG00000074340	serf2	4.0571745	0.83681886	0.04669993

Appendix table 6.3 The list of human orthologues of cluster one-specific zebrafish genes.

No	Zebrafish Cluster 1 genes ID	Zebrafish Cluster 1 genes name	Human homologous genes ID	Human homologous genes name
1.	ENSDARG00000003599	rpl3	ENSG00000100316	RPL3
2.	ENSDARG00000003795	idh2	ENSG00000182054	IDH2
3.	ENSDARG00000005230	ssr2	ENSG00000163479	SSR2
4.	ENSDARG00000005821	ncf2	ENSG00000116701	NCF2
5.	ENSDARG00000005926	ak2	ENSG00000004455	AK2
6.	ENSDARG00000007320	rpl7	ENSG00000147604	RPL7
7.	ENSDARG00000007697	fabp7a	ENSG00000164434	FABP7
8.	ENSDARG00000008155	sms	ENSG00000102172	SMS
9.	ENSDARG00000009285	rpl15	ENSG00000174748	RPL15
10.	ENSDARG00000010244	rpl22l1	ENSG00000163584	RPL22L1
11.	ENSDARG00000014165	ssr3	ENSG00000114850	SSR3
12.	ENSDARG00000014867	rpl8	ENSG00000161016	RPL8
13.	ENSDARG00000015862	rpl5b	ENSG00000122406	RPL5
14.	ENSDARG00000019230	rpl7a	ENSG00000148303	RPL7A
15.	ENSDARG00000019444	ssr4	ENSG00000180879	SSR4
16.	ENSDARG00000019778	rps6	ENSG00000137154	RPS6
17.	ENSDARG00000020197	rpl5a	ENSG00000122406	RPL5
18.	ENSDARG00000021339	cpa5	ENSG00000091704	CPA1
19.	ENSDARG00000021864	rplp1	ENSG00000137818	RPLP1
20.	ENSDARG00000023290	fabp3	ENSG00000121769	FABP3
21.	ENSDARG00000025073	rpl18a	ENSG00000105640	RPL18A
22.	ENSDARG00000026369	dbi	ENSG00000155368	DBI
23.	ENSDARG00000029500	rpl34	ENSG00000109475	RPL34
24.	ENSDARG00000029533	rpl18	ENSG00000063177	RPL18
25.	ENSDARG00000030408	rps26l	ENSG00000197728	RPS26
26.	ENSDARG00000034291	rpl37	ENSG00000145592	RPL37
27.	ENSDARG00000035692	rps3a	ENSG00000145425	RPS3A
28.	ENSDARG00000036629	rps14	ENSG00000164587	RPS14
29.	ENSDARG00000036875	rps12	ENSG00000112306	RPS12
30.	ENSDARG00000036966	hsd3b7	ENSG00000099377	HSD3B7
31.	ENSDARG00000037350	rpl9	ENSG00000163682	RPL9
32.	ENSDARG00000039579	cfp	ENSG00000197766	CFD
33.	ENSDARG00000041182	rpl4	ENSG00000174444	RPL4
34.	ENSDARG00000041619	rack1	ENSG00000204628	RACK1
35.	ENSDARG00000042566	rps7	ENSG00000171863	RPS7
36.	ENSDARG00000042905	rpl10a	ENSG00000198755	RPL10A
37.	ENSDARG00000043453	rps5	ENSG00000083845	RPS5
38.	ENSDARG00000043509	rpl11	ENSG00000142676	RPL11
39.	ENSDARG00000043848	sod1	ENSG00000142168	SOD1
40.	ENSDARG00000044521	eeef1b2	ENSG00000114942	EEF1B2

41.	ENSDARG00000046157	RPS17	ENSG00000182774	RPS17
42.	ENSDARG00000051783	rplp0	ENSG00000089157	RPLP0
43.	ENSDARG00000053457	rpl23	ENSG00000125691	RPL23
44.	ENSDARG00000054155	pcna	ENSG00000132646	PCNA
45.	ENSDARG00000054818	rpl32	ENSG00000144713	RPL32
46.	ENSDARG00000055996	rps8a	ENSG00000142937	RPS8
47.	ENSDARG00000056119	eef1g	ENSG00000254772	EEF1G
48.	ENSDARG00000056200	abcb9	ENSG00000150967	ABCB9
49.	ENSDARG00000056600	papss2b	ENSG00000198682	PAPSS2
50.	ENSDARG00000057556	rpl17	ENSG00000265681	RPL17
51.	ENSDARG00000058451	rpl6	ENSG00000089009	RPL6
52.	ENSDARG00000058734	prdx1	ENSG00000117450	PRDX1
53.	ENSDARG00000062519	abcc13	ENSG00000243064	ABCC13
54.	ENSDARG00000069100	aldh9a1a.1	ENSG00000143149	ALDH9A1
55.	ENSDARG00000070426	chac1	ENSG00000128965	CHAC1
56.	ENSDARG00000076568	sec61b	ENSG00000106803	SEC61B
57.	ENSDARG00000080010	adh5	ENSG00000197894	ADH5
58.	ENSDARG00000090697	eif3ea	ENSG00000104408	EIF3E
59.	ENSDARG00000094845	zmp:000000 1323	ENSG00000166482	MFAP4
60.	ENSDARG00000099022	faua	ENSG00000149806	FAU
61.	ENSDARG00000099380	rpl13	ENSG00000167526	RPL13
62.	ENSDARG00000099572	hmgn2	ENSG00000118418	HMGN3
63.	ENSDARG00000102291	eef1da	ENSG00000104529	EEF1D
64.	ENSDARG00000102640	pdia3	ENSG00000167004	PDIA3
65.	ENSDARG00000104011	rps17	ENSG00000182774	RPS17
66.	ENSDARG00000105116	p4hb	ENSG00000185624	P4HB
67.	ENSDARG00000116491	h2az2b	ENSG00000105968	H2AZ2
68.	ENSDARG00000033170	sult2st1	ENSG00000197165	SULT1A2
69.	ENSDARG00000033170	sult2st1	ENSG00000196502	SULT1A1
70.	ENSDARG00000014690	rps4x	ENSG00000198034	RPS4X
71.	ENSDARG00000034897	rps10	ENSG00000124614	RPS10
72.	ENSDARG00000034897	rps10	ENSG00000270800	RPS10-NUDT3

Appendix table 6.4 The list of human orthologues of cluster two-specific zebrafish genes.

No	Zebrafish Cluster 1 genes ID	Zebrafish Cluster 1 genes name	Human homologues genes ID	Human homologues genes name
1.	ENSDARG00000000767	spi1b	ENSG00000066336	SPI1
2.	ENSDARG00000000906	skap2	ENSG00000005020	SKAP2
3.	ENSDARG00000001452	adam8a	ENSG00000151651	ADAM8
4.	ENSDARG00000004311	ldlrp1a	ENSG00000157978	LDLRAP1
5.	ENSDARG00000006029	lta4h	ENSG00000111144	LTA4H
6.	ENSDARG00000006468	grap2a	ENSG00000100351	GRAP2
7.	ENSDARG00000007098	itm2ba	ENSG00000136156	ITM2B
8.	ENSDARG00000007347	degs1	ENSG00000143753	DEGS1
9.	ENSDARG00000007377	odc1	ENSG00000115758	ODC1
10.	ENSDARG00000007823	atf3	ENSG00000162772	ATF3
11.	ENSDARG00000008363	mcl1b	ENSG00000143384	MCL1
12.	ENSDARG00000008735	atp6ap2	ENSG00000182220	ATP6AP2
13.	ENSDARG00000009753	sf3b6	ENSG00000115128	SF3B6
14.	ENSDARG00000010317	gpr183a	ENSG00000169508	GPR183
15.	ENSDARG00000010384	lpar5a	ENSG00000184574	LPAR5
16.	ENSDARG00000011934	gyg1a	ENSG00000163754	GYG1
17.	ENSDARG00000012513	sdcbp2	ENSG00000125775	SDCBP2
18.	ENSDARG00000012874	snap23.1	ENSG00000092531	SNAP23
19.	ENSDARG00000013598	tnfb	ENSG00000232810	TNF
20.	ENSDARG00000013598	tnfb	ENSG00000226979	LTA
21.	ENSDARG00000013655	tpd52l2b	ENSG00000101150	TPD52L2
22.	ENSDARG00000014348	stk17b	ENSG00000081320	STK17B
23.	ENSDARG00000015551	fth1a	ENSG00000167996	FTH1
24.	ENSDARG00000016939	itgb2	ENSG00000160255	ITGB2
25.	ENSDARG00000017320	f11r.1	ENSG00000158769	F11R
26.	ENSDARG00000018283	cyba	ENSG00000051523	CYBA
27.	ENSDARG00000018569	tnfrsf1a	ENSG00000067182	TNFRSF1A
28.	ENSDARG00000019033	tmem59	ENSG00000116209	TMEM59
29.	ENSDARG00000021059	alas1	ENSG00000023330	ALAS1
30.	ENSDARG00000021647	gnai1	ENSG00000127955	GNAI1
31.	ENSDARG00000021924	hsp70.3	ENSG00000126803	HSPA2
32.	ENSDARG00000024746	hsp90aa1.2	ENSG00000080824	HSP90AA1
33.	ENSDARG00000025254	s100a10b	ENSG00000163221	S100A12
34.	ENSDARG00000025428	socs3a	ENSG00000184557	SOCS3
35.	ENSDARG00000025903	lgals9l1	ENSG00000006659	LGALS14
36.	ENSDARG00000026500	xkr9	ENSG00000221947	XKR9
37.	ENSDARG00000026726	anxa1a	ENSG00000135046	ANXA1
38.	ENSDARG00000027249	btg1	ENSG00000133639	BTG1

39.	ENSDARG00000029688	hsp70.1	ENSG00000126803	HSPA2
40.	ENSDARG00000030012	lrrfip1a	ENSG00000124831	LRRFIP1
41.	ENSDARG00000030440	rsrp1	ENSG00000117616	RSRP1
42.	ENSDARG00000030938	fermt3b	ENSG00000149781	FERMT3
43.	ENSDARG00000032631	ltb4r	ENSG00000213903	LTB4R
44.	ENSDARG00000033614	rasgef1ba	ENSG00000138670	RASGEF1B
45.	ENSDARG00000033735	ncf1	ENSG00000158517	NCF1
46.	ENSDARG00000035018	thy1	ENSG00000154096	THY1
47.	ENSDARG00000035858	cnm2	ENSG00000064666	CNM2
48.	ENSDARG00000036967	smox	ENSG00000088826	SMOX
49.	ENSDARG00000037091	asah1a	ENSG00000104763	ASAH1
50.	ENSDARG00000038068	ddx5	ENSG00000108654	DDX5
51.	ENSDARG00000039269	arg2	ENSG00000081181	ARG2
52.	ENSDARG00000039490	pitpnaa	ENSG00000174238	PITPNA
53.	ENSDARG00000039830	gng5	ENSG00000174021	GNNG5
54.	ENSDARG00000041505	itm2bb	ENSG00000136156	ITM2B
55.	ENSDARG00000041878	rab11ba	ENSG00000185236	RAB11B
56.	ENSDARG00000042816	mmp9	ENSG00000100985	MMP9
57.	ENSDARG00000042824	nfe2l2a	ENSG00000116044	NFE2L2
58.	ENSDARG00000042927	mapre1a	ENSG00000101367	MAPRE1
59.	ENSDARG00000043555	tmem30ab	ENSG00000112697	TMEM30A
60.	ENSDARG00000044039	stx11a	ENSG00000135604	STX11
61.	ENSDARG00000044125	txn	ENSG00000136810	TXN
62.	ENSDARG00000044694	fybb	ENSG00000082074	FYB1
63.	ENSDARG00000044852	wbp2nl	ENSG00000183066	WBP2NL
64.	ENSDARG00000045490	ndufb2	ENSG00000090266	NDUFB2
65.	ENSDARG00000045549	bik	ENSG00000183066	WBP2NL
66.	ENSDARG00000052088	cxcr1	ENSG00000163464	CXCR1
67.	ENSDARG00000053136	b2m	ENSG00000166710	B2M
68.	ENSDARG00000053979	chmp2a	ENSG00000130724	CHMP2A
69.	ENSDARG00000054543	samsn1a	ENSG00000155307	SAMSN1
70.	ENSDARG00000055120	ctsba	ENSG00000164733	CTSB
71.	ENSDARG00000055186	ccr9a	ENSG00000173585	CCR9
72.	ENSDARG00000055504	si:ch211-212k18.7	ENSG00000129226	CD68
73.	ENSDARG00000055723	hsp70l	ENSG00000126803	HSPA2
74.	ENSDARG00000056499	ca6	ENSG00000131686	CA6
75.	ENSDARG00000056532	serinc2	ENSG00000168528	SERINC2
76.	ENSDARG00000056615	cybb	ENSG00000165168	CYBB
77.	ENSDARG00000057853	atp6v0ca	ENSG00000185883	ATP6V0C
78.	ENSDARG00000057867	lasp1	ENSG00000002834	LASP1

79.	ENSDARG00000059049	zgc:174904	ENSG00000152672	CLEC4F
80.	ENSDARG00000062049	arhgap45b	ENSG00000180448	ARHGAP45
81.	ENSDARG00000063295	myh9a	ENSG00000100345	MYH9
82.	ENSDARG00000067916	lyn	ENSG00000254087	LYN
83.	ENSDARG00000068214	ccni	ENSG00000118816	CCNI
84.	ENSDARG00000068233	zgc:64051	ENSG00000143119	CD53
85.	ENSDARG00000068784	vsir	ENSG00000107738	VSIR
86.	ENSDARG00000069175	vps4b	ENSG00000119541	VPS4B
87.	ENSDARG00000069461	rnaseka	ENSG00000219200	RNASEK
88.	ENSDARG00000070792	lrrc15	ENSG00000164342	TLR3
89.	ENSDARG00000071251	ppp1r18	ENSG00000146112	PPP1R18
90.	ENSDARG00000074656	ctss2.1	ENSG00000163131	CTSS
91.	ENSDARG00000074782	adgrg3	ENSG00000182885	ADGRG3
92.	ENSDARG00000074851	s1pr4	ENSG00000125910	S1PR4
93.	ENSDARG00000075261	timp2b	ENSG00000035862	TIMP2
94.	ENSDARG00000075666	tsc22d3	ENSG00000157514	TSC22D3
95.	ENSDARG00000077499	plekho1b	ENSG00000023902	PLEKHO1
96.	ENSDARG00000078547	si:ch211-264f5.2	ENSG00000162897	FCAMR
97.	ENSDARG00000078619	pnp5a	ENSG00000198805	PNP
98.	ENSDARG00000078824	si:ch211-66e2.3	ENSG00000076662	ICAM3
99.	ENSDARG00000086332	bri3	ENSG00000164713	BRI3
100.	ENSDARG00000087152	sowahd	ENSG00000187808	SOWAHD
101.	ENSDARG00000089368	hopx	ENSG00000171476	HOPX
102.	ENSDARG00000089706	si:ch211-276a23.5	ENSG00000172901	LVRN
103.	ENSDARG00000091902	b3gnt2b	ENSG00000170340	B3GNT2
104.	ENSDARG00000092362	hsp70.2	ENSG00000126803	HSPA2
105.	ENSDARG00000093313	ubb	ENSG00000150991	UBC
106.	ENSDARG00000000069	dap	ENSG00000112977	DAP
107.	ENSDARG000000000690	sypl2b	ENSG00000143028	SYPL2
108.	ENSDARG00000002369	UBB	ENSG00000170315	UBB
109.	ENSDARG00000004034	arhgdig	ENSG00000111348	ARHGDIB
110.	ENSDARG00000004301	rhogb	ENSG00000177105	RHOG
111.	ENSDARG00000005162	tpm3	ENSG00000143549	TPM3
112.	ENSDARG00000007323	chmp4bb	ENSG00000101421	CHMP4B
113.	ENSDARG00000009313	napsa	ENSG00000131400	NAPSA
114.	ENSDARG00000009505	prelid3b	ENSG00000101166	PRELID3B
115.	ENSDARG00000010279	scamp2	ENSG00000140497	SCAMP2
116.	ENSDARG00000012688	eif1	ENSG00000173812	EIF1
117.	ENSDARG00000012729	hcls1	ENSG00000180353	HCLS1
118.	ENSDARG00000012987	gpia	ENSG00000105220	GPI

119.	ENSDARG00000013968	psap	ENSG00000197746	PSAP
120.	ENSDARG00000013979	ndfip1	ENSG00000131507	NDFIP1
121.	ENSDARG00000015343	pgd	ENSG00000142657	PGD
122.	ENSDARG00000017602	ccng2	ENSG00000138764	CCNG2
123.	ENSDARG00000018174	gnai2a	ENSG00000114353	GNAI2
124.	ENSDARG00000019062	arpc5b	ENSG00000162704	ARPC5
125.	ENSDARG00000019881	bsg	ENSG00000172270	BSG
126.	ENSDARG00000020645	slc7a3a	ENSG00000165349	SLC7A3
127.	ENSDARG00000022315	atp6v1g1	ENSG00000136888	ATP6V1G1
128.	ENSDARG00000022456	eno1a	ENSG00000074800	ENO1
129.	ENSDARG00000025522	sgk1	ENSG00000118515	SGK1
130.	ENSDARG00000026829	cotl1	ENSG00000103187	COTL1
131.	ENSDARG00000026865	fam107b	ENSG00000065809	FAM107B
132.	ENSDARG00000029036	rab32a	ENSG00000118508	RAB32
133.	ENSDARG00000029353	serpine2	ENSG00000135919	SERPINE2
134.	ENSDARG00000033144	psme2	ENSG00000100911	PSME2
135.	ENSDARG00000033466	tagln2	ENSG00000149591	TAGLN
136.	ENSDARG00000036456	anxa4	ENSG00000196975	ANXA4
137.	ENSDARG00000038010	rac2	ENSG00000128340	RAC2
138.	ENSDARG00000039914	gapdhs	ENSG00000105679	GAPDHS
139.	ENSDARG00000039980	ppt1	ENSG00000131238	PPT1
140.	ENSDARG00000041959	cxcr4b	ENSG00000121966	CXCR4
141.	ENSDARG00000043154	ucp2	ENSG00000175567	UCP2
142.	ENSDARG00000043608	eif4ebp1	ENSG00000187840	EIF4EBP1
143.	ENSDARG00000044254	anxa3b	ENSG00000138772	ANXA3
144.	ENSDARG00000044573	cdc42	ENSG00000070831	CDC42
145.	ENSDARG00000045248	h3f3d	ENSG00000188375	H3-5
146.	ENSDARG00000052438	actr2a	ENSG00000138071	ACTR2
147.	ENSDARG00000052674	csnk1a1	ENSG00000113712	CSNK1A1
148.	ENSDARG00000054063	arpc4	ENSG00000241553	ARPC4
149.	ENSDARG00000054755	alox5ap	ENSG00000132965	ALOX5AP
150.	ENSDARG00000055064	prdx5	ENSG00000126432	PRDX5
151.	ENSDARG00000058128	msna	ENSG00000147065	MSN
152.	ENSDARG00000058593	sri	ENSG00000075142	SRI
153.	ENSDARG00000063911	mt-atp6	ENSG00000198899	MT-ATP6
154.	ENSDARG00000068367	nfkbie	ENSG00000146232	NFKBIE
155.	ENSDARG00000070161	vamp3	ENSG00000049245	VAMP3
156.	ENSDARG00000070822	cnp	ENSG00000173786	CNP
157.	ENSDARG00000071203	sptssa	ENSG00000165389	SPTSSA
158.	ENSDARG00000071384	slc7a11	ENSG00000151012	SLC7A11
159.	ENSDARG00000074340	serf2	ENSG00000140264	SERF2

160.	ENSDARG00000075758	ywhabb	ENSG00000166913	YWHAB
161.	ENSDARG00000075768	sdhb	ENSG00000117118	SDHB
162.	ENSDARG00000077777	tmsb4x	ENSG00000205542	TMSB4X
163.	ENSDARG00000088439	gm2a	ENSG00000196743	GM2A
164.	ENSDARG00000089043	ptpn6	ENSG00000111679	PTPN6
165.	ENSDARG00000090454	gnb1a	ENSG00000078369	GNB1
166.	ENSDARG00000099624	chmp1b	ENSG00000255112	CHMP1B
167.	ENSDARG00000099672	capgb	ENSG00000042493	CAPG
168.	ENSDARG00000099690	vaspa	ENSG00000125753	VASP
169.	ENSDARG00000099766	myl12.1	ENSG00000101335	MYL9
170.	ENSDARG00000100823	selenok	ENSG00000113811	SELENOK
171.	ENSDARG00000101169	grap2b	ENSG00000100351	GRAP2
172.	ENSDARG00000101626	arf6a	ENSG00000165527	ARF6
173.	ENSDARG00000102320	mob1a	ENSG00000114978	MOB1A
174.	ENSDARG00000102632	ubc	ENSG00000150991	UBC
175.	ENSDARG00000102741	sumo2b	ENSG00000188612	SUMO2
176.	ENSDARG00000103340	clic1	ENSG00000213719	CLIC1
177.	ENSDARG00000103735	elovl1b	ENSG00000066322	ELOVL1
178.	ENSDARG00000104068	gstp1	ENSG00000084207	GSTP1
179.	ENSDARG00000104372	gnb1b	ENSG00000078369	GNB1
180.	ENSDARG00000104782	oaz1b	ENSG00000143450	OAZ3
181.	ENSDARG00000068708	ifrd1	ENSG00000006652	IFRD1
182.	ENSDARG00000101127	map1lc3b	ENSG00000258102	MAP1LC3B2
183.	ENSDARG00000101127	map1lc3b	ENSG00000140941	MAP1LC3B
184.	ENSDARG00000086337	si:dkey-102g19.3	ENSG00000234465	PINLYP
185.	ENSDARG00000091996	tcnbb	ENSG00000134827	TCN1
186.	ENSDARG00000022303	higd1a	ENSG00000131097	HIGD1B
187.	ENSDARG00000018146	gpx1a	ENSG00000211445	GPX3
188.	ENSDARG00000007682	ppdpfa	ENSG00000125534	PPDPF
189.	ENSDARG00000034187	calm1b	ENSG00000160014	CALM3
190.	ENSDARG00000036282	rnaset2	ENSG00000026297	RNASET2
191.	ENSDARG00000036282	rnaset2	ENSG00000249141	AL159163.1
192.	ENSDARG00000041394	dnajb1b	ENSG00000172404	DNAJB7
193.	ENSDARG00000041394	dnajb1b	ENSG00000132002	DNAJB1
194.	ENSDARG00000057698	ctsd	ENSG00000117984	CTSD
195.	ENSDARG00000057698	ctsd	ENSG00000250644	AC068580.4
196.	ENSDARG00000067797	spi1a	ENSG00000269404	SPIB
197.	ENSDARG00000022303	higd1a	ENSG00000181061	HIGD1A
198.	ENSDARG00000026712	rab5c	ENSG00000108774	RAB5C
199.	ENSDARG00000026712	rab5c	ENSG00000267261	AC099811.2
200.	ENSDARG00000054578	arl6ip1	ENSG00000170540	ARL6IP1

201.	ENSDARG00000054578	arl6ip1	ENSG00000260342	AC138811.2
202.	ENSDARG00000112087	ube2l3a	ENSG00000185651	UBE2L3
203.	ENSDARG00000090552	si:dkey- 7j14.6	ENSG00000149516	MS4A3
204.	ENSDARG00000090552	si:dkey- 7j14.6	ENSG00000110077	MS4A6A
205.	ENSDARG00000090552	si:dkey- 7j14.6	ENSG00000166927	MS4A7

Appendix table 6.5 The list of the human orthologues of cluster one-specific zebrafish genes that were found within the Dinh et al dataset in SeqMonk.

No	Gene Name	Gene ID
1-	RPL11	ENSG00000142676
2-	FABP3	ENSG00000121769
3-	AK2	ENSG00000004455
4-	RPS8	ENSG00000142937
5-	PRDX1	ENSG00000117450
6-	RPL5	ENSG00000122406
7-	SSR2	ENSG00000163479
8-	ALDH9A1	ENSG00000143149
9-	NCF2	ENSG00000116701
10-	RPS7	ENSG00000171863
11-	DBI	ENSG00000155368
12-	EEF1B2	ENSG00000114942
13-	RPL32	ENSG00000144713
14-	RPL15	ENSG00000174748
15-	SSR3	ENSG00000114850
16-	RPL22L1	ENSG00000163584
17-	RPL9	ENSG00000163682
18-	ADH5	ENSG00000197894
19-	RPL34	ENSG00000109475
20-	RPS3A	ENSG00000145425
21-	RPL37	ENSG00000145592
22-	RPS14	ENSG00000164587
23-	RACK1	ENSG00000204628
24-	RPS10-NUDT3	ENSG00000270800
25-	RPS10	ENSG00000124614
26-	RPL10A	ENSG00000198755
27-	HMGN3	ENSG00000118418
28-	FABP7	ENSG00000164434
29-	RPS12	ENSG00000112306
30-	CPA1	ENSG00000091704
31-	RPL7	ENSG00000147604
32-	EIF3E	ENSG00000104408
33-	EEF1D	ENSG00000104529
34-	RPL8	ENSG00000161016
35-	RPS6	ENSG00000137154
36-	SEC61B	ENSG00000106803
37-	RPL7A	ENSG00000148303
38-	PAPSS2	ENSG00000198682
39-	EEF1G	ENSG00000254772
40-	FAU	ENSG00000149806
41-	RPS26	ENSG00000197728
42-	RPLP0	ENSG00000089157
43-	ABCB9	ENSG00000150967
44-	CHAC1	ENSG00000128965
45-	PDIA3	ENSG00000167004
46-	RPL4	ENSG00000174444
47-	RPLP1	ENSG00000137818
48-	RPS17	ENSG00000182774

49-	IDH2	ENSG00000182054
50-	SULT1A2	ENSG00000197165
51-	SULT1A1	ENSG00000196502
52-	HSD3B7	ENSG00000099377
53-	RPL13	ENSG00000167526
54-	MFAP4	ENSG00000166482
55-	RPL23	ENSG00000125691
56-	P4HB	ENSG00000185624
57-	RPL17	ENSG00000265681
58-	CFD	ENSG00000197766
59-	RPL18A	ENSG00000105640
60-	RPL18	ENSG00000063177
61-	RPS5	ENSG00000083845
62-	PCNA	ENSG00000132646
63-	SOD1	ENSG00000142168
64-	RPL3	ENSG00000100316
65-	SMS	ENSG00000102172
66-	RPS4X	ENSG00000198034
67-	SSR4	ENSG00000180879

Appendix table 6.6 List of human orthologues of cluster one-specific zebrafish genes that were significantly upregulated in human ePreNs relative to Neuts in the Dinh et al. dataset.

No	Gene Name	Gene ID	Log2 Fold Change (<i>qmc85-CL1</i> on DESeq $p < 0.05$)
1-	NCF2	ENSG00000116701	4.160583973
2-	SULT1A2	ENSG00000197165	3.584739447
3-	SULT1A1	ENSG00000196502	2.923308611
4-	RPS10- NUDT3	ENSG00000270800	-1.803654075
5-	RPS10	ENSG00000124614	-2.145505905
6-	SOD1	ENSG00000142168	-2.365692377
7-	PDIA3	ENSG00000167004	-2.509667635
8-	RPL9	ENSG00000163682	-2.548565626
9-	PCNA	ENSG00000132646	-2.602244377
10-	PRDX1	ENSG00000117450	-2.668878794
11-	SSR3	ENSG00000114850	-2.800568819
12-	RPL7	ENSG00000147604	-2.818050146
13-	RPL11	ENSG00000142676	-2.835545301
14-	RPL7A	ENSG00000148303	-3.077988625
15-	RPS14	ENSG00000164587	-3.104066661
16-	AK2	ENSG00000004455	-3.108836174
17-	RPL18	ENSG00000063177	-3.23916769
18-	RPL32	ENSG00000144713	-3.289279222
19-	RPL23	ENSG00000125691	-3.291668892
20-	RPL15	ENSG00000174748	-3.330591202
21-	RACK1	ENSG00000204628	-3.398004055
22-	RPS3A	ENSG00000145425	-3.495939493
23-	RPL8	ENSG00000161016	-3.496892452
24-	RPS6	ENSG00000137154	-3.563730955
25-	EIF3E	ENSG00000104408	-3.887967587
26-	RPL3	ENSG00000100316	-4.032348156
27-	RPS5	ENSG00000083845	-4.100712776
28-	RPS8	ENSG00000142937	-4.546203136
29-	EEF1G	ENSG00000254772	-4.554638863
30-	SSR4	ENSG00000180879	-4.604718685
31-	IDH2	ENSG00000182054	-4.669121265
32-	P4HB	ENSG00000185624	-4.732108116
33-	RPL10A	ENSG00000198755	-4.835008144
34-	RPS4X	ENSG00000198034	-4.844734192
35-	RPLP1	ENSG00000137818	-4.871033192
36-	EEF1B2	ENSG00000114942	-4.879534721
37-	RPL4	ENSG00000174444	-4.892151356
38-	RPL18A	ENSG00000105640	-4.926548958
39-	RPS12	ENSG00000112306	-5.037570953
40-	RPL13	ENSG00000167526	-5.282867908
41-	RPL5	ENSG00000122406	-5.833559513
42-	RPLP0	ENSG00000089157	-6.536200523
43-	HMGN3	ENSG00000118418	-6.678925514

Appendix table 6.7 The list of the human orthologues of cluster two-specific zebrafish genes that were found within the Dinh et al dataset in SeqMonk.

No	Gene Name	Gene ID
1-	FTH1	ENSG00000167996
2-	S100A12	ENSG00000163221
3-	B2M	ENSG00000166710
4-	MMP9	ENSG00000100985
5-	VSIR	ENSG00000107738
6-	ARHGDIB	ENSG00000111348
7-	CTSD	ENSG00000117984
8-	ITGB2	ENSG00000160255
9-	EIF1	ENSG00000173812
10-	COTL1	ENSG00000103187
11-	CYBA	ENSG00000051523
12-	CNN2	ENSG00000064666
13-	PSAP	ENSG00000197746
14-	ALOX5AP	ENSG00000132965
15-	ICAM3	ENSG00000076662
16-	RAC2	ENSG00000128340
17-	AC068580.4	ENSG00000250644
18-	BTG1	ENSG00000133639
19-	ADGRG3	ENSG00000182885
20-	RNASET2	ENSG00000026297
21-	GNAI2	ENSG00000114353
22-	HCLS1	ENSG00000180353
23-	SPI1	ENSG00000066336
24-	TSC22D3	ENSG00000157514
25-	ITM2B	ENSG00000136156
26-	RHOG	ENSG00000177105
27-	TPM3	ENSG00000143549
28-	NCF1	ENSG00000158517
29-	PGD	ENSG00000142657
30-	CTSS	ENSG00000163131
31-	TMSB4X	ENSG00000205542
32-	UBC	ENSG00000150991
33-	MAP1LC3B	ENSG00000140941
34-	ADAM8	ENSG00000151651
35-	ATP6V0C	ENSG00000185883
36-	MCL1	ENSG00000143384
37-	PPT1	ENSG00000131238
38-	RNASEK	ENSG00000219200

39-	AL159163.1	ENSG00000249141
40-	ASAH1	ENSG00000104763
41-	DDX5	ENSG00000108654
42-	RAB5C	ENSG00000108774
43-	LYN	ENSG00000254087
44-	MYH9	ENSG00000100345
45-	VASP	ENSG00000125753
46-	CHMP2A	ENSG00000130724
47-	CXCR4	ENSG00000121966
48-	ARPC4	ENSG00000241553
49-	CAPG	ENSG00000042493
50-	PTPN6	ENSG00000111679
51-	SERF2	ENSG00000140264
52-	LRRFIP1	ENSG00000124831
53-	TXN	ENSG00000136810
54-	BSG	ENSG00000172270
55-	ARPC5	ENSG00000162704
56-	IFRD1	ENSG00000006652
57-	YWHAB	ENSG00000166913
58-	MSN	ENSG00000147065
59-	STK17B	ENSG00000081320
60-	CCNI	ENSG00000118816
61-	AC099811.2	ENSG00000267261
62-	MT-ATP6	ENSG00000198899
63-	CDC42	ENSG00000070831
64-	SOCS3	ENSG00000184557
65-	CALM3	ENSG00000160014
66-	S1PR4	ENSG00000125910
67-	PRDX5	ENSG00000126432
68-	FYB1	ENSG00000082074
69-	CD53	ENSG00000143119
70-	MAPRE1	ENSG00000101367
71-	UBB	ENSG00000170315
72-	ENO1	ENSG00000074800
73-	RAB11B	ENSG00000185236
74-	CXCR1	ENSG00000163464
75-	LASP1	ENSG00000002834
76-	TPD52L2	ENSG00000101150
77-	DNAJB1	ENSG00000132002
78-	FERMT3	ENSG00000149781
79-	SAMSN1	ENSG00000155307
80-	STX11	ENSG00000135604

81-	TMEM59	ENSG00000116209
82-	ATP6V1G1	ENSG00000136888
83-	SELENOK	ENSG00000113811
84-	ARF6	ENSG00000165527
85-	ATP6AP2	ENSG00000182220
86-	RAB32	ENSG00000118508
87-	GYG1	ENSG00000163754
88-	BRI3	ENSG00000164713
89-	ELOVL1	ENSG00000066322
90-	CSNK1A1	ENSG00000113712
91-	ANXA1	ENSG00000135046
92-	CCNG2	ENSG00000138764
93-	ACTR2	ENSG00000138071
94-	GNG5	ENSG00000174021
95-	SKAP2	ENSG00000005020
96-	SCAMP2	ENSG00000140497
97-	TNFRSF1A	ENSG00000067182
98-	DAP	ENSG00000112977
99-	SDHB	ENSG00000117118
100-	UCP2	ENSG00000175567
101-	VPS4B	ENSG00000119541
102-	GPX3	ENSG00000211445
103-	VAMP3	ENSG00000049245
104-	ANXA3	ENSG00000138772
105-	CD68	ENSG00000129226
106-	HSP90AA1	ENSG00000080824
107-	SF3B6	ENSG00000115128
108-	MOB1A	ENSG00000114978
109-	CYBB	ENSG00000165168
110-	LTA4H	ENSG00000111144
111-	PLEKHO1	ENSG00000023902
112-	F11R	ENSG00000158769
113-	SNAP23	ENSG00000092531
114-	RSRP1	ENSG00000117616
115-	ODC1	ENSG00000115758
116-	FAM107B	ENSG00000065809
117-	GM2A	ENSG00000196743
118-	NFKBIE	ENSG00000146232
119-	CHMP1B	ENSG00000255112
120-	UBE2L3	ENSG00000185651
121-	GPI	ENSG00000105220
122-	AC138811.2	ENSG00000260342

123-	NDFIP1	ENSG00000131507
124-	NFE2L2	ENSG00000116044
125-	HIGD1A	ENSG00000181061
126-	PITPNA	ENSG00000174238
127-	B3GNT2	ENSG00000170340
128-	DEGS1	ENSG00000143753
129-	GNB1	ENSG00000078369
130-	ARL6IP1	ENSG00000170540
131-	SGK1	ENSG00000118515
132-	TCN1	ENSG00000134827
133-	SRI	ENSG00000075142
134-	TIMP2	ENSG00000035862
135-	SUMO2	ENSG00000188612
136-	GSTP1	ENSG00000084207
137-	CNP	ENSG00000173786
138-	MAP1LC3B2	ENSG00000258102
139-	MS4A6A	ENSG00000110077
140-	TMEM30A	ENSG00000112697
141-	SERINC2	ENSG00000168528
142-	PRELID3B	ENSG00000101166
143-	CHMP4B	ENSG00000101421
144-	PSME2	ENSG00000100911
145-	EIF4EBP1	ENSG00000187840
146-	PPDPF	ENSG00000125534
147-	ARG2	ENSG00000081181
148-	CTSB	ENSG00000164733
149-	ALAS1	ENSG00000023330
150-	ANXA4	ENSG00000196975
151-	NDUFB2	ENSG00000090266
152-	SOWAHD	ENSG00000187808
153-	SPIB	ENSG00000269404
154-	HSPA2	ENSG00000126803
155-	SDCBP2	ENSG00000125775
156-	HOPX	ENSG00000171476
157-	TAGLN	ENSG00000149591
158-	SMOX	ENSG00000088826
159-	PNP	ENSG00000198805
160-	SLC7A11	ENSG00000151012
161-	DNAJB7	ENSG00000172404
162-	GPR183	ENSG00000169508
163-	SPTSSA	ENSG00000165389
164-	MYL9	ENSG00000101335

165-	GNAI1	ENSG00000127955
166-	RASGEF1B	ENSG00000138670
167-	NAPSA	ENSG00000131400
168-	CLIC1	ENSG00000213719
169-	OAZ3	ENSG00000143450
170-	LDLRAP1	ENSG00000157978
171-	GRAP2	ENSG00000100351
172-	MS4A3	ENSG00000149516
173-	SYPL2	ENSG00000143028
174-	ATF3	ENSG00000162772
175-	SERPINE2	ENSG00000135919
176-	LVRN	ENSG00000172901
177-	LGALS14	ENSG00000006659
178-	LPAR5	ENSG00000184574
179-	FCAMR	ENSG00000162897
180-	THY1	ENSG00000154096
181-	WBP2NL	ENSG00000183066
182-	SLC7A3	ENSG00000165349
183-	HIGD1B	ENSG00000131097
184-	PINLYP	ENSG00000234465
185-	XKR9	ENSG00000221947
186-	MS4A7	ENSG00000166927
187-	GAPDHS	ENSG00000105679
188-	CA6	ENSG00000131686
189-	CCR9	ENSG00000173585
190-	LTB4R	ENSG00000213903
191-	CLEC4F	ENSG00000152672
192-	TLR3	ENSG00000164342
193-	PPP1R18	ENSG00000146112
194-	LTA	ENSG00000226979
195-	TNF	ENSG00000232810

Appendix table 6.8 List of human orthologues of cluster two-specific zebrafish genes that were significantly upregulated in human Neuts relative to ePreNs in the Dinh et al. dataset.

No	Gene Name	Gene ID	Log2 Fold Change (<i>qmc85</i> -CL2 on DESeq $p < 0.05$)
1.	CXCR1	ENSG00000163464	11.6435661
2.	MMP9	ENSG00000100985	9.15347195
3.	NCF1	ENSG00000158517	8.31669617
4.	ADAM8	ENSG00000151651	8.30516911
5.	FTH1	ENSG00000167996	7.83515692
6.	S100A12	ENSG00000163221	7.52459621
7.	CD53	ENSG00000143119	7.07229948
8.	SOCS3	ENSG00000184557	6.30383968
9.	CTSS	ENSG00000163131	6.27276659
10.	CYBB	ENSG00000165168	5.71180725
11.	IFRD1	ENSG00000006652	5.70418787
12.	SGK1	ENSG00000118515	5.48315382
13.	ITGB2	ENSG00000160255	5.30169725
14.	CCNG2	ENSG00000138764	5.0957799
15.	PLEKHO1	ENSG00000023902	5.0955267
16.	HCLS1	ENSG00000180353	5.02369213
17.	CXCR4	ENSG00000121966	4.84272289
18.	SAMSN1	ENSG00000155307	4.83191538
19.	VSIR	ENSG00000107738	4.76494741
20.	STK17B	ENSG00000081320	4.47524977
21.	ITM2B	ENSG00000136156	4.28421259
22.	LYN	ENSG00000254087	3.93346882
23.	COTL1	ENSG00000103187	3.90150738
24.	MCL1	ENSG00000143384	3.79999948
25.	MAP1LC3B2	ENSG00000258102	3.77855134
26.	ALOX5AP	ENSG00000132965	3.45562172
27.	GPX3	ENSG00000211445	3.41144419
28.	ICAM3	ENSG00000076662	3.32475519
29.	SKAP2	ENSG00000005020	3.2325685
30.	SELENOK	ENSG00000113811	3.22844458
31.	ADGRG3	ENSG00000182885	3.22148561
32.	MAP1LC3B	ENSG00000140941	3.21004939
33.	FYB1	ENSG00000082074	3.16017032
34.	CHMP1B	ENSG00000255112	3.1503768
35.	BTG1	ENSG00000133639	3.04834247
36.	TCN1	ENSG00000134827	3.01011825
37.	S1PR4	ENSG00000125910	2.95764041
38.	CNN2	ENSG00000064666	2.80556202
39.	VASP	ENSG00000125753	2.78748322
40.	VPS4B	ENSG00000119541	2.73154044
41.	MYH9	ENSG00000100345	2.66449857
42.	CHMP4B	ENSG00000101421	2.65666389

43.	PSAP	ENSG00000197746	2.62612081
44.	VAMP3	ENSG00000049245	2.62347651
45.	CHMP2A	ENSG00000130724	2.61271286
46.	SNAP23	ENSG00000092531	2.59181261
47.	ASAH1	ENSG00000104763	2.55898738
48.	ARPC5	ENSG00000162704	2.55402398
49.	RAC2	ENSG00000128340	2.47340345
50.	LASP1	ENSG00000002834	2.43020105
51.	TPD52L2	ENSG00000101150	2.3720932
52.	MSN	ENSG00000147065	2.30453134
53.	B3GNT2	ENSG00000170340	2.29135203
54.	STX11	ENSG00000135604	2.14012074
55.	RHOG	ENSG00000177105	2.09859133
56.	SPI1	ENSG00000066336	1.94632483
57.	NFE2L2	ENSG00000116044	1.93107986
58.	PITPNA	ENSG00000174238	1.8041141
59.	RAB32	ENSG00000118508	-1.7882421
60.	CD68	ENSG00000129226	-1.8551776
61.	UBB	ENSG00000170315	-2.0249305
62.	SERF2	ENSG00000140264	-2.240526
63.	ATF3	ENSG00000162772	-2.6856265
64.	ANXA1	ENSG00000135046	-2.7538369
65.	PNP	ENSG00000198805	-2.9758272
66.	MT-ATP6	ENSG00000198899	-3.0278051
67.	ENO1	ENSG00000074800	-3.1827567
68.	PINLYP	ENSG00000234465	-3.5205438
69.	HIGD1B	ENSG00000131097	-3.8843133
70.	GPI	ENSG00000105220	-4.2378497
71.	MS4A7	ENSG00000166927	-4.3068972
72.	NDUFB2	ENSG00000090266	-4.3869758
73.	EIF4EBP1	ENSG00000187840	-4.5625925
74.	GSTP1	ENSG00000084207	-7.5476856
75.	MS4A3	ENSG00000149516	-8.6479082

Appendix table 6.9 List of transcription factors (TFs) harbouring conserved binding sites in conserved non-coding element one (CNE1).

Coloured boxes represent the transcription factor binding motifs that were screened against JASPAR and TRANSFAC database (Figure 3.23C). Note: All genes encoding the TFs are expressed in the subset of GFP+ kidney marrow cells of Tg(*c-myb*-zRE1:PF5:*egfp*)^{qmc85} of scRNA-Seq gene expression profile.

TFs	Description	Databases
Pu.1	Purine-rich termed the PU box.1	JASPAR
Gabpa	GA binding protein transcription factor subunit alpha	JASPAR
Hsf1	Heat shock factor 1	JASPAR
Cebpa	CCAAT enhancer-binding protein alpha	TRANSFAC
Zfx	Zinc finger protein X-linked	JASPAR
Tp53	Tumour protein p53	JASPAR
Meis1	myeloid ecotropic viral integration site-1	TRANSFAC
Cebpa	CCAAT enhancer-binding protein alpha	JASPAR
Runx1	Runt-related transcription factor 1	JASPAR
Atf4	Activating transcription factor 4a	TRANSFAC
Cad	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	JASPAR
Pou6f1	POU class 6 homeobox 1	TRANSFAC
Tbp	TATA box binding protein	TRANSFAC
Irf1	Interferon (IFN) regulatory factor 1	JASPAR TRANSFAC
Irf2	Interferon (IFN) regulatory factor 2	JASPAR TRANSFAC
Stat1	Signal transducer and activator of transcription	JASPAR

7 References

- ADAY, A. W., ZHU, L. J., LAKSHMANAN, A., WANG, J. & LAWSON, N. D. 2011. Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. *Dev Biol*, 357, 450-62.
- AHITUV, N., ZHU, Y., VISEL, A., HOLT, A., AFZAL, V., PENNACCHIO, L. A. & RUBIN, E. M. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS biology*, 5.
- AKKERS, R. C., VAN HEERINGEN, S. J., JACOBI, U. G., JANSSEN-MEGENS, E. M., FRANCOIJS, K. J., STUNNENBERG, H. G. & VEENSTRA, G. J. 2009. A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in *Xenopus* embryos. *Dev Cell*, 17, 425-34.
- AKTUNA, A. Y. 2018. Tendency of K562 chronic myeloid leukemia cells towards cell reprogramming. *Turkish Journal of Hematology*, 35, 260.
- ALBERT, B., JOHNSON, A., LEWIS, J., MORGAN, D., RAFF, M., ROBERTS, K., WALTER, P. 2015. *The Cell*, New York, US, Garland Science.
- ALLEN III, R. D., KIM, H. K., SARAFOVA, S. D. & SIU, G. 2001. Negative regulation of CD4 gene expression by a HES-1-c-Myb complex. *Molecular and cellular biology*, 21, 3071.
- ALLISON, L. A. 2007. *Fundamental molecular biology*, Blackwell Pub.
- ALSAYEGH, K., CORTÉS-MEDINA, L. V., RAMOS-MANDUJANO, G., BADRAIQ, H. & LI, M. 2019. Hematopoietic differentiation of human pluripotent stem cells: HOX and GATA transcription factors as master regulators. *Current genomics*, 20, 438-452.
- AMANO, T., SAGAI, T., TANABE, H., MIZUSHINA, Y., NAKAZAWA, H. & SHIROISHI, T. 2009. Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Developmental cell*, 16, 47-57.
- ANDERSSON, L. C., NILSSON, K. & GAHMBERG, C. G. 1979. K562—a human erythroleukemic cell line. *International journal of cancer*, 23, 143-147.
- ANDERSSON, R. & SANDELIN, A. 2019. Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics*, 1-17.
- ANTOSOVA, B., SMOLIKOVA, J., KLIMOVA, L., LACHOVA, J., BENDOVA, M., KOZMIKOVA, I., MACHON, O. & KOZMIK, Z. 2016. The gene regulatory network of lens induction is wired through Meis-dependent shadow enhancers of Pax6. *PLoS genetics*, 12, e1006441.
- ATHANASIADIS, E. I., BOTTHOF, J. G., ANDRES, H., FERREIRA, L., LIO, P. & CVEJIC, A. 2017. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nature communications*, 8, 1-11.
- AZCOITIA, V., ARACIL, M., MARTÍNEZ-A, C. & TORRES, M. 2005. The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Developmental biology*, 280, 307-320.
- BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W. & NOBLE, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, 37, W202-W208.
- BAINTON, D. F., ULLYOT, J. L. & FARQUHAR, M. G. 1971. The development of neutrophilic polymorphonuclear leukocytes in human bone marrow origin and content of azurophil and specific granules. *Journal of Experimental Medicine*, 134, 907-934.

- BALUDA, M. A. & REDDY, E. P. 1994. Anatomy of an integrated avian myeloblastosis provirus: structure and function. *Oncogene*, 9, 2761-2774.
- BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, 129, 823-37.
- BEJERANO, G., PHEASANT, M., MAKUNIN, I., STEPHEN, S., KENT, W. J., MATTICK, J. S. & HAUSSLER, D. 2004. Ultraconserved elements in the human genome. *Science*, 304, 1321-1325.
- BENGTSSEN, M., KLEPPER, K., GUNDERSEN, S., CUERVO, I., DRABLØS, F., HOVIG, E., SANDVE, G. K., GABRIELSEN, O. S. & ESKELAND, R. 2015. c-Myb binding sites in haematopoietic chromatin landscapes. *PLoS One*, 10, e0133280.
- BENNETT, C. M., KANKI, J. P., RHODES, J., LIU, T. X., PAW, B. H., KIERAN, M. W., LANGENAU, D. M., DELAHAYE-BROWN, A., ZON, L. I. & FLEMING, M. D. 2001. Myelopoiesis in the zebrafish, *Danio rerio*. *Blood, The Journal of the American Society of Hematology*, 98, 643-651.
- BERGHMANS, S., JETTE, C., LANGENAU, D., HSU, K., STEWART, R., LOOK, T. & KANKI, J. P. 2005. Making waves in cancer research: new models in the zebrafish. *Biotechniques*, 39, 227.
- BERTRAND, J. Y., CHI, N. C., SANTOSO, B., TENG, S., STAINIER, D. Y. & TRAVER, D. 2010. Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature*, 464, 108-11.
- BERTRAND, J. Y., KIM, A. D., VIOLETTE, E. P., STACHURA, D. L., CISSON, J. L. & TRAVER, D. 2007. Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. *Development*, 134, 4147-56.
- BHAYA, D., DAVISON, M. & BARRANGOU, R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet*, 45, 273-97.
- BIRBRAIR, A. & FRENETTE, P. S. 2016. Niche heterogeneity in the bone marrow. *Annals of the new York Academy of Sciences*, 1370, 82.
- BOISSET, J.-C., VAN CAPPELLEN, W., ANDRIEU-SOLER, C., GALJART, N., DZIERZAK, E. & ROBIN, C. 2010. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*, 464, 116-120.
- BONN, S., ZINZEN, R. P., GIRARDOT, C., GUSTAFSON, E. H., PEREZ-GONZALEZ, A., DELHOMME, N., GHAVI-HELM, Y., WILCZYŃSKI, B., RIDDELL, A. & FURLONG, E. E. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature genetics*, 44, 148-156.
- BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 132, 311-322.
- BROWN, M. & WITWER, C. 2000. Flow cytometry: principles and clinical applications in hematology. *Clin Chem*, 46, 1221-9.
- BROWNLIE, A., HERSEY, C., OATES, A. C., PAW, B. H., FALICK, A. M., WITKOWSKA, H. E., FLINT, J., HIGGS, D., JESSEN, J. & BAHARY, N. 2003. Characterization of embryonic globin genes of the zebrafish. *Developmental biology*, 255, 48-61.
- BRUDNO, M., DO, C. B., COOPER, G. M., KIM, M. F., DAVYDOV, E., GREEN, E. D., SIDOW, A., BATZOGLOU, S. & PROGRAM, N. C. S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome research*, 13, 721-731.
- BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*, 10, 1213.

- BURGER, A., LINDSAY, H., FELKER, A., HESS, C., ANDERS, C., CHIAVACCI, E., ZAUGG, J., WEBER, L. M., CATENA, R. & JINEK, M. 2016. Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development*, 143, 2025-2037.
- CAI, Z., DE BRUIJN, M., MA, X., DORTLAND, B., LUTEIJN, T., DOWNING, J. R. & DZIERZAK, E. 2000. Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity*, 13, 423-431.
- CANTOR, A. B. & ORKIN, S. H. 2002. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene*, 21, 3368.
- CARROLL, K. J. & NORTH, T. E. 2014. Oceans of opportunity: Exploring vertebrate hematopoiesis in zebrafish. *Experimental hematology*, 42, 684-696.
- CHANG, H. H. Y., PANNUNZIO, N. R., ADACHI, N. & LIEBER, M. R. 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol*, 18, 495-506.
- CHANG, N., SUN, C., GAO, L., ZHU, D., XU, X., ZHU, X., XIONG, J.-W. & XI, J. J. 2013. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell research*, 23, 465.
- CHEN, A. T. & ZON, L. I. 2009. Zebrafish blood stem cells. *Journal of cellular biochemistry*, 108, 35-42.
- CHEN, M. J., LI, Y., DE OBALDIA, M. E., YANG, Q., YZAGUIRRE, A. D., YAMADA-INAGAWA, T., VINK, C. S., BHANDOOOLA, A., DZIERZAK, E. & SPECK, N. A. 2011. Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell stem cell*, 9, 541-552.
- CHEN, M. J., YOKOMIZO, T., ZEIGLER, B. M., DZIERZAK, E. & SPECK, N. A. 2009. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*, 457, 887-891.
- CHOO, A., PALLADINETTI, P., PASSIOURA, T., SHEN, S., LOCK, R., SYMONDS, G. & DOLNIKOV, A. 2006. The role of IRF1 and IRF2 transcription factors in leukaemogenesis. *Current gene therapy*, 6, 543-550.
- CIAU-UITZ, A., MONTEIRO, R., KIRMIZITAS, A. & PATIENT, R. 2014. Developmental hematopoiesis: ontogeny, genetic programming and conservation. *Exp Hematol*, 42, 669-83.
- CLARKE, S. L., VANDERMEER, J. E., WENGER, A. M., SCHAAR, B. T., AHITUV, N. & BEJERANO, G. 2012. Human developmental enhancers conserved between deuterostomes and protostomes. *PLoS genetics*, 8.
- CLÉMENT, Y., TORBEY, P., GILARDI-HEBENSTREIT, P. & CROLLIUS, H. R. 2018. Enhancer-gene maps in the human and zebrafish genomes using evolutionary linkage conservation. *Nucleic Acids Research*.
- CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819-23.
- CONSORTIUM, E. P. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *nature*, 447, 799.
- COSTA, G., KOUSKOFF, V. & LACAUD, G. 2012. Origin of blood cells and HSC production in the embryo. *Trends in immunology*, 33, 215-223.
- COWLAND, J. B. & BORREGAARD, N. 2016. Granulopoiesis and granules of human neutrophils. *Immunological reviews*, 273, 11-28.
- CUMANO, A. & GODIN, I. 2007. Ontogeny of the hematopoietic system. *Annu Rev Immunol*, 25, 745-85.
- CVEJIC, A., SERBANOVIC-CANIC, J., STEMPEL, D. L. & OUWEHAND, W. H. 2011. The role of meis1 in primitive and definitive hematopoiesis during zebrafish development. *Haematologica*, 96, 190.
- DAVIDSON, A. J. & ZON, L. I. 2004. The 'definitive'(and 'primitive') guide to zebrafish hematopoiesis. *Oncogene*, 23, 7233-7246.

- DE BRUIJN, M. F., SPECK, N. A., PEETERS, M. C. & DZIERZAK, E. 2000a. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *Embo j*, 19, 2465-74.
- DE BRUIJN, M. F., SPECK, N. A., PEETERS, M. C. & DZIERZAK, E. 2000b. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *The EMBO journal*, 19, 2465-2474.
- DE JONG, J. L. & ZON, L. I. 2005. Use of the zebrafish system to study primitive and definitive hematopoiesis. *Annu Rev Genet*, 39, 481-501.
- DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. *science*, 295, 1306-1311.
- DI, Q., LIN, Q., HUANG, Z., CHI, Y., CHEN, X., ZHANG, W. & ZHANG, Y. 2017. Zebrafish nephrosin helps host defence against Escherichia coli infection. *Open biology*, 7, 170040.
- DINH, H. Q., EGGERT, T., MEYER, M. A., ZHU, Y. P., OLINGY, C. E., LLEWELLYN, R., WU, R. & HEDRICK, C. C. 2020. Coexpression of CD71 and CD117 Identifies an Early Unipotent Neutrophil Progenitor Population in Human Bone Marrow. *Immunity*, 53, 319-334. e6.
- DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485, 376-380.
- DOBZYCKI, T., MAHONY, C. B., KRECSMARIK, M., KOYUNLAR, C., RISPOLI, R., PEULEN-ZINK, J., GUSSINKLO, K., FEDLAOUI, B., DE PATER, E. & PATIENT, R. 2020. Deletion of a conserved Gata2 enhancer impairs haemogenic endothelium programming and adult Zebrafish haematopoiesis. *Communications biology*, 3, 1-14.
- DOLFINI, D., ZAMBELLI, F., PAVESI, G. & MANTOVANI, R. 2009. A perspective of promoter architecture from the CCAAT box. *Cell cycle*, 8, 4127-4137.
- DOOLEY, K. A., DAVIDSON, A. J. & ZON, L. I. 2005. Zebrafish scl functions independently in hematopoietic and endothelial development. *Developmental biology*, 277, 522-536.
- DOOLITTLE, W. F. 2013. Is junk DNA bunk? A critique of ENCODE. *Proceedings of the National Academy of Sciences*, 110, 5294-5300.
- DRISSEN, R., BUZA-VIDAS, N., WOLL, P., THONGJUEA, S., GAMBARDELLA, A., GIUSTACCHINI, A., MANCINI, E., ZRIWIL, A., LUTTEROPP, M. & GROVER, A. 2016. Distinct myeloid progenitor–differentiation pathways identified through single-cell RNA sequencing. *Nature immunology*, 17, 666.
- DUROUX-RICHARD, I., GIOVANNANGELI, C. & APPARAILLY, F. 2017. CRISPR-Cas9: A revolution in genome editing in rheumatic diseases. *Joint Bone Spine*, 1, 1-4.
- DVOŘÁK, M., URBÁNEK, P., BARTYUÈK, P., PAČES, V., VLACH, J., PEČNKA, V., ARNOLD, L., TRÁVNÍČEK, M. & PIMAN, J. 1989. Transcription of the chicken myb proto-oncogene starts within a CpG island. *Nucleic acids research*, 17, 5651-5664.
- DZIERZAK, E. & MEDVINSKY, A. 1995. Mouse embryonic hematopoiesis. *Trends Genet*, 11, 359-66.
- EMA, M., FALON, P., ZHANG, W. J., HIRASHIMA, M., REID, T., STANFORD, W. L., ORKIN, S., CHOI, K. & ROSSANT, J. 2003. Combinatorial effects of Flk1 and Tal1 on vascular and hematopoietic development in the mouse. *Genes & development*, 17, 380-393.
- ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R. & COYNE, M. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473, 43.
- EVRARD, M., KWOK, I. W., CHONG, S. Z., TENG, K. W., BECHT, E., CHEN, J., SIEOW, J. L., PENNY, H. L., CHING, G. C. & DEVI, S. 2018. Developmental analysis of bone marrow neutrophils reveals populations

- specialized in expansion, trafficking, and effector functions. *Immunity*, 48, 364-379. e8.
- FERKOWICZ, M. J., STARR, M., XIE, X., LI, W., JOHNSON, S. A., SHELLEY, W. C., MORRISON, P. R. & YODER, M. C. 2003. CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. *Development*, 130, 4393-4403.
- FRAME, J. M., MCGRATH, K. E. & PALIS, J. 2013. Erythro-myeloid progenitors: "definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. *Blood Cells, Molecules, and Diseases*, 51, 220-225.
- FRANKE, M., DE LA CALLE-MUSTIENES, E., NETO, A., ACEMEL, R. D., TENA, J. J., SANTOS-PEREIRA, J. M. & GOMEZ-SKARMETA, J. L. 2020. CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression. *bioRxiv*.
- FRASER, S. T. & BARON, M. H. 2009. Embryonic fates for extraembryonic lineages: New perspectives. *J Cell Biochem*, 107, 586-91.
- FUJIWARA, Y., BROWNE, C. P., CUNNIFF, K., GOFF, S. C. & ORKIN, S. H. 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proceedings of the National Academy of Sciences*, 93, 12355-12358.
- FURZE, R. C. & RANKIN, S. M. 2008. Neutrophil mobilization and clearance in the bone marrow. *Immunology*, 125, 281-288.
- GAINES, P., TIEN, C. W., OLINS, A. L., OLINS, D. E., SHULTZ, L. D., CARNEY, L. & BERLINER, N. 2008. Mouse neutrophils lacking lamin B-receptor expression exhibit aberrant development and lack critical functional responses. *Experimental hematology*, 36, 965-976.
- GALLOWAY, J. L., WINGERT, R. A., THISSE, C., THISSE, B. & ZON, L. I. 2005. Loss of gata1 but not gata2 converts erythropoiesis to myelopoiesis in zebrafish embryos. *Developmental cell*, 8, 109-116.
- GERING, M. & 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., GOTTGENS, B., PATIENT, R. K. & GREEN, A. R. 1998. The SCL gene specifies haemangioblast development from early mesoderm. *Embo j*, 17, 4029-45.
- GERLACH, G. F., SCHRADER, L. N. & WINGERT, R. A. 2011. Dissection of the adult zebrafish kidney. *J Vis Exp*.
- GODBAY, W. T. 2014. *An Introduction to Biotechnology : The Science, Technology and Medical Applications*, Oxford, UK, Elsevier Science & Technology.
- GÓMEZ-MARÍN, C., TENA, J. J., ACEMEL, R. D., LÓPEZ-MAYORGA, M., NARANJO, S., DE LA CALLE-MUSTIENES, E., MAESO, I., BECCARI, L., ANEAS, I. & VIELMAS, E. 2015. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proceedings of the National Academy of Sciences*, 112, 7542-7547.
- GONDA, T. J. & METCALF, D. 1984. Expression of myb, myc and fos proto-oncogenes during the differentiation of a murine myeloid leukaemia. *Nature*, 310, 249-251.
- GORDON-KEYLOCK, S., SOBIESIAK, M., RYBTSOV, S., MOORE, K. & MEDVINSKY, A. 2013. Mouse extraembryonic arterial vessels harbor precursors capable of maturing into definitive HSCs. *Blood, The Journal of the American Society of Hematology*, 122, 2338-2345.
- GORDON, S., TODD, J. & COHN, Z. 1974. In vitro synthesis and secretion of lysozyme by mononuclear phagocytes. *The Journal of experimental medicine*, 139, 1228-1248.
- GORE, A. V., MONZO, K., CHA, Y. R., PAN, W. & WEINSTEIN, B. M. 2012. Vascular development in the zebrafish. *Cold Spring Harbor perspectives in medicine*, 2, a006684.

- GORE, A. V., PILLAY, L. M., VENERO GALANTERNIK, M. & WEINSTEIN, B. M. 2018. The zebrafish: A fantastic model for hematopoietic development and disease. *Wiley Interdisciplinary Reviews: Developmental Biology*, 7, e312.
- GRABHER, C., PAYNE, E. M., JOHNSTON, A. B., BOLLI, N., LECHMAN, E., DICK, J. E., KANKI, J. P. & LOOK, A. T. 2011. Zebrafish microRNA-126 determines hematopoietic cell fate through c-Myb. *Leukemia*, 25, 506-514.
- GRASSI, L., POURFARZAD, F., ULLRICH, S., MERKEL, A., WERE, F., CARRILLO-DE-SANTA-PAU, E., YI, G., HIEMSTRA, I. H., TOOL, A. T. & MUL, E. 2018. Dynamics of transcription regulation in human bone marrow myeloid differentiation to mature blood neutrophils. *Cell reports*, 24, 2784-2794.
- GREIG, K. T., CAROTTA, S. & NUTT, S. L. 2008. Critical roles for c-Myb in hematopoietic progenitor cells. *Semin Immunol*, 20, 247-56.
- GRITZ, E. & HIRSCHI, K. K. 2016. Specification and function of hemogenic endothelium during embryogenesis. *Cellular and Molecular Life Sciences*, 73, 1547-1567.
- GUPTA, R. M. & MUSUNURU, K. 2014. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest*, 124, 4154-61.
- HADLAND, B. K., HUPPERT, S. S., KANUNGO, J., XUE, Y., JIANG, R., GRIDLEY, T., CONLON, R. A., CHENG, A. M., KOPAN, R. & LONGMORE, G. D. 2004. A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. *Blood*, 104, 3097-3105.
- HALL, C., FLORES, M. V., STORM, T., CROSIER, K. & CROSIER, P. 2007. The zebrafish lysozyme C promoter drives myeloid-specific expression in transgenic fish. *BMC developmental biology*, 7, 42.
- HESS, I., IWANAMI, N., SCHORPP, M. & BOEHM, T. 2013. Zebrafish model for allogeneic hematopoietic cell transplantation not requiring preconditioning. *Proceedings of the National Academy of Sciences*, 110, 4327-4332.
- HILLER, M., AGARWAL, S., NOTWELL, J. H., PARIKH, R., GUTURU, H., WENGER, A. M. & BEJERANO, G. 2013. Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: application to zebrafish. *Nucleic acids research*, 41, e151-e151.
- HOFFMANN, K., SPERLING, K., OLINS, A. L. & OLINS, D. E. 2007. The granulocyte nucleus and lamin B receptor: avoiding the ovoid. *Chromosoma*, 116, 227-235.
- HOLMES, W. E., LEE, J., KUANG, W.-J., RICE, G. C. & WOOD, W. I. 1991. Structure and functional expression of a human interleukin-8 receptor. *Science*, 253, 1278-1280.
- HOWE, K., CLARK, M. D., TORROJA, C. F., TORRANCE, J., BERTHELOT, C., MUFFATO, M., COLLINS, J. E., HUMPHRAY, S., MCLAREN, K. & MATTHEWS, L. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496, 498-503.
- HSU, J.-C. 2010. *Transcriptional regulation of the stem cell/progenitor c-myb gene*. University of Nottingham.
- HUANG, B., QI, Z. T., XU, Z. & NIE, P. 2010. Global characterization of interferon regulatory factor (IRF) genes in vertebrates: glimpse of the diversification in evolution. *BMC immunology*, 11, 1-28.
- HUANG, Z., CHEN, K., CHI, Y., JIN, H., LI, L., ZHANG, W., XU, J. & ZHANG, Y. 2021. Runx1 regulates zebrafish neutrophil maturation via synergistic interaction with c-Myb. *Journal of Biological Chemistry*, 296, 100272.
- HUBER, T. L., KOUSKOFF, V., FEHLING, H. J., PALIS, J. & KELLER, G. 2004. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature*, 432, 625.
- HUGO, H., CURES, A., SURAWEERA, N., DRABSCH, Y., PURCELL, D., MANTAMADIOTIS, T., PHILLIPS, W., DOBROVIC, A., ZUPI, G. & GONDA, T. J. 2006. Mutations in the MYB intron I regulatory sequence increase

- transcription in colon cancers. *Genes, Chromosomes and Cancer*, 45, 1143-1154.
- HUO, X., LI, H., LI, Z., YAN, C., AGRAWAL, I., MATHAVAN, S., LIU, J. & GONG, Z. 2019. Transcriptomic profiles of tumor-associated neutrophils reveal prominent roles in enhancing angiogenesis in liver tumorigenesis in zebrafish. *Scientific reports*, 9, 1-11.
- HWANG, W. Y., FU, Y., REYON, D., GONZALES, A. P., JOUNG, J. K. & YEH, J.-R. J. 2015. Targeted mutagenesis in zebrafish using CRISPR RNA-guided nucleases. *CRISPR*. Springer.
- HWANG, W. Y., FU, Y., REYON, D., MAEDER, M. L., TSAI, S. Q., SANDER, J. D., PETERSON, R. T., YEH, J. J. & JOUNG, J. K. 2013a. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology*, 31, 227-229.
- HWANG, W. Y., FU, Y., REYON, D., MAEDER, M. L., TSAI, S. Q., SANDER, J. D., PETERSON, R. T., YEH, J. R. & JOUNG, J. K. 2013b. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*, 31, 227-9.
- IRION, U., KRAUSS, J. & NUSSLEIN-VOLHARD, C. 2014. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development*, 141, 4827-30.
- ITOH, M., KIM, C.-H., PALARDY, G., ODA, T., JIANG, Y.-J., MAUST, D., YEO, S.-Y., LORICK, K., WRIGHT, G. J. & ARIZA-MCNAUGHTON, L. 2003. Mind bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by Delta. *Developmental cell*, 4, 67-82.
- IWASAKI, H. & AKASHI, K. 2007. Myeloid lineage commitment from the hematopoietic stem cell. *Immunity*, 26, 726-740.
- JAFFREDO, T., GAUTIER, R., EICHMANN, A. & DIETERLEN-LIÈVRE, F. 1998. Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development*, 125, 4575-4583.
- JAGANNATHAN-BOGDAN, M. & ZON, L. I. 2013. Hematopoiesis. *Development*, 140, 2463-2467.
- JAO, L. E., WENTE, S. R. & CHEN, W. 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A*, 110, 13904-9.
- JEEVANI, T. 2011. Stem cell transplantation-types, risks and benefits. *J Stem Cell Res Ther*, 1, 114.
- JENNE, C. N., LIAO, S. & SINGH, B. 2018. Neutrophils: multitasking first responders of immunity and tissue homeostasis. *Cell and Tissue Research*, 371, 395-397.
- JIN, H., HUANG, Z., CHI, Y., WU, M., ZHOU, R., ZHAO, L., XU, J., ZHEN, F., LAN, Y. & LI, L. 2016. c-Myb acts in parallel and cooperatively with Cebp1 to regulate neutrophil maturation in zebrafish. *Blood*, 128, 415-426.
- JIN, H., LI, L., XU, J., ZHEN, F., ZHU, L., LIU, P. P., ZHANG, M., ZHANG, W. & WEN, Z. 2012. Runx1 regulates embryonic myeloid fate choice in zebrafish through a negative feedback loop inhibiting Pu. 1 expression. *Blood, The Journal of the American Society of Hematology*, 119, 5239-5249.
- JIN, H., XU, J. & WEN, Z. 2007. Migratory path of definitive hematopoietic stem/progenitor cells during zebrafish development. *Blood*, 109, 5208-14.
- JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-21.
- JOHNSON, D. S., MORTAZAVI, A., MYERS, R. M. & WOLD, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316, 1497-1502.

- KAAIJ, L. J., VAN DER WEIDE, R. H., KETTING, R. F. & DE WIT, E. 2018. Systemic loss and gain of chromatin architecture throughout zebrafish development. *Cell reports*, 24, 1-10. e4.
- KALEV-ZYLINSKA, M. L., HORSFIELD, J. A., FLORES, M. V. C., POSTLETHWAIT, J. H., VITAS, M. R., BAAS, A. M., CROSIER, P. S. & 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, 2015-2030.
- KARSUNKY, H., ZENG, H., SCHMIDT, T., ZEVIK, B., KLUGE, R., SCHMID, K. W., DÜHRSEN, U. & MÖRÖY, T. 2002. Inflammatory reactions and severe neutropenia in mice lacking the transcriptional repressor Gfi1. *Nature genetics*, 30, 295-300.
- KIMMEL, C. B., BALLARD, W. W., KIMMEL, S. R., ULLMANN, B. & SCHILLING, T. F. 1995. Stages of embryonic development of the zebrafish. *Developmental dynamics*, 203, 253-310.
- KINGSLEY, P. D., MALIK, J., FANTAUZZO, K. A. & PALIS, J. 2004. Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis. *Blood*, 104, 19-25.
- KISHI, H., JIN, Z.-X., WEI, X.-C., NAGATA, T., MATSUDA, T., SAITO, S. & MURAGUCHI, A. 2002. Cooperative binding of c-Myb and Pax-5 activates the RAG-2 promoter in immature B cells. *Blood, The Journal of the American Society of Hematology*, 99, 576-583.
- KISSA, K. & HERBOMEL, P. 2010. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*, 464, 112-115.
- KISSA, K., MURAYAMA, E., ZAPATA, A., CORTÉS, A., PERRET, E., MACHU, C. & HERBOMEL, P. 2008. Live imaging of emerging hematopoietic stem cells and early thymus colonization. *Blood, The Journal of the American Society of Hematology*, 111, 1147-1156.
- KITAGUCHI, T., KAWAKAMI, K. & KAWAHARA, A. 2009. Transcriptional regulation of a myeloid-lineage specific gene lysozyme C during zebrafish myelopoiesis. *Mechanisms of development*, 126, 314-323.
- KRIVEGA, I. & DEAN, A. 2012. Enhancer and promoter interactions—long distance calls. *Current opinion in genetics & development*, 22, 79-85.
- KULKEAW, K. & SUGIYAMA, D. 2012. Zebrafish erythropoiesis and the utility of fish as models of anemia. *Stem cell research & therapy*, 3, 1-11.
- KUMAR, S., XU, J., PERKINS, C., GUO, F., SNAPPER, S., FINKELMAN, F. D., ZHENG, Y. & FILIPPI, M.-D. 2012. Cdc42 regulates neutrophil migration via crosstalk between WASp, CD11b, and microtubules. *Blood, The Journal of the American Society of Hematology*, 120, 3563-3574.
- LATCHMAN, D. S. 2015 *Gene Control*, New York USA, Garland Science.
- LAUDER, A., CASTELLANOS, A. & WESTON, K. 2001. c-Myb transcription is activated by protein kinase B (PKB) following interleukin 2 stimulation of T cells and is required for PKB-mediated protection from apoptosis. *Molecular and Cellular Biology*, 21, 5797.
- LAURENTI, E. & GÖTTGENS, B. 2018. From haematopoietic stem cells to complex differentiation landscapes. *Nature*, 553, 418-426.
- LE GUYADER, D., REDD, M. J., COLUCCI-GUYON, E., MURAYAMA, E., KISSA, K., BRIOLAT, V., MORDELET, E., ZAPATA, A., SHINOMIYA, H. & HERBOMEL, P. 2008. Origins and unconventional behavior of neutrophils in developing zebrafish. *Blood, The Journal of the American Society of Hematology*, 111, 132-141.
- LENHARD, B., SANDELIN, A. & CARNINCI, P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, 13, 233-245.
- LETTICE, L. A., HEANEY, S. J., PURDIE, L. A., LI, L., DE BEER, P., OOSTRA, B. A., GOODE, D., ELGAR, G., HILL, R. E. & DE GRAAFF, E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin

- and is associated with preaxial polydactyly. *Human molecular genetics*, 12, 1725-1735.
- LETTICE, L. A., HILL, A. E., DEVENNEY, P. S. & HILL, R. E. 2008. Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Human molecular genetics*, 17, 978-985.
- LI, Q., RITTER, D., YANG, N., DONG, Z., LI, H., CHUANG, J. H. & GUO, S. 2010. A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Developmental biology*, 337, 484-495.
- LIAN, J., CHEN, J., WANG, K., ZHAO, L., MENG, P., YANG, L., WEI, J., MA, N., XU, J. & ZHANG, W. 2018. *Alas1* is essential for neutrophil maturation in zebrafish. *haematologica*, 103, 1785.
- LIAO, E. C., PAW, B. H., OATES, A. C., PRATT, S. J., POSTLETHWAIT, J. H. & ZON, L. I. 1998. SCL/Tal-1 transcription factor acts downstream of cloche to specify hematopoietic and vascular progenitors in zebrafish. *Genes & development*, 12, 621-626.
- LIEBERMAN-AIDEN, E., VAN BERKUM, N. L., WILLIAMS, L., IMAKAEV, M., RAGOCZY, T., TELLING, A., AMIT, I., LAJOIE, B. R., SABO, P. J. & DORSCHNER, M. O. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326, 289-293.
- LIESCHKE, G. J., OATES, A. C., CROWHURST, M. O., WARD, A. C. & LAYTON, J. E. 2001. Morphologic and functional characterization of granulocytes and macrophages in embryonic and adult zebrafish. *Blood, The Journal of the American Society of Hematology*, 98, 3087-3096.
- LIESCHKE, G. J., OATES, A. C., PAW, B. H., THOMPSON, M. A., HALL, N. E., WARD, A., HO, R. K., ZON, L. I. & LAYTON, J. E. 2002. Zebrafish SPI-1 marks a site of myeloid development independent of primitive erythropoiesis: implications for axial patterning. *Developmental biology*, 246, 274-295.
- LIEU, Y. K. & REDDY, E. P. 2009. Conditional c-myb knockout in adult hematopoietic stem cells leads to loss of self-renewal due to impaired proliferation and accelerated differentiation. *Proc Natl Acad Sci U S A*, 106, 21689-94.
- LINK, D. C. 2005. Neutrophil homeostasis. *Immunologic research*, 32, 169-178.
- LIU, F. & PATIENT, R. 2008. Genome-wide analysis of the zebrafish ETS family identifies three genes required for hemangioblast differentiation or angiogenesis. *Circulation research*, 103, 1147-1154.
- LIU, F. & WEN, Z. 2002. Cloning and expression pattern of the lysozyme C gene in zebrafish. *Mechanisms of development*, 113, 69-72.
- LORENZO, P. I., BREDEFORD, E. M., GILFILLAN, S., GAVRILOV, A. A., LEEDSAK, M., RAZIN, S. V., ESKELAND, R., SAETHER, T. & GABRIELSEN, O. S. 2011. Identification of c-Myb Target Genes in K562 Cells Reveals a Role for c-Myb as a Master Regulator. *Genes Cancer*, 2, 805-17.
- MAHONY, S. & BENOS, P. V. 2007. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic acids research*, 35, W253-W258.
- MATYS, V., FRICKE, E., GEFERS, R., GÖBLING, E., HAUBROCK, M., HEHL, R., HORNISCHER, K., KARAS, D., KEL, A. E. & KEL-MARGOULIS, O. V. 2003. TRANSFAC®: transcriptional regulation, from patterns to profiles. *Nucleic acids research*, 31, 374-378.
- MAZO, I. B., MASSBERG, S. & VON ANDRIAN, U. H. 2011. Hematopoietic stem and progenitor cell trafficking. *Trends in immunology*, 32, 493-503.
- MCGRATH, K. E., FRAME, J. M., FROMM, G. J., KONISKI, A. D., KINGSLEY, P. D., LITTLE, J., BULGER, M. & PALIS, J. 2011. A transient definitive erythroid lineage with unique regulation of the beta-globin locus in the mammalian embryo. *Blood*, 117, 4600-8.

- MEDVINSKY, A. & DZIERZAK, E. 1996. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, 86, 897-906.
- MEDVINSKY, A., RYBTSOV, S. & TAOUDI, S. 2011. Embryonic origin of the adult hematopoietic system: advances and questions. *Development*, 138, 1017-1031.
- MITRA, P. 2018. Transcription regulation of MYB: a potential and novel therapeutic target in cancer. *Annals of translational medicine*, 6.
- MOHAMED, T. 2015. *The Transcriptional Regulation of c-myb in Zebrafish Haematopoiesis* MRes, University of Nottingham.
- MOORE, C., RICHENS, J. L., HOUGH, Y., UCANOK, D., MALLA, S., SANG, F., CHEN, Y., ELWORTHY, S., WILKINSON, R. N. & GERING, M. 2018. Gfi1aa and Gfi1b set the pace for primitive erythroblast differentiation from hemangioblasts in the zebrafish embryo. *Blood advances*, 2, 2589-2606.
- MOORE, F. E., GARCIA, E. G., LOBBARDI, R., JAIN, E., TANG, Q., MOORE, J. C., CORTES, M., MOLODTSOV, A., KASHETA, M. & LUO, C. C. 2016. Single-cell transcriptional analysis of normal, aberrant, and malignant hematopoiesis in zebrafish. *Journal of Experimental Medicine*, 213, 979-992.
- MORENO-MATEOS, M. A., VEJNAR, C. E., BEAUDOIN, J.-D., FERNANDEZ, J. P., MIS, E. K., KHOKHA, M. K. & GIRALDEZ, A. J. 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature methods*, 12, 982.
- MÖRÖY, T., VASSEN, L., WILKES, B. & KHANDANPOUR, C. 2015. From cytopenia to leukemia: the role of Gfi1 and Gfi1b in blood formation. *Blood, The Journal of the American Society of Hematology*, 126, 2561-2569.
- MUCENSKI, M. L., MCLAIN, K., KIER, A. B., SWERDLOW, S. H., SCHREINER, C. M., MILLER, T. A., PIETRYGA, D. W., SCOTT JR, W. J. & POTTER, S. S. 1991. A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell*, 65, 677-689.
- MUKAI, H. Y., MOTOHASHI, H., OHNEDA, O., SUZUKI, N., NAGANO, M. & YAMAMOTO, M. 2006. Transgene Insertion in Proximity to the c-myb Gene Disrupts Erythroid-Megakaryocytic Lineage Bifurcation. *Molecular and cellular biology*, 26, 7953-7965.
- MURAYAMA, E., KISSA, K., ZAPATA, A., MORDELET, E., BRIOLAT, V., LIN, H. F., HANDIN, R. I. & HERBOMEL, P. 2006. Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development. *Immunity*, 25, 963-75.
- NELSON, A. C. & WARDLE, F. C. 2013. Conserved non-coding elements and cis regulation: actions speak louder than words. *Development*, 140, 1385-95.
- NESS, S. A. 1999. Myb binding proteins: regulators and cohorts in transformation. *Oncogene*, 18, 3039-3046.
- NESS, S. A., KOWENZ-LEUTZ, E., CASINI, T., GRAF, T. & LEUTZ, A. 1993. Myb and NF-M: combinatorial activators of myeloid genes in heterologous cell types. *Genes & development*, 7, 749-759.
- NIEMANN, C. U., ABRINK, M., PEJLER, G., FISCHER, R. L., CHRISTENSEN, E. I., KNIGHT, S. D. & BORREGAARD, N. 2007. Neutrophil elastase depends on serglycin proteoglycan for localization in granules. *Blood*, 109, 4478-4486.
- NORTH, T., GU, T.-L., STACY, T., WANG, Q., HOWARD, L., BINDER, M., MARÍN-PADILLA, M. & SPECK, N. A. 1999. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development*, 126, 2563-2575.
- NORTH, T. E., GOESSLING, W., PEETERS, M., LI, P., CEOL, C., LORD, A. M., WEBER, G. J., HARRIS, J., CUTTING, C. C. & HUANG, P. 2009. Hematopoietic stem cell development is dependent on blood flow. *Cell*, 137, 736-748.

- NOWICK, K. & STUBBS, L. 2010. Lineage-specific transcription factors and the evolution of gene regulatory networks. *Briefings in functional genomics*, 9, 65-78.
- OEHLERS, S. H., FLORES, M. V., HALL, C. J., O'TOOLE, R., SWIFT, S., CROSIER, K. E. & CROSIER, P. S. 2010. Expression of zebrafish cxcl8 (interleukin-8) and its receptors during development and in response to immune stimulation. *Developmental & Comparative Immunology*, 34, 352-359.
- OHTEKI, T., YOSHIDA, H., MATSUYAMA, T., DUNCAN, G. S., MAK, T. W. & OHASHI, P. S. 1998. The Transcription Factor Interferon Regulatory Factor 1 (IRF-1) Is Important during the Maturation of Natural Killer 1.1+ T Cell Receptor- α/β +(NK1+ T) Cells, Natural Killer Cells, and Intestinal Intraepithelial T Cells. *The Journal of experimental medicine*, 187, 967-972.
- OKUDA, T., VAN DEURSEN, J., HIEBERT, S. W., GROSVELD, G. & DOWNING, J. R. 1996. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, 84, 321-30.
- ONG, C.-T. & CORCES, V. G. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, 12, 283-293.
- ORKIN, S. H. & ZON, L. I. 2008a. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, 132, 631-44.
- ORKIN, S. H. & ZON, L. I. 2008b. SnapShot: hematopoiesis. *Cell*, 132, 712. e1-712. e2.
- OSTERWALDER, M., BAROZZI, I., TISSIÈRES, V., FUKUDA-YUZAWA, Y., MANNION, B. J., AFZAL, S. Y., LEE, E. A., ZHU, Y., PLAJZER-FRICK, I. & PICKLE, C. S. 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 554, 239-243.
- OTTERSBAACH, K. 2019. Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochemical Society Transactions*, 47, 591-601.
- PAIK, E. J. & ZON, L. I. 2010. Hematopoietic development in the zebrafish. *Int J Dev Biol*, 54, 1127-37.
- PALIS, J., ROBERTSON, S., KENNEDY, M., WALL, C. & KELLER, G. 1999. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development*, 126, 5073-84.
- PATTERSON, L. J., GERING, M., ECKFELDT, C. E., GREEN, A. R., VERFAILLIE, C. M., EKKER, S. C. & PATIENT, R. 2007. The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. *Blood*, 109, 2389-98.
- PENNACCHIO, L. A., BICKMORE, W., DEAN, A., NOBREGA, M. A. & BEJERANO, G. 2013. Enhancers: five essential questions. *Nat Rev Genet*, 14, 288-95.
- PERDIGUERO, E. G., KLAPPROTH, K., SCHULZ, C., BUSCH, K., AZZONI, E., CROZET, L., GARNER, H., TROUILLET, C., DE BRUIJN, M. F. & GEISSMANN, F. 2015. Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature*, 518, 547-551.
- PILLAY, L. M., FORRESTER, A. M., ERICKSON, T., BERMAN, J. N. & WASKIEWICZ, A. J. 2010. The Hox cofactors Meis1 and Pbx act upstream of gata1 to regulate primitive hematopoiesis. *Developmental biology*, 340, 306-317.
- POWERS, D. A. 1991. Evolutionary genetics of fish. *Advances in genetics*. Elsevier.
- RAAB, J. R. & KAMAKAKA, R. T. 2010. Insulators and promoters: closer than we think. *Nature Reviews Genetics*, 11, 439-446.
- RADA-IGLESIAS, A., BAJPAI, R., SWIGUT, T., BRUGMANN, S. A., FLYNN, R. A. & WYSOCKA, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470, 279-283.

- RAMSAY, R. G. 2005. c-Myb a stem-progenitor cell regulator in multiple tissue compartments. *Growth Factors*, 23, 253-261.
- RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8, 2281-2308.
- RAZ, E. 2003. Primordial germ-cell development: the zebrafish perspective. *Nature Reviews Genetics*, 4, 690-700.
- REISCHAUER, S., STONE, O. A., VILLASENOR, A., CHI, N., JIN, S.-W., MARTIN, M., LEE, M. T., FUKUDA, N., MARASS, M. & WITTY, A. 2016. Cloche is a bHLH-PAS transcription factor that drives haemato-vascular specification. *Nature*, 535, 294-298.
- RENSHAW, S. A., LOYNES, C. A., TRUSHELL, D. M., ELWORTHY, S., INGHAM, P. W. & WHYTE, M. K. 2006. A transgenic zebrafish model of neutrophilic inflammation. *Blood*, 108, 3976-3978.
- RHODES, J., HAGEN, A., HSU, K., DENG, M., LIU, T. X., LOOK, A. T. & KANKI, J. P. 2005. Interplay of pu. 1 and gata1 determines myelo-erythroid progenitor cell fate in zebrafish. *Developmental cell*, 8, 97-108.
- ROBERTSON, A. L., AVAGYAN, S., GANSNER, J. M. & ZON, L. I. 2016. Understanding the regulation of vertebrate hematopoiesis and blood disorders—big lessons from a small fish. *FEBS letters*, 590, 4016-4033.
- ROBERTSON, G., HIRST, M., BAINBRIDGE, M., BILENKY, M., ZHAO, Y., ZENG, T., EUSKIRCHEN, G., BERNIER, B., VARHOL, R. & DELANEY, A. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods*, 4, 651-657.
- ROJANO, E., SEOANE, P., RANEA, J. A. & PERKINS, J. R. 2019. Regulatory variants: from detection to predicting impact. *Briefings in bioinformatics*, 20, 1639-1654.
- ROSOWSKI, E. E., DENG, Q., KELLER, N. P. & HUTTENLOCHER, A. 2016. Rac2 functions in both neutrophils and macrophages to mediate motility and host defense in larval zebrafish. *The Journal of Immunology*, 197, 4780-4790.
- RYAN, G. E. & FARLEY, E. K. 2020. Functional genomic approaches to elucidate the role of enhancers during development. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 12, e1467.
- RYBTSOV, S., IVANOV, A., ZHAO, S. & MEDVINSKY, A. 2016. Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver. *Development*, 143, 1284-1289.
- SAGAI, T., AMANO, T., MAENO, A., KIYONARI, H., SEO, H., CHO, S.-W. & SHIROISHI, T. 2017. SHH signaling directed by two oral epithelium-specific enhancers controls tooth and oral development. *Scientific reports*, 7, 1-11.
- SAGAI, T., HOSOYA, M., MIZUSHINA, Y., TAMURA, M. & SHIROISHI, T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development*, 132, 797-803.
- SANDELIN, A., ALKEMA, W., ENGSTRÖM, P., WASSERMAN, W. W. & LENHARD, B. 2004a. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic acids research*, 32, D91-D94.
- SANDELIN, A., WASSERMAN, W. W. & LENHARD, B. 2004b. ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic acids research*, 32, W249-W252.
- SANDER, J. D. & JOUNG, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, 32, 347-55.
- SAVAGE, A. 2012. *MRes Thesis*. University of Nottingham.
- SHAFIZADEH, E. & PAW, B. H. 2004. Zebrafish as a model of human hematologic disorders. *Current opinion in hematology*, 11, 255-261.

- SHEN, L.-J., CHEN, F.-Y., ZHANG, Y., CAO, L.-F., KUANG, Y., ZHONG, M., WANG, T. & ZHONG, H. 2013. MYCN transgenic zebrafish model with the characterization of acute myeloid leukemia and altered hematopoiesis. *PLoS one*, 8, e59070.
- SHMAKOV, S., ABUDAYYEH, O. O., MAKAROVA, K. S., WOLF, Y. I., GOOTENBERG, J. S., SEMENOVA, E., MINAKHIN, L., JOUNG, J., KONERMANN, S., SEVERINOV, K., ZHANG, F. & KOONIN, E. V. 2015. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell*, 60, 385-97.
- SILVER, L. & PALIS, J. 1997. Initiation of murine embryonic erythropoiesis: a spatial analysis. *Blood*, 89, 1154-1164.
- SITZMANN, J., NOBEN-TRAUTH, K. & KLEMPNAUER, K.-H. 1995. Expression of mouse c-myb during embryonic development. *Oncogene*, 11, 2273-2279.
- SOARES-DA-SILVA, F., FREYER, L., ELSAID, R., BURLIN-DEFRANOUX, O., ITURRI, L., SISMEIRO, O., PINTO-DO-Ó, P., GOMEZ-PERDIGUERO, E. & CUMANO, A. 2021. Yolk sac, but not hematopoietic stem cell-derived progenitors, sustain erythropoiesis throughout murine embryonic life. *Journal of Experimental Medicine*, 218.
- SONG, H., YAN, Y.-L., TITUS, T., HE, X. & POSTLETHWAIT, J. H. 2011. The role of stat1b in zebrafish hematopoiesis. *Mechanisms of development*, 128, 442-456.
- SOZA-RIED, C., HESS, I., NETUSCHIL, N., SCHORPP, M. & BOEHM, T. 2010. Essential role of c-myb in definitive hematopoiesis is evolutionarily conserved. *Proceedings of the National Academy of Sciences*, 107, 17304-17308.
- SPITZ, F. & FURLONG, E. E. 2012. Transcription factors: from enhancer binding to developmental control. *Nature reviews genetics*, 13, 613-626.
- STACHURA, D. L. & TRAVER, D. 2011. Cellular dissection of zebrafish hematopoiesis. *Methods in cell biology*, 101, 75-110.
- STADHOUDERS, R., THONGJUEA, S., ANDRIEU-SOLER, C., PALSTRA, R. J., BRYNE, J. C., VAN DEN HEUVEL, A., STEVENS, M., DE BOER, E., KOCKX, C. & VAN DER SLOOT, A. 2012. Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *The EMBO journal*, 31, 986-999.
- STAINIER, D., WEINSTEIN, B. M., DETRICH, H. R., ZON, L. I. & FISHMAN, M. C. 1995. Cloche, an early acting zebrafish gene, is required by both the endothelial and hematopoietic lineages. *Development*, 121, 3141-3150.
- STRYDOM, N. & RANKIN, S. M. 2013. Regulation of circulating neutrophil numbers under homeostasis and in disease. *Journal of innate immunity*, 5, 304-314.
- STURGEON, C. M., DITADI, A., CLARKE, R. L. & KELLER, G. 2013. Defining the path to hematopoietic stem cells. *Nat Biotechnol*, 31, 416-8.
- SUGIYAMA, D., INOUE-YOKOO, T., FRASER, S. T., KULKEAW, K., MIZUOCHI, C. & HORIO, Y. 2011. Embryonic regulation of the mouse hematopoietic niche. *The Scientific World Journal*, 11, 1770-1780.
- SUMNER, R., CRAWFORD, A., MUCENSKI, M. & FRAMPTON, J. 2000. Initiation of adult myelopoiesis can occur in the absence of c-Myb whereas subsequent development is strictly dependent on the transcription factor. *Oncogene*, 19, 3335-3342.
- TAMPLIN, O. J., DURAND, E. M., CARR, L. A., CHILDS, S. J., HAGEDORN, E. J., LI, P., YZAGUIRRE, A. D., SPECK, N. A. & ZON, L. I. 2015. Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche. *Cell*, 160, 241-252.
- TANAKA, Y., PATESTOS, N. P., MAEKAWA, T. & ISHII, S. 1999. B-myb is required for inner cell mass formation at an early stage of development. *Journal of Biological Chemistry*, 274, 28067-28070.

- TAOUDI, S. & MEDVINSKY, A. 2007. Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta. *Proceedings of the National Academy of Sciences*, 104, 9399-9403.
- THAMBYRAJAH, R., UCANOK, D., JALALI, M., HOUGH, Y., WILKINSON, R. N., MCMAHON, K., MOORE, C. & GERING, M. 2016. A gene trap transposon eliminates haematopoietic expression of zebrafish Gfi1aa, but does not interfere with haematopoiesis. *Developmental biology*, 417, 25-39.
- THOMAS, M. D., KREMER, C. S., RAVICHANDRAN, K. S., RAJEWSKY, K. & BENDER, T. P. 2005. c-Myb is critical for B cell development and maintenance of follicular B cells. *Immunity*, 23, 275-286.
- THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H. & VERNOT, B. 2012. The accessible chromatin landscape of the human genome. *Nature*, 489, 75-82.
- TOBER, J., KONISKI, A., MCGRATH, K. E., VEMISHETTI, R., EMERSON, R., DE MESY-BENTLEY, K. K., WAUGH, R. & PALIS, J. 2006. The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood*, 109, 1433-1441.
- TOSCANI, A., METTUS, R. V., COUPLAND, R., SIMPKINS, H., LITVIN, J., ORTH, J., HATTON, K. S. & REDDY, E. P. 1997. Arrest of spermatogenesis and defective breast development in mice lacking A-myb. *Nature*, 386, 713-717.
- TRAVER, D., HERBOMEL, P., PATTON, E. E., MURPHEY, R. D., YODER, J. A., LITMAN, G. W., CATIC, A., AMEMIYA, C. T., ZON, L. I. & TREDE, N. S. 2003a. The zebrafish as a model organism to study development of the immune system. *Advances in immunology*, 81, 253.
- TRAVER, D., PAW, B. H., POSS, K. D., PENBERTHY, W. T., LIN, S. & ZON, L. I. 2003b. Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol*, 4, 1238-46.
- TSIKANDELOVA, R. 2013. *Cis-regulation of cmyb in zebrafish*. MRes University of Nottingham.
- TULOTTA, C., STEFANESCU, C., BELETKAIA, E., BUSSMANN, J., TARBASHEVICH, K., SCHMIDT, T. & SNAAR-JAGALSKA, B. E. 2016. Inhibition of signaling between human CXCR4 and zebrafish ligands by the small molecule IT1t impairs the formation of triple-negative breast cancer early metastases in a zebrafish xenograft model. *Disease models & mechanisms*, 9, 141-153.
- VEGIOPOULOS, A., GARCÍA, P., EMAMBOKUS, N. & FRAMPTON, J. 2006. Coordination of erythropoiesis by the transcription factor c-Myb. *Blood*, 107, 4703-4710.
- VELDMAN, M. B. & LIN, S. 2008. Zebrafish as a developmental model organism for pediatric research. *Pediatr Res*, 64, 470-6.
- VISEL, A., BRISTOW, J. & PENNACCHIO, L. A. Enhancer identification through comparative genomics. *Seminars in cell & developmental biology*, 2007. Elsevier, 140-152.
- VISEL, A., RUBIN, E. M. & PENNACCHIO, L. A. 2009. Genomic views of distant-acting enhancers. *Nature*, 461, 199-205.
- WAHLBERG, K., JIANG, J., ROOKS, H., JAWAID, K., MATSUDA, F., YAMAGUCHI, M., LATHROP, M., THEIN, S. L. & BEST, S. 2009. The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells. *Blood, The Journal of the American Society of Hematology*, 114, 1254-1262.
- WANG, Q., STACY, T., BINDER, M., MARIN-PADILLA, M., SHARPE, A. H. & SPECK, N. A. 1996. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive

- hematopoiesis. *Proceedings of the National Academy of Sciences*, 93, 3444-3449.
- WEI, Y., XU, J., ZHANG, W., WEN, Z. & LIU, F. 2016. RNA polymerase III component Rpc9 regulates hematopoietic stem and progenitor cell maintenance in zebrafish. *Development*, 143, 2103-2110.
- WESTERFIELD, M. 1993. The zebrafish book; a guide for the laboratory use of zebrafish (*brachydanio rerio*). 2nd ed. Eugene, University of Oregon Press.
- WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153, 307-19.
- WIEDENHEFT, B., STERNBERG, S. H. & DOUDNA, J. A. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, 482, 331-8.
- WILLETT, C. E., ZAPATA, A. G., HOPKINS, N. & STEINER, L. A. 1997. Expression of zebrafish rag genes during early development identifies the thymus. *Dev Biol*, 182, 331-41.
- WOOLFE, A., GOODSON, M., GOODE, D. K., SNELL, P., MCEWEN, G. K., VAVOURI, T., SMITH, S. F., NORTH, P., CALLAWAY, H. & KELLY, K. 2004. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*, 3, e7.
- WRIGHT, J. 2019. *Gene Control*, United Kingdom, Essex, ED-Tech Press.
- YAMANE, T. 2018. Mouse yolk sac hematopoiesis. *Frontiers in Cell and Developmental Biology*, 6, 80.
- YANG, C.-T., CAMBIER, C., DAVIS, J. M., HALL, C. J., CROSIER, P. S. & RAMAKRISHNAN, L. 2012. Neutrophils exert protection in the early tuberculous granuloma by oxidative killing of mycobacteria phagocytosed from infected macrophages. *Cell host & microbe*, 12, 301-312.
- YOSHIMOTO, M., MONTECINO-RODRIGUEZ, E., FERKOWICZ, M. J., PORAYETTE, P., SHELLEY, W. C., CONWAY, S. J., DORSHKIND, K. & 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. *Proceedings of the National Academy of Sciences*, 108, 1468-1473.
- YZAGUIRRE, A. D., DE BRUIJN, M. F. & SPECK, N. A. 2017. The role of Runx1 in embryonic blood cell formation. *RUNX proteins in development and cancer*, 47-64.
- ZHANG, J., HAN, B., LI, X., BIES, J., JIANG, P., KOLLER, R. P. & WOLFF, L. 2016. Distal regulation of c-myb expression during IL-6-induced differentiation in murine myeloid progenitor M1 cells. *Cell death & disease*, 7, e2364-e2364.
- ZHANG, Y., JIN, H., LI, L., QIN, F. X. & WEN, Z. 2011. cMyb regulates hematopoietic stem/progenitor cell mobilization during zebrafish hematopoiesis. *Blood*, 118, 4093-101.
- ZHANG, Y., ZHANG, Z. & GE, W. 2018. An efficient platform for generating somatic point mutations with germline transmission in the zebrafish by CRISPR/Cas9-mediated gene editing. *Journal of Biological Chemistry*, 293, 6611-6622.
- ZHU, Y. P., PADGETT, L., DINH, H. Q., MARCOVECCHIO, P., BLATCHLEY, A., WU, R., EHINGER, E., KIM, C., MIKULSKI, Z. & SEUMOIS, G. 2018. Identification of an early unipotent neutrophil progenitor with pro-tumoral activity in mouse and human bone marrow. *Cell reports*, 24, 2329-2341. e8.