

University of Nottingham
UK | CHINA | MALAYSIA

Investigating the Roles of Long-Non-Coding RNA Genes in Zebrafish Haematopoiesis

Fatemah Alzaqqah

BSc, MSc

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

September 2019

Acknowledgements

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

Nobody has been more important to me in the pursuit of this PhD journey than the members of my family who have provided me through moral and emotional, whose love and guidance are with me in whatever I pursue. They are the ultimate role models. I would like to thank my parents (**my father, Mr Najem** and **my mother, Mrs Aliya**) and siblings (**my sisters Anwar**, **Amna**, **Zainab**, **Muna**, **Mariyam** and **my brother Abdullah**).

I would like to express my sincere gratitude 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. I could not have imagined having a better supervisor and mentor for my PhD study.

I am grateful to all friends who were supporting me spiritually throughout writing this thesis and my PhD journey in general. In particular, I am grateful to my best friends (**Jawaher Almulhim** and **Amna Jasim**).

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

I thank my fellow lab mates in for all help, support, stimulating discussions, and for all the fun we have had in the last four years. With a special mention to Dr Chris Moore, Dr Joanna Richens, Mariam

Jalali, Yasmin Hough, Peter Jessop, Mariam Jaddah, Farah Alnassay and Aisha Alshanqiti. It was fantastic to have the opportunity to work with you.

List of abbreviations

AB	Angioblast
AGM	Aorta-Gonad-Meshonephros
ALM	Anterior Lateral Mesoderm
AML1	Acute Myelogenous Leukaemia 1
AMP	Ampicillin
CHT	Caudal Hematopoietic Tissue
clo	*Cloche*
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CRISPR/Cas	Clustered, Regularly Interspaced, Short Palindromic Repeats/ CRISPR associated system
crRNAs	CRISPR RNAs
DA	Dorsal aorta
dH_2O	Distilled Water
dHSC	Definitive haematopoietic stem cell
dpf	Days Post Fertilization
DSB	double strand break
E	Embryonic
EC	Endothelial cell
eGFP	Enhanced Green Fluorescent Protein
EHT	Endothelial to Haematopoietic Transition

EMkMPP	Erythroid-megakaryocyte-primed multipotent progenitor
EMPs	Erythromyeloid Progenitors
FACS	Fluorescence activated cell sorting
gDNA	Genomic DNA
GFP	Green fluorescent protein
gRNA	Guide Ribonucleic Acid
HCs	Haematopoietic Cells
HDR	homology-directed repair
hpf	Hours Post Fertilisation
HSC	Haematopoietic Stem Cell
HSPC	haematopoietic stem progenitor cell
Hybe	Hybridization Buffer
ICM	Intermediate Cell Mass
IVT	In vitro translation
KM	Kidney marrow
LB medium	Luria-Bertani Medium
LMPP	Lymphoid-myeloid primed progenitor
lncRNA	Long non-coding RNA
mib	Mindbomb
MPP	Multipotent progenitor
NHEJ	non-homologous end joining
NK	Natural killer
PAM	protospacer adjacent motif
PBI	Posterior Blood Island

PBS	Phosphate Buffered Saline
PCV	Posterior Cardinal Vein
PFA	Paraformaldehyde
PLM	Posterior Lateral Mesoderm
preNM	Pre-neutrophil-monocyte
prRBCs	Primitive Red Blood Cells
RPKM	Read per kilobase per million
RPKMs	reads per kilobase millions
RT	Room Temperature
***RUNX*1**	Runt-related transcription factor
SMO	smoothened
TAE	Tris Acetate Edta Buffer
TE	Tris Edta Buffer
VBI	Ventral blood island
vDA	Ventral Wall of The Dorsal Aorta
VEGF	Vascular endothelial growth factor
WISH	whole mount in situ hybridisation
wt	Wild-Type
YS	Yolk Sac

Table of contents

List of abbreviations ---------- **iii**

Table of contents ---------- **vi**

List of figures ---------- **x**

List of tables ---------- **xii**

Abstract ---------- **xiii**

1. Introduction 1

1.1 Haematopoiesis in vertebrates 4

1.1.1 The primitive wave of murine haematopoiesis give rise to haemangioblast 7

1.1.2 Erythroid-myeloid progenitors generation in the yolk sac as a transient wave 10

1.1.3 The definitive wave gives rise a long term haematopoietic stem cells in the murine embryos 10

1.1.4 Embryogenesis and haematopoiesis of zebrafish 12

1.1.5 The embryogenesis and primitive haematopoiesis wave of zebrafish 14

1.1.6 Erythroid myeloid progenitors are raised in the posterior blood island (PBI) as an intermediate wave of zebrafish 17

1.1.7 vDA of zebrafish embryos generate long-term haematopoietic stem cells 17

1.2 The zebrafish gene trap line *qmc551* eliminates primitive and definitive haematopoietic gfi1aa expression without interfering with normal haematopoiesis 20

1.3 Gfi1aa and Gfi1b promote late erythroid gene expression by suppress endothelial gene expression in immature primitive red blood cells 22

1.4 Gfi1ab is upregulated in the vDA in the absence of *gfi1aa* 23

1.5 RNA-seq on qmc551:GFP+ cells identifies two lncRNA genes that are specifically expressed in embryonic blood cells 25

1.5.1 lncRNA genes play essential regulatory roles in numerous cells types, including haematopoietic cells 31

1.5.1.1 long non-coding RNAs 31

1.5.1.2 lncRNA genes play important roles during the regulation of blood cells development 34

1.5.1.3 lncRNA in diseases and disorders 36
1.6 The CRISPR/Cas9 system as a tool for lncRNA genes deletion . 38
1.7 Project aims and Objectives 41

2. Materials and methods 42

2.1 Materials 42
2.1.1 Buffers and solutions 42
2.1.2 Zebrafish maintenance and breeding 44
2.1.3 Mutant and transgenic zebrafish lines used for whole amount in situ hybridisation in this study 44
2.1.4 Gene information, Synteny and Alignment 45
2.1.5 Primer design 45
2.2 Molecular biology methods 46
2.2.1 Embryo lysis or fin clip biopsies for DNA extraction and genotype 46
2.2.2 Genotyping embryos after *in situ* hybridisation 46
2.2.3 PCR amplification 47
2.2.4 Agarose gel electrophoresis 48
2.2.5 DNA extraction from agarose gel & measurement of nucleic acid concentrations 49
2.2.6 DNA sequencing 49
2.3 Whole mount *in situ* RNA Hybridization (WISH) 50
2.3.1 Making probe 50
2.3.1.1 cDNA ligation, transformation and extraction 50
2.3.1.2 *In vitro* transcription to make antisense RNA probe 51
2.3.1.3 Embryo collection, chorion removal and fixation 53
2.3.1.4 Whole mount RNA *in situ* hybridisation (WISH) 53
2.3.1.5 Imaging 55
2.4 Targeting genes using the CRISPR/Cas9 technique 56
2.4.1 Single-guide RNA (sgRNA) design 56
2.4.2 Cloning of the sgRNA-coding linker in pDR274 58
2.4.3 sgRNA synthesis for microinjection 59
2.4.4 Microinjection needle preparation 60
2.4.5 Microinjection 61
2.5 Adult zebrafish fin clipping 62
2.6 Caudal fin clipping at three days post-fertilization 62

3. Results 63

3.1 The *lncRNA1* and *lncRNA2* genes are expressed in primitive and definitive haematopoietic cells in zebrafish embryos 63

3.1.1 Introduction 63

3.1.2 *lncRNA1* is expressed in primitive and definitive haematopoietic cells in zebrafish embryos 63

3.1.3 *lncRNA2* is expressed in primitive and definitive haematopoietic cells in zebrafish embryos 66

3.2 Loss of *lncRNA1* and *lncRNA2* expression in cloche mutants confirms that the *lncRNA*s are expressed in blood and endothelial cells in the wt zebrafish embryo 68

3.3 The loss of *lncRNA1* and *lncRNA2* expressions in a subset of cells in *mib* mutant reveals that these cells are definitive blood cell progenitors 72

3.4 *lncRNA2* is upregulated in the vDA and PBI of qmc551 homozygous mutants 76

3.5 *lncRNA2* expression is lost in the onset of definitive haematopoietic cells in Gfi1aa/1ab double mutants 79

3.6 *lncRNA1* expression is depleted in the vDA and adult haematopoietic cells of *runx1* homozygous zebrafish embryos 81

3.7 lncRNA1 gene expression was observed in T cells of zebrafish embryos 85

3.8 lncRNA1 and lncRNA2 have homologues in two other members of the carp family of freshwater fish but not in other vertebrate species 90

3.9 The genomic map of *lncRNA1* and *lncRNA2* of wt zebrafish 94

3.10 Deletion of *lncRNA* genes from the wild-type zebrafish genome using CRISPR/Cas9 97

3.11 Using CRISPR/Cas9 to delete lncRNA genes from *qmc551/qmc554* double mutant zebrafish 106

3.12 CRISPR/Cas9 was used to delete lncRNA genes from the wt zebrafish genome 112

3.13 CRISPR/Cas9 system was used to delete *lncRNA1* from the zebrafish genome 118

3.13.1 Introduction 118

3.14 CRISPR/Cas9 system was used to generate a *lncRNA1* promoter/*lncRNA2* mutant allele 122

3.14.1 Introduction 122

3.15 Mutant *lncRNA2* alleles were generated using the CRISPR/Cas9 system 132

4. Discussion 140

4.1 lncRNA1 and lncRNA2 expression is associated with haematopoiesis during zebrafish development 141

4.2 *lncRNA1* and *2* expression is blood-specific but variable in different haematopoietic lineages 141
4.3 *lncRNA2* expression in prRBC is repressed by Gfi1aa 144
4.4 *qmc222* homozygous fish survive to adulthood and are fertile 146
4.5 qmc221 allele is not transmitted to the next generation 147
4.6 investigation the role of lncRNA2 in prRBC by generating triple mutant of qmc222/qmc551/qmc554 homozygous..... 148
4.7 Future perspectives 149

5. Appendix 150

5.1 Appendix A 150
5.2 Appendix B 155
5.3 Appendix C 157
5.4 Appendix D 158
5.5 Appendix E 159

List of Figures

Figure 1.1 Adult haematopoiesis hierarchy 3

Figure 1.2 Conserved of haematopoiesis development 6

Figure 1.3 Haematopoietic sites development in murine embryos 9

Figure 1.4 Haematopoiesis in zebrafish 13

Figure 1.5 Haematopoiesis sites of zebrafish 16

Figure 1.6 Expression of mouse Gfi1/1b and zebrafish Gfi1/1b/ab during primitive erythropoiesis and definitive haematopoiesis in wild-type embryos 21

Figure 1.7 Expression of zebrafish gfi1/1b paralogs during primitive erythropoiesis and definitive haematopoiesis in wild-type and homozygous qmc551 embryos. 24

Figure 1.8 30 hpf zebrafish gene trap line qmc551 27

Figure 1.9 Flowchart summarizing the main steps in the RNA-seq procedure for data analysis of differentially expressed genes 28

Figure 1.10 lncRNA1 and lncRNA2 are expressed in sites associated with primitive and definitive haematopoiesis of zebrafish 30

Figure 1.11 Long non-coding RNAs have the capacity to regulate gene expression at variant levels 33

Figure 1.12 Expression of RUNXOR lncRNA in AML 37

Figure 1.13 The process used by CRISPR/Cas to cleave foreign DNA as part of the acquired immune system 39

Figure 3.1 *lncRNA1* is expressed in sites associated with primitive and definitive haematopoiesis during zebrafish development 65

Figure 3.2 *lncRNA2* expression is detected at sites associated with primitive and definitive haematopoiesis during zebrafish development 67

Figure 3.3 Loss of *lncRNA1* and *lncRNA2* expression in *cloche* mutant embryos confirms that the lncRNAs are expressed in blood and/or endothelial cells in the wt zebrafish embryo 70

Figure 3.4 The loss of *lncRNA1* and *lncRNA2* expressions in a subset of cells in *mib* mutant embryos reveals that these cells are definitive blood cell progenitors 74

Figure 3.5 *lncRNA2* is upregulated in the vDA and PBI of *qmc551* homozygous mutants 77

Figure 3.6 *lncRNA2* gene expression was lost in the definitive blood cells of gfi*1aa/gfi1ab* double mutants 80

Figure 3.7 *lncRNA1* expression is strongly reduced in the vDA of *runx1* homozygous zebrafish embryos 83

Figure 3.8 *lncRNA1* gene expression is lost due to loss of the T cells of cmyb mutant embryos 86

Figure 3.9 The alignment of lncRNA1 and lncRNA2 transcripts between zebrafish and members of the carp family of freshwater fish 92

Figure 3.10 Conservation shared synteny genes human and zebrafish .. 93

Figure 3.11 Conservation shared synteny genes mouse and zebrafish .. 93

Figure 3.12 Mapping of *lncRNA1* and *lncRNA2* wt zebrafish genes 95

Figure 3.13 Generation of gRNAs to knockout *lncRNA1* and *lncRNA2* genes of zebrafish using the CRISPR/Cas9 system 99

Figure 3.14 The CRISPR/Cas9 system was used to delete the *lncRNA1* and *lncRNA2* genes from zebrafish genome 104

Figure 3.15 CRISPR/Cas9 system was used to delete the *lncRNA* genes from qmc551 +/-; qmc554 +/- zebrafish embryos..... 108

Figure 3.16 Identification of qmc221 heterozygous larvae..... 111

Figure 3.17 Knockout *lncRNA1* and *lncRNA2* genes of zebrafish using the CRISPR/Cas9 system 115

Figure 3.18 Generation and identification of *qmc224* carrier..... 116

Figure 3.19 Genomic deletion of the zebrafish *lncRNA1* gene using the CRISPR/Cas9 system 120

Figure 3.20 Knockout *lncRNA2* gene of zebrafish using the CRISPR/Cas9 system 125

Figure 3.21 Generation and identification of qmc222 carrier..... 127

Figure 3.22 Generation and identification of qmc222 homozygous larvae 128

Figure 3.23 Loss of *lncRNA2* expression in *qmc222* mutant confirmed that the CRISPR/Cas9 system deleted *lncRNA2* gene of zebrafish 130

Figure 3.24 *lncRNA1* expression is retained in thymic in *qmc222* 131

Figure 3.25 Knockout *lncRNA2* gene of zebrafish using the CRISPR/Cas9 system 133

Figure 3.26 Generation and identification of *qmc223* heterozygous larvae 136

Figure 3.27 Loss of lncRNA2 expression in *qmc223* mutant confirmed that the CRISPR/Cas9 system deleted *lncRNA2* gene of zebrafish 138

Figure 4.1 The levels of *lncRNA1* and *lncRNA2* expression in prRBCs at 20hpf in Gfi1aa and/or Gfi1b-depleted embryos 145

Figure 5.1 *lncRNA1* and *lncRNA2* probes synthesis for whole-mount in situ hybridisation 160

Figure 5.2 Continue *lncRNA1* and *lncRNA2* probes synthesis for whole-mount in situ hybridisation. (A) *lncRNA1* and (B) *lncRNA2* 162

List of Tables

Table 1.1 List of the some interesting genes expression in qmc551:GFP+ cells relative to qmc551:GFP- and an RPKM value in qmc551:GFP+ cells. 29

Table 2.1 List of buffers and solutions. The list provides a description of their ingredients and gives the temperature they were stored at 42

Table 2.2 Reaction components for a standard PCR experiment using NEB standard Taq polymerase 48

Table 2.3 General PCR thermocycler steps and conditions using NEB Standard Taq Polymerase 48

Table 2.4 The reaction set up to ligate the PCR fragment with pGEM-Teasy 50

Table 2.5 The reaction set up for in vitro transcription..... 52

Table 2.6 Proteinase K treatment durations according to the age of embryos..... 53

Table 2.7 The sequences and oligonucleotides used for targets site..... 57

Table 2.8 The reaction mix to insert the linker into the expression vector 58

Table 2.9 The mixture reaction to make *in vitro* transcription..... 60

Table 3.1 illustrates sets of primers used in PCR amplification (GRCz11) 96

Table 3.2 The attempts to knockout *lncRNA1* and *lncRNa2* from wt zebrafish genome by injected with TS14 and TS65 105

Abstract

In vertebrates, haematopoiesis, the process of blood formation, is governed by a complex network of gene regulators that include numerous transcription factors, members of signalling pathways as well as regulatory RNA species. In a previous study, a transposon-mediated gene trap screen performed in zebrafish identified the gene trap line *qmc551* that expresses green fluorescent protein (GFP) in primitive red blood cells (prRBCs) and in definitive haematopoietic progenitors. This line carries the transposon in intron 1 of the gene that encodes the transcriptional repressor Gfi1aa. The gene trap interferes with its haematopoietic expression and expresses GFP in its place. A transcriptome analysis performed on GFP-positive and GFP-negative cells isolated from 30 hour-old *qmc551* heterozygous embryos revealed that two long-noncoding RNAs (si:ch211-214p16.1 (lncRNA1) and si:ch211-214p16.2 (lncRNA2)) are specifically expressed in GFP-positive cells. Whole-mount RNA in situ hybridisation on wt and *cloche* mutant embryos confirmed the blood-specific expression patterns of lncRNA1 and 2, and analysis of *mindbomb* mutant embryos showed that both prRBCs, as well as immature definitive blood progenitors expressed the two lncRNAs. In the prRBCs, lncRNA1 and 2 expressions are subject to Gfi1aa-mediated repression, as embryos homozygous for the *qmc551* mutant allele of *gfi1aa* displayed elevated lncRNA1 and 2 levels. While all early embryonic blood cells express higher levels of lncRNA2, lncRNA1 alone displays high-level expression in T-cell progenitors in the thymus. Altogether, their expression patterns suggested a specific role for the lncRNAs in early blood development. To study this role, the CRISPR/Cas9 technology was used to delete their genes individually and in combination from zebrafish chromosome 17, where they occupy adjacent positions at a common lncRNA1/2 locus. Here, the generation and preliminary

characterisation of the lncRNA1/2 locus mutant alleles *qmc221*, *qm222*, *qmc223* and *qmc224* are described. These experiments form the foundation for future investigations into the roles that these lncRNAs play within the gene regulatory network that controls haematopoiesis.

1. Introduction

Haematopoietic stem cells (HSCs) are immature cells. They have the capacity to self-renew and the ability to give rise to committed progenitor cells that give rise to all cellular components of the blood. This complicated process, called haematopoiesis, first occurs during embryonic development and continues throughout adulthood (Pallister and Watson, 2011). It is well regulated intrinsically and extrinsically controlled by transcription factors as well as microRNA, lncRNAs and others, and an array of growth factors, respectively (Jagannathan-Bogdan and Zon, 2013, Hoffbrand et al., 2010). Growth factors stimulate the proliferation and differentiation of haematopoietic progenitor cells to ensure that different types of blood cell are produced in the precise proportions required to maintain normal counts of circulating blood cells (Jagannathan-Bogdan and Zon, 2013, Hoffbrand et al., 2010). The haematopoietic progenitors differentiate into a variety of specialised cells. Each blood cell type follows a slightly different path of differentiation. HSCs give rise into multipotent progenitors (MPPs) which have limited self-renewal capacity. According to the composite model of the haematopoietic hierarchy, MPPs differentiate into common myeloid progenitors (CMPs) and lymphoid primed multipotent progenitors (LMPPs). First, CMPs give rise to two main progenitors: erythroid-megakaryocyte-primed multipotent progenitors (EMkMPPs) and pre-neutrophil-monocyte (preNM). Platelets, erythrocytes, basophil/master cells and eosinophils are placed downstream of EMkMPPs (Figure 1.1). Whereas, neutrophils and macrophages are placed downstream of preNM. Second, LMPPs are placed upstream of preNM and common lymphoid progenitors (CLPs), which differentiate into B-cells, T-cells and natural killer (NK) cells (Drissen et al., 2016) (Figure 1.1). Haematopoiesis in adult mammals is conducted at the bone marrow while the kidney marrow

is the main haematopoietic organ in zebrafish (Jagannathan-Bogdan and Zon, 2013, Hoffbrand et al., 2010).

There is a great deal of interest in haematopoietic stem cells because of their potential applications in tissue engineering, regenerative medicine, cell therapy, and gene therapy (Seita and Weissman, 2010). Despite improvements in transplant medicine and antigen matching, "graft versus host disease" remains a major complication of haematopoietic stem cells transplantation. In allogeneic transplant settings, 80% of those who receive an antigen-matched transplant still show signs of graft versus host disease (Petersdorf, 2013). Therefore, a better understanding of the processes involved in haematopoietic stem cell programming in the embryo can provide future medicine with the potential ability of reprogrammed haematopoietic stem cell as cell replacement therapies. Such an idea could serve patients with an alternative source of antigen matched HSCs transplant with minimum to none graft versus host disease onset (Brunstein and Wagner, 2006, Riddell et al., 2014).

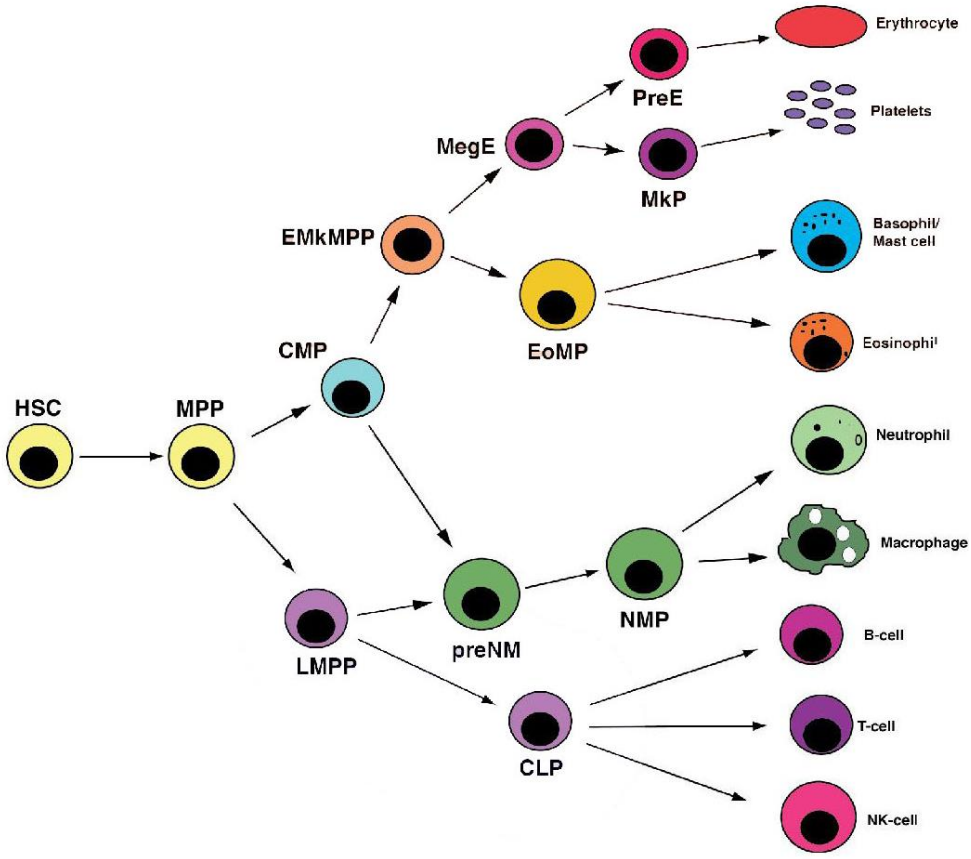


Figure 1.1 Adult haematopoiesis hierarchy

HSC multipotent on top of the hierarchy. HSCs give rise to multipotent progenitors (MPP), which differentiate into common myeloid progenitors (CMPs) and lymphoid primed multipotent, (LMPPs). CMPs are upstream of erythroid-megakaryocyte-primed multipotent progenitors (EMkMPPs), and pre-neutrophil-monocyte (preNM). EMkMPPs give rise to megakaryocyte-erythrocyte lineage (MegE) that differentiate into Platelets, erythrocytes. Eosinophil-mast cell progenitors (EoMPs) are placed upstream of basophil/master cells and eosinophils. LMPPs is placed upstream of common lymphoid progenitors (CLPs) and pre-neutrophil-monocyte (preNM). The CLP differentiate into B-cells, T-cells and natural killer (NK) cells. The preNMs give rise to neutrophil-monocyte progenitors (NMPs) which differentiate into neutrophils and macrophages (Drissen et al., 2016).

1.1 Haematopoiesis in vertebrates

Haematopoiesis is conserved across all vertebrates with only a few differences existing. In vertebrates' haematopoiesis involves three waves, the primitive wave, transient wave and the definitive wave. In the primitive wave, mesodermal progenitors give rise to haemangioblasts that give rise to two types of cells; a transient haematopoietic cells and endothelial cells. The haematopoietic cells such as macrophages and primitive erythrocytes (prRBCs). The primitive wave takes place in the yolk sac (YS) blood islands in the mouse, the dorsal aorta and the ventral blood island in the amphibian embryo, anterior and posterior ventral blood islands (VBI) in the Xenopus and anterior and posterior lateral mesoderm in the zebrafish. The primitive wave is temporary and does not generate HSCs with self-renewal capability (Figure 1.2). The transient (late fetal and early adult) wave is believed to generate erythroid–myeloid progenitors (EMPs) from haemogenic endothelium in the wall of blood vessels in the yolk sac (mouse). The EMPs migrate to the fetal liver. In the Xenopus and zebrafish, the EMPs arise from posterior lateral VBI and posterior blood islands (PBI), respectively (Figure 1.2). The definitive haematopoiesis occurs later in embryo development in which it generates haemogenic endothelial cells (HECs). These *de novo* HECs are a specialised population of endothelial cells that have haemogenic potential. Such cells are found in the intraembryonic aorta-gonad-mesonephros (AGM) region (Figure 1.2). Haemogenic endothelium is termed the endothelial to haematopoietic transition (EHT) that sit on the inner lining of the dorsal aorta (DA) and this has been conserved across mammalian and zebrafish. The HSCs later migrate to the fetal liver and then to the bone marrow. The adult HSCs reside in the bone marrow are multipotent and self-renewable in the mouse. Whereas, the HSCs later migrate to the caudal hematopoietic tissue (CHT) and

then to the kidney in zebrafish (it is equivalent in the mouse, the fetal liver and bone morrow, respectively) (Figure 1.2). (Cumano and Godin, 2007, Jagannathan-Bogdan and Zon, 2013, Hoffbrand et al., 2010, Frame et al., 2016, Chen et al., 2011, Ciau-Uitz et al., 2014, Medvinsky et al., 2011, Orkin and Zon, 2008, Paik and Zon, 2010, McGrath et al., 2011, Palis et al., 1999).

In this chapter, it will focus on the different waves of haematopoiesis, the different haematopoietic sites and the genetic mechanisms that underlie the formation of HSCs in the mouse and in the zebrafish.

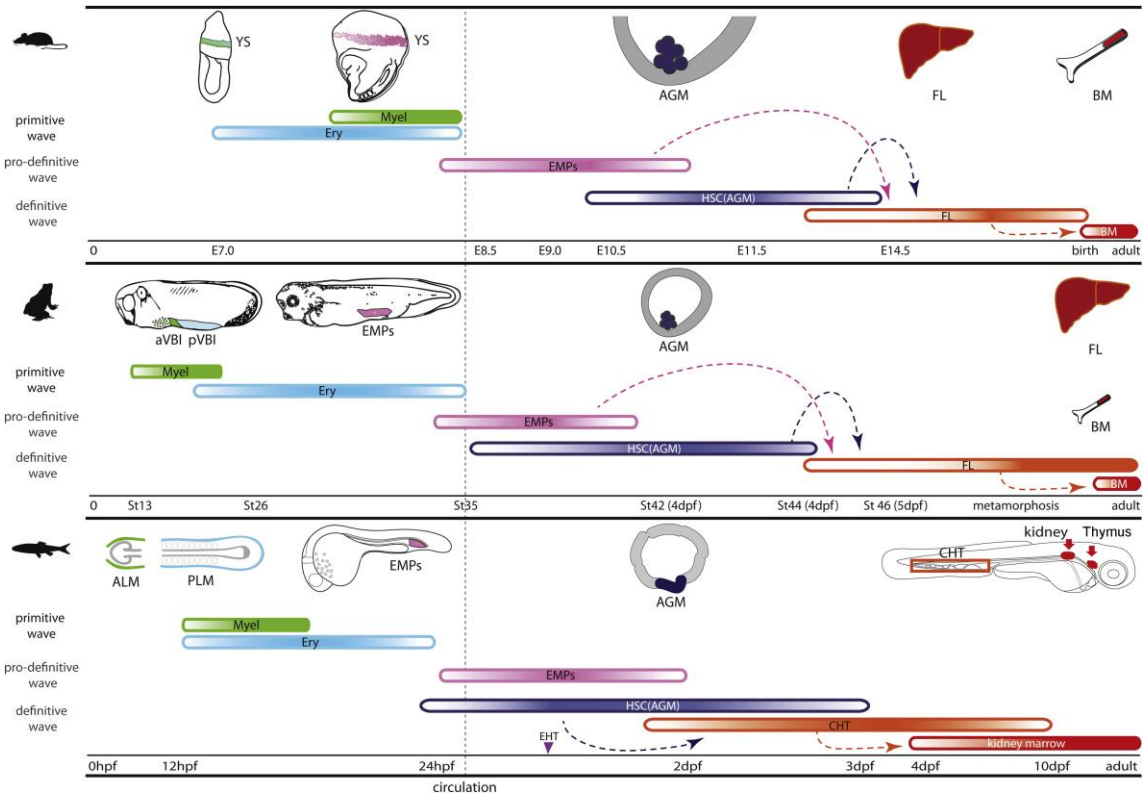


Figure 1.2 Conserved of haematopoiesis development

The haematopoiesis occurs in different waves arising from distinct anatomical sites during embryo development in mouse, Xenopus and zebrafish. Shaded bars represent the timing of haematopoietic activity. Broken arrows migration and seeding of haematopoietic sites. YS (yolk sac), aVBI (anterior ventral blood islands), pVBI (posterior ventral blood islands), ALM (anterior lateral mesoderm), PLM (posterior lateral mesoderm), PBI (posterior blood islands), FL (fetal liver), CHT (caudal hematopoietic tissue), AGM (aorta/gonad/mesonephros), BM (bone marrow), Ery (erythroid, light blue), Myel (myeloid, green), EMPs (erythromyeloid progenitors, pink) (Ciau-Uitz et al., 2014).

1.1.1 The primitive wave of murine haematopoiesis give rise to haemangioblast

The primitive wave is transitory in mouse embryos occurs after embryonic day (E) 7.0 generating cells from extra-embryonically mesoderm ingresses through the posterior primitive streak and differentiate in the blood islands of the yolk sac (Silver and Palis, 1997, Moore and Metcalf, 1970) (Figure 1.3). The extraembryonic sites develop haemangioblasts, mesenchymal cells that are bipotent progenitors for blood and endothelial cells (ECs). The existence of the haemangioblast was indicated by the finding that these cells express genes associated with haematopoietic such as *Scl*, *Gata1* and *βH1 globin*, as well as genes involved with vascular development including *Flk-1* and *VE-cadherin* (Huber et al., 2004). SCL null mice leads to absences of sac haematopoiesis (Robb et al., 1995) and mice lacking Flk-1, fail to generate both vasculature and haematopoietic cells (Shalaby et al., 1995). Between E8.0 and E9.0, haemangioblast starts to differentiate into endothelial cells or primitive haematopoietic cells. The haematopoietic cells are believed to give rise to erythrocytes, macrophage and megakaryocytes. These progenitors are not pluripotent and do not have self-renewal capability (Palis and Yoder, 2001, Tober et al., 2007). Gata1 and Pu.1 transcription factors primarily regulate the formation of primitive erythroid and myeloid fates, respectively (Cantor and Orkin, 2002). At this stage, coinciding with the onset of the heartbeat, the immature nucleated erythroid cells get into embryonic circulation (Ji et al., 2003). Between 12.5-16.5 embryo ages, these primitive erythroid cells become more mature and eventually enucleate. Their main role is to supply oxygen to the early embryonic tissues (Kingsley et al., 2004). At E9.5, megakaryocyte markers were observed to be expressed in cells of the yolk sac. These megakaryocytes release primitive platelets into the bloodstream at

around E11. These primitive platelets are larger in size compared to definitive ones (Tober et al., 2007). In addition to these two progenitor types, immature macrophages have also been observed in the E9 yolk sac (Takahashi et al., 1989).

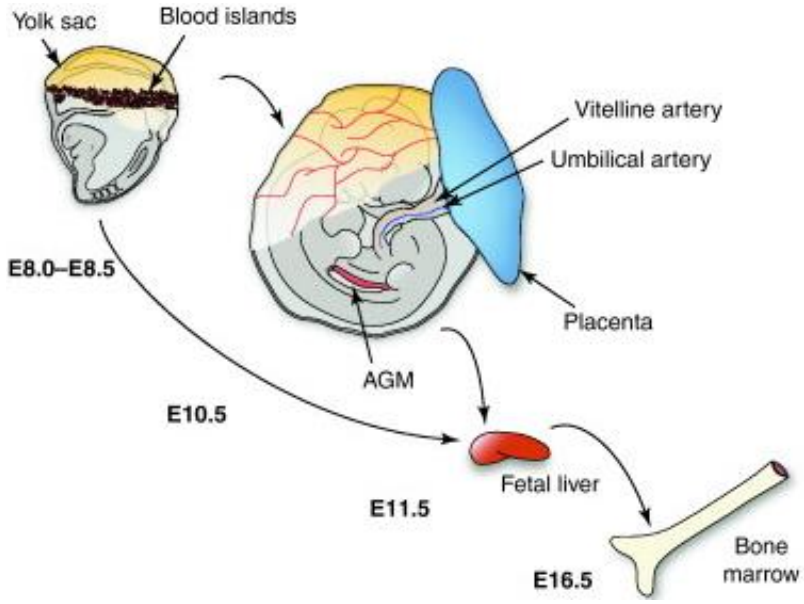


Figure 1.3 Haematopoietic sites development in murine embryos

The primitive haematopoietic progenitors are found extra-embryonically at E8.0 in the yolk sac blood islands. After E9.5, pre-definitive yolk sac progenitors (EMPs) rise and migrate into the fetal liver where blood progenitors expand and/or mature. Finally, at E16.5, the bone marrow remains the main hematopoietic niche throughout adult life (Costa et al., 2012).

1.1.2 Erythroid-myeloid progenitors generation in the yolk sac as a transient wave

The capacity of yolk sac cells to generate blood cell lineages is not limited to primitive erythroid cells, platelets, and macrophages. At E9.5, erythromyeloid progenitors (EMPs) are produced in the yolk sac after producing primitive cells in the transient wave (Palis et al., 1999, Ferkowicz et al., 2003). These cells are believed to be non-definitive haematopoietic cells due to lack of long-term potential; therefore, this wave considered as pre-definitive haematopoiesis (Yamane et al., 2013). This wave consists of red blood cells, platelets, myeloid, and multipotent progenitors (Chen et al., 2011). The EMPs emerge from the YS then migrate to the fetal liver where the differentiation takes place (McGrath et al., 2015) (Figure 1.3). Furthermore, it has revealed the presence of lymphoid lineage potentials that generate T and B cells at the yolk sac cells around E9.5 (Yamane et al., 2009). Boiers and her colleagues, found the cells in E9.5 yolk sacs to express the IL-7 receptors, which are expressed exclusively in lymphoid progenitors (Boiers et al., 2013).

1.1.3 The definitive wave gives rise a long term haematopoietic stem cells in the murine embryos

The definitive haematopoiesis is the third wave of haematopoiesis in the murine embryos. It is the wave that forms long-term haematopoietic cells. These cells emerge at E10.5 to E11 in the intraembryonic site, called the aorta-gonad-mesonephros (AGM) region (Medvinsky and Dzierzak, 1996) (Figure 1.3). The endothelial lining, which is term haemogenic endothelium, of the embryonic ventral wall of the DA is clustered to give raise into HSCs (Medvinsky et al., 2011, Medvinsky and Dzierzak, 1996, Zovein et al., 2010).

This process is known as endothelial to haematopoietic transition (EHT) (Boisset et al., 2010). The HSCs later migrate to the fetal liver and then to the bone marrow (Figure 1.3). The adult HSCs reside in the bone marrow are multipotent and self-renewable (Cumano and Godin, 2007). It has been found that the transcription factor Runx1 is involved in endothelial-hematopoietic transition in the dorsal aorta (North et al., 1999, Yokomizo et al., 2001). Moreover, it has been found that most of the definitive haematopoietic cells are lost in *Runx1* knockout mice (Wang et al., 1996). Although Runx1 protein is first detected during primitive haematopoiesis (E7.5) in mesenchymal cells in the blood islands of the yolk sac in both primitive erythrocytes and endothelial cells (Marshall and Thrasher, 2001), primitive erythropoiesis is normal in Runx1-null mouse embryos (Okuda et al., 1996, Yokomizo et al., 2008). Together, these data indicated that Runx1 has a very specific essential role during definitive haematopoiesis (Chen et al., 2009). Interestingly, the HSC in AGM region can be rescued by overexpression of Runx1 in *Runx1*$^{-/-}$ mouse (Goyama et al., 2004). Notch signalling is required for cell fate specification and pattern formation for dHSCs (Hadland et al., 2004). It has been shown that in *Notch1*$^{-/-}$ mice the lack of contribution of embryonic stem (ES) cells to the dHSCs (Kumano et al., 2003). Beside Runx1 and Notch signalling that plays key regulators of haematopoiesis is the transcription factor Cmyb. It regulates haematopoietic cells differentiation and maintain hematopoietic progenitors. *Cmyb* mutant mice die with failures of erythroid and myeloid development however, primitive stages of blood cell development is dispensable (Mucenski et al., 1991).

1.1.4 Embryogenesis and haematopoiesis of zebrafish

Over recent years zebrafish has become a model as an alternative to mammalian species for studying such as haematopoiesis. The zebrafish model has several advantages such as high fecundity, external fertilisation, ease to introduce genetic change, and optical transparency through embryonic stages. The embryo's transparency is ideal for microscopic imaging of life processes. These features have led to fundamental new insights into the development and differentiation of hematopoietic cells (Gore et al., 2018). As in others vertebrates, during zebrafish embryogenesis, haematopoiesis divides into three waves from mesodermal germ layer with shifting sites; primitive, transient and definitive wave, respectively as shown in Figure 1.4 (Ciau-Uitz et al., 2010). Embryogenesis and haematopoiesis and the genetic regulatory mechanisms of HSCs formation in the zebrafish will be explained in this section.

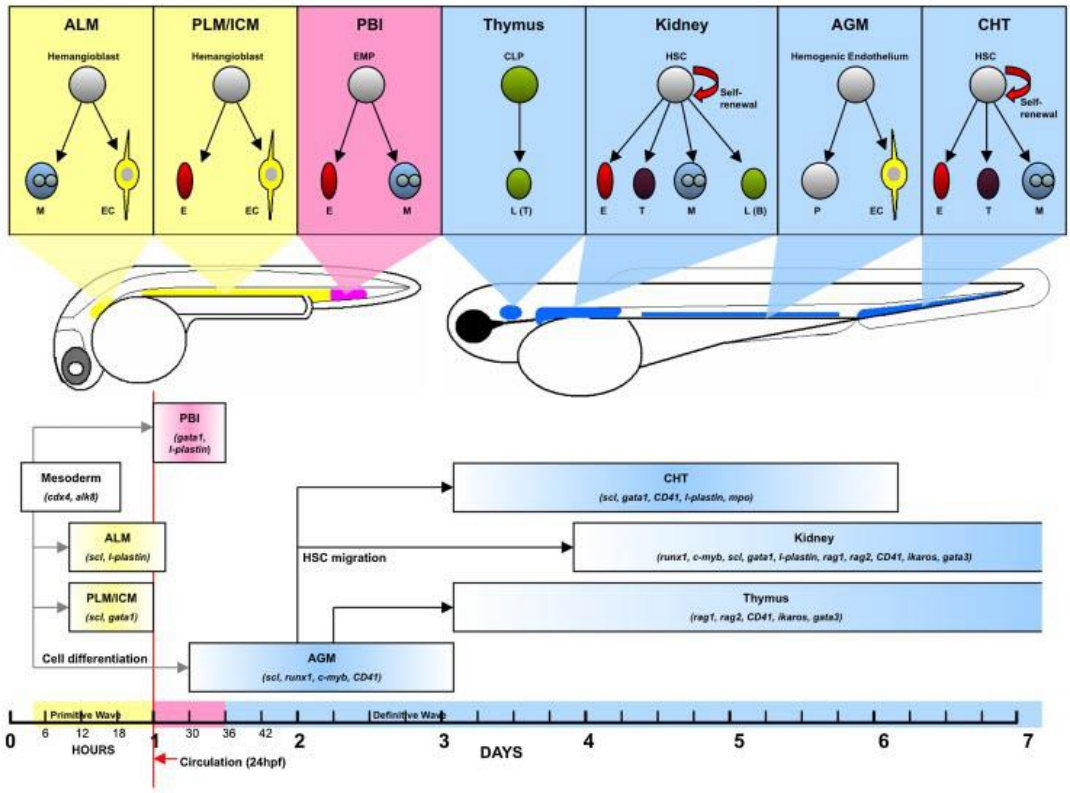


Figure 1.4 Haematopoiesis in zebrafish

The development of blood cells of the primitive wave starts at first 12-22 hpf when prRBCs progenitors emerge in the posterior lateral mesoderm (PLM) and the macrophages and neutrophils progenitors initiate in anterior lateral mesoderm (ALM). After circulation starts at 24 hpf, haematopoiesis shifts temporarily to the posterior blood island (PBI). The definitive wave begins in the aorta-gonad-mesonephros (AGM) at 30 hpf, where haemogenic endothelial (HE) cells transition occurs. By day 3pf haematopoietic stem cells (HSCs) migrates to caudal haematopoietic tissue (CHT) and thymus to differentiation. By day 4 pf migrate to the kidney (Chen and Zon, 2009).

1.1.5 The embryogenesis and primitive haematopoiesis wave of zebrafish

At 10-hour post-fertilisation (hpf) the primitive wave of zebrafish haematopoiesis is started. It is believed that blood cells and endothelial cells (ECs) arise from mesenchymal bipotent haemangioblasts in intraembryonic sites of the lateral mesoderm (Figure 1.5A) (Vogeli et al., 2006). Both the endothelial and blood lineages express several genes may in fact function at the level of the haemangioblast. The co-expression of *scl*, *gata-2* and *flk-1* within the same population of cells at this stage suggests that these cells are bipotent and not specific yet (Gering et al., 1998). The *Scl* gene is expressed in blood and endothelium (Green et al., 1992, Drake et al., 1997). The scl is a transcription factor that is required for the development of the hematopoietic lineage (Porcher et al., 1996). *Gata-2* is expressed in an early haematopoietic progenitor (Detrich et al., 1995). The *Flk-1* is a receptor tyrosine kinase whose expression is restricted to endothelial cells and play critical roles in vascular development (Liao et al., 1997). In zebrafish, *scl* is first detected at 11 hpf in the anterior lateral plate mesoderm (ALM) and posterior lateral plate mesoderm (PLM), regions which give rise to both endothelial and blood cells (Figure 1.4) and (Figure 1.5A). The co-expression of *scl* and *flk-1* were indicated at 12 hpf in the anterior and posterior lateral plate mesoderm by *in situ* hybridization in zebrafish. In addition, at the same stage and sites the co-expression of *scl* and *gata-2* were detected (Gering et al., 1998). A recent study shows that transcription factor npas4l acts in early to regulate the endothelial and haematopoietic progenitor differentiation. *npas4l* gene that when mutated in zebrafish leads to the loss of most endothelial and haematopoietic (Reischauer et al., 2016).

The anterior lateral mesoderm (ALM) where the primitive myeloid and angioblasts (ABs) are born with the expression of *pu.1* and *flk1/kdrl*, respectively. Pu.1 is transcription factor essential for normal B-lymphoid and macrophage development (Bennett et al., 2001). On the other hand, the posterior lateral mesoderm (PLM) where the primitive erythroid progenitors that arise to red blood cells (prRBCs) and angioblasts are born with expression zinc finger transcription factor gata1. It is a master regulator in erythrocyte development (Paik and Zon, 2010). Later by 18 hpf, the PLM cells migrate to the midline between the notochord, the endoderm and the somites to form the intermediate cell mass (ICM) (Figure 1.5B) (Aladhami and Kunz, 1977). The ICM forms which are the equivalent site of the extraembryonic yolk sac blood islands in mammals. As soon as the ICM is formed, the PLM-derived ABs differentiate into ECs that form the major blood vessels of the trunk, the dorsal aorta (DA) and the axial vein, also called the posterior cardinal vein (PCV). The ALM ABs give rise to the head vessels and differentiate into the inner epithelial cell layer of the heart, the endocardium. Posteriorly, the primitive erythroid progenitors then differentiate into primitive red blood. The circulation of primitive erythroid cells starts at ~24 hpf (Gering et al., 1998, Liao et al., 1998, Vogeli et al., 2006, Chen and Zon, 2009, Paik and Zon, 2010, Ciau-Uitz et al., 2014).

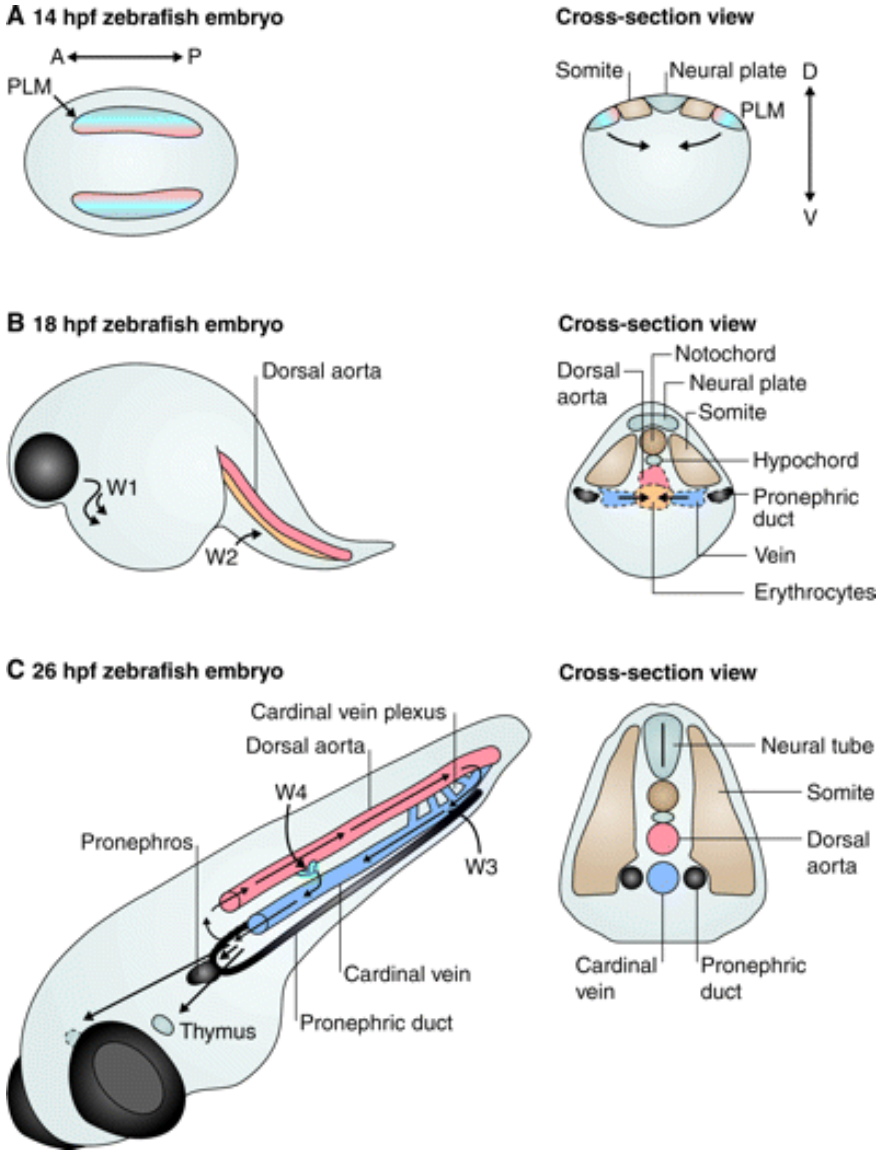


Figure 1.5 Haematopoiesis sites of zebrafish

Zebrafish hematopoiesis occurs in different sites during development. **(A)** zebrafish presented in above and in cross-section, left and right respectively, showing lateral stripes of PLM (the curved arrows). **(B)** Side viewed with dorsal uppermost and in cross-section view of embryo with the two black arrows migration of the PLM. Orange shading: red blood cells, shown in pink: angioblast cord. **(C)** Embryo HSCs from the floor of the dorsal aorta (left, curved arrow w4) to PBI which later on called CHT with cross-section view shown on the right. Dorsal (D), Ventral (V), Anterior (A) and posterior (P) (Medvinsky et al., 2011).

1.1.6 Erythroid myeloid progenitors are raised in the posterior blood island (PBI) as an intermediate wave of zebrafish

The second wave of zebrafish haematopoiesis is a transient (intermediate) wave that occurs in the PBI, which has also been described as the beginning of definitive wave since progenitors form to rise to multiple blood lineages. Haematopoiesis shifts to the posterior blood island (PBI) after 24 hpf for a brief period (Bertrand et al., 2007) (Figure 1.4) and (Figure 1.5 C). At that region, erythromyeloid progenitors (EMPs) have erythroid and myeloid differentiation potential but lack lymphoid potential. The EMPs arise before HSCs were detected in the aorta-gonad-mesonephros (AGM). The EMPs serve as transient progenitors that are born independently of HSCs to initiate definitive haematopoiesis in the developing zebrafish embryo (Bertrand et al., 2007, Sood and Liu, 2012). However, EMPs were not observed to populate the kidney marrow (KM) or thymus (Carroll and North, 2014).

1.1.7 vDA of zebrafish embryos generate long-term haematopoietic stem cells

At around 30 hpf, the definitive wave starts. A subset of endothelial cells known as haemogenic endothelium cells (HECs), undergo the transition from spindle-shaped endothelial cells lining the ventral wall of the dorsal aorta (vDA) to mesenchymal round shaped cells as they turn into haematopoietic cells by a process known as (EHT); they enter the blood circulation via the axial vein (Bertrand et al., 2010a). Emerging definitive blood cells enter the space between the DA and the PCV and do not form clusters of cells in the lumen of the DA. The HECs are the lineages of a common precursor definitive

haematopoietic cells (dHSCs) emerging from the ventral wall of the dorsal aorta in (AGM) equivalent to other vertebrates (Figure 1.5 C) (Chen et al., 2011, Gordon-Keylock et al., 2013). The population of cells between the DA and the PCV are expressed the transcription factor *runx1*, which is associated with definitive haematopoiesis (Kalev-Zylinska et al., 2002, Burns et al., 2002, Gering and Patient, 2005). Researchers found that the definitive haematopoietic cells absence completely in Runx1 gene deletion (Kalev-Zylinska et al., 2002). While the *runx1* is expressed in the ventral wall of the dorsal aorta, the transcription factor *cmyb* is expressed at around 36 hpf. It has been found that loss of Runx1 in Runx1 morphants causes loss of *cmyb* expressing definitive progenitors (Gering and Patient, 2005). Studies demonstrated that *runx1* and *cmyb* expression are main regulators of HSC development in the AGM region of other vertebrates too (Bertrand et al., 2010).

Regulation of HSC formation also occurs via cell-cell communication. Notch signalling that involves direct interaction between neighbouring cells also plays an important role in HSC emergence. The Notch signalling plays a critical role in the regulation of HSCs differentiation, proliferation, survival, and apoptosis (Bray, 2006, Fiuza and Arias, 2007). In 2005, a study by Gering and Patient showed that zebrafish mutants such as the *mindbomb* (*mib*); which is a deficit in the Notch signalling pathway (Itoh et al., 2003) leds to defect in definitive haematopoiesis, without affecting primitive haematopoiesis. Besides, Smoothened (*Smo*) is a transmembrane protein that is essential for transduction of Hedgehog activity, and embryos treated with a VEGF (vascular endothelial growth factor) receptor inhibitor failed to express *runx1* in the vDA and did not develop definitive haematopoietic cells, showing that that Notch, Hedgehog and Vegf signalling are required for definitive haematopoiesis, but not for primitive.

These HSCs later migrate to transient larval site the caudal hematopoietic tissue (CHT) in the ventral region of the tail. The CHT is where the haematopoietic stem and progenitor cells (HSPCs) occupy perivascular niches (Tamplin et al., 2015). At day three post fertilisation, the T cells progenitors migrate to the thymus where they differentiate into T-lymphocytes, and one day later, the HSCs migrate to the kidney marrow. The kidney in the zebrafish is responsible for the maintaining of adult HSCs. Its role is analogous to the bone marrow in mammals (Figure 1.2) (Gering and Patient, 2005, Chen and Zon, 2009, Paik and Zon, 2010, Martin et al., 2011, Medvinsky et al., 2011, Kulkeaw and Sugiyama, 2012, Swiers et al., 2013, Jagannathan-Bogdan and Zon, 2013).

1.2 The zebrafish gene trap line *qmc551* eliminates primitive and definitive haematopoietic gfi1aa expression without interfering with normal haematopoiesis

Gfi1 and Gfi1b play important overlapping roles in adult HSCs (Möröy et al., 2015). *Gfi1b*$^{-/-}$ mice die during gestation, due to the failure to produce red blood cells and platelets (Saleque et al., 2002). In contrast, *Gfi1*$^{-/-}$ mice show significant reductions in neutrophils and inner ear defects (Karsunky et al., 2002, Hock et al., 2003, Wallis et al., 2003). Zebrafish has three *gfi1/1b* paralogues. *Gfi1aa* and *gfi1ab* are orthologues of mammalian *Gfi1* (Dufourcq et al., 2004, Wei et al., 2008), whereas *gfi1b* corresponds to mouse *Gfi1b* (Cooney et al., 2013) (Figure 1.6).

Gfi1 and Gfi1b are transcriptional repressors, with a conserved *N*-terminal domain, called the SNAG domain, a variant linker, and a conserved C-terminal Zn-finger DNA binding domain (Zweidler-Mckay et al., 1996, Grimes et al., 1996, Tong et al., 1998). Gfi1 and Gfi1b are both involved in many cell biological processes, such the inhibition or promotion of cell proliferation, influencing cell fate, terminal differentiation, and the inhibition of apoptosis. The zebrafish line *qmc551* carries a gene trap transposon in intron 1 of *gfi1aa*, a zebrafish paralogue that encodes a homologous transcriptional repressor to the mammalian Gfi1 and Gfi1b proteins. The transposon interferes with normal *gfi1aa* expression, and instead, a gfp reporter gene is expressed under the control of the endogenous *gfi1aa* promoter, in transient primitive erythrocytes, erythromyeloid progenitors, and definitive blood cells. In *qmc551* homozygous zebrafish, normal primitive and definitive haematopoiesis can be observed, due to the functional redundancy of *gfi1b*, whereas *gfi1ab* expression remains unaffected. At day 3 pf,

qmc551 homozygous zebrafish embryos injected with *Gfi1b* morpholinos showed a dramatic reduction in DA displayed and the morphology of prRBCs appeared larger, with less condensed nuclei. These results suggested that in the absence of Gfi1aa, Gfi1b is sufficient to rescue primitive erythropoiesis.

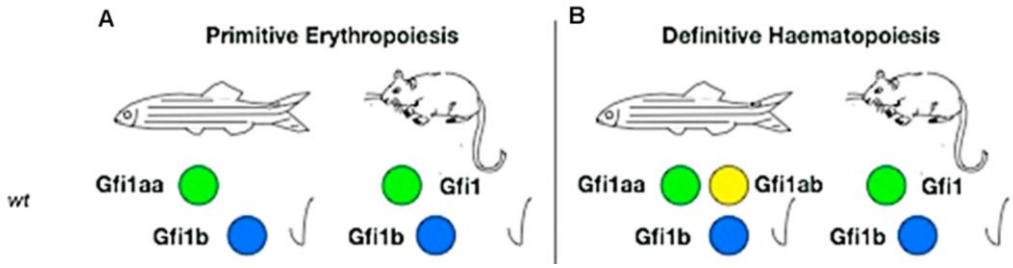


Figure 1.6 Expression of mouse Gfi1/1b and zebrafish gfi1/1b/ab during primitive erythropoiesis and definitive haematopoiesis, in wild-type embryos (Thambyrajah et al., 2016)

1.3 Gfi1aa and Gfi1b promote late erythroid gene expression by suppress endothelial gene expression in immature primitive red blood cells

The *qmc554* line is a *gfi1b* mutant allele, created in our lab by Dr Chris Moore. The *gfi1b*^{qmc554} allele a CRISPR/Cas9-induced deletion, results in a frameshift mutation in exon 4. Moore and his colleagues (2018) showed that in zebrafish, Gfi1aa and Gfi1b suppress endothelial gene expression and promote late erythroid gene expression in immature primitive erythroblasts. Therefore, mutations in both genes result in zebrafish with the delayed differentiation of primitive erythroblasts from mesenchymal haemangioblasts, although this differentiation is not blocked entirely. At early erythroid progenitors the Gfi1aa suppresses endothelial gene expression. These suggested that the early expression of Gfi1aa in prRBC progenitors is to suppress the EC gene regulatory network. In its absence, the network transiently commands the expression of endothelial markers. The loss of EC gene expression revealed the presence of additional mechanisms that can suppress EC gene expression in maturing prRBCs. The loss of both *gfi1aa* and *gfi1b* gene copies leads to progressive delays in red blood cells maturation, but without any defects in definitive haematopoiesis, this indicate that either loss Gfi1b alone or in combination with Gfi1aa, does not affect with the onset of definitive haematopoiesis in zebrafish (Figure 1.7). However, Gfi1b alone is essential for adult erythropoiesis. RNA sequencing, performed in embryos 20 hpf showed that prRBCs displayed reduced expression levels of the late red blood cell transcription gene (*slc4a1a*) and elevated levels of early red blood cell (*zfp36l2*) and endothelial genes (*etv2*) when Gfi1aa/1b is depleted (Moore et al., 2018).

1.4 Gfi1ab is upregulated in the vDA in the absence of *gfi1aa*

The zebrafish *gfi1ab* is a paralogue of *gfi1aa*. In the absence of *gif1aa* expression, g*fi1ab* is upregulated in the vDA, suggesting that *gfi1aa* and *gfi1ab* play redundant roles during definitive haematopoiesis (Ucanok, previous PhD student). Thus, a double mutant was generated in our lab, by introducing a *gfi1ab* mutant allele into the *qmc551* line, to generate the *qmc553* zebrafish line (Dr Moore and Mariam Jalali, a previous PhD student), to gather further evidence for the functional redundancy of these genes and to explore the function of *gfi1a* during HEC and HSC formation. The *qmc553* allele effectively fused exons 2 and 3 of zebrafish chromosome 6. The mutation shifts the *gfi1ab* reading frame, introducing a premature stop codon. In addition, *qmc553* zebrafish encode truncated Gfi1ab proteins that retain the SNAG repressor domain but lack the DNA binding domain. These truncated proteins are unlikely to be capable of fulfilling those Gfi1ab functions that require active DNA binding. Our previous PhD student (Mariam Jalali) found that the *qmc551/qmc553* double mutant displayed a reduced number of qmc551:eGFP-positive definitive blood cells in the CHT. These results are similar to those reported for the *qmc551/sh320* double mutant. The *sh320* line, created by Dr. Robert Wilkinson's lab in Sheffield, expresses a *gfi1ab* mutant allele, consisting of a 4 bp sequence insertion into exon 2 of the *gfi1ab* locus on chromosome 6, which results in a frameshift in the *gfi1ab* coding sequence. The *qmc551/sh320* double mutant showed a significant reduction in cmyb-positive cells in the CHT, thymus, and kidney at 5 dpf (Figure 1.7).

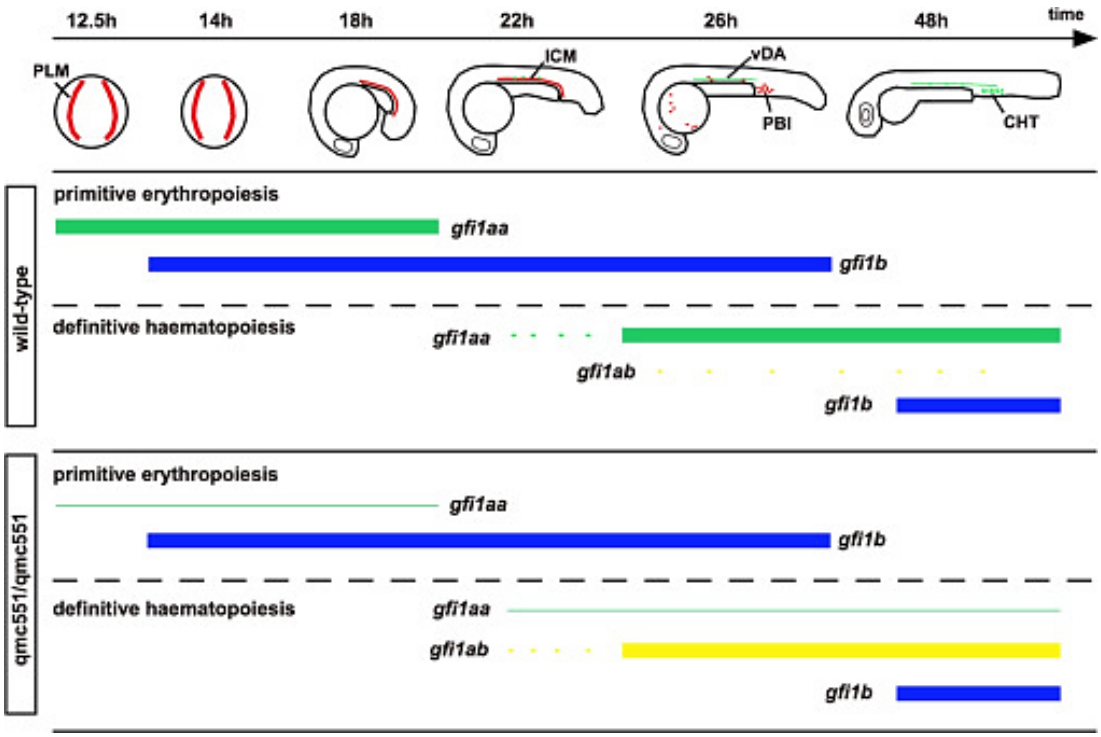


Figure 1.7 *gfi1/1b/ab* expression during primitive erythropoiesis and definitive haematopoiesis, in wt and homozygous *qmc551* embryos

The images show posterior and lateral views of early and late embryos, respectively. Red: primitive erythropoiesis sites; green: definitive haematopoiesis. Expression of *gfi1* paralogs in wt and *qmc551* homozygous embryos. Solid boxes represent strong expression. Dots show scattered expression. Lines represent the apparent loss of expression (Thambyrajah et al., 2016).

1.5 RNA-seq on qmc551:GFP+ cells identifies two lncRNA genes that are specifically expressed in embryonic blood cells

In an attempt to identify new genes involved in haematopoiesis, a transposon-mediated gene trap screen was performed in zebrafish. This screen identified a zebrafish line, called *qmc551*, that expressed the gene trap's GFP reporter in primitive red blood cells and in haemogenic endothelial cells that underwent endothelial to mesenchymal transition as they turned into definitive blood cells. Genomic analyses revealed that the line's gene trap transposon was inserted in intron 1 of the gene for the transcriptional repressor Gfi1aa and that its gfp reporter was expressed under the control of the endogenous *gfi1aa* promoter (Thambyrajah et al., 2016) (Figure 1.8). This line was subsequently used as a source of GFP+ blood cells. To find out which genes were expressed in developing blood cells, a former post-doc in the lab, Dr Chris Moore, performed a transcriptome analysis on the GFP-positive and GFP-negative cells that he isolated from 30 hours post fertilisation (hpf) embryos. The cell populations were sorted to GFP+ and GFP- using fluorescence-activated cell sorting (FACS). The RNA was extracted and prepared for Next Generation Sequencing. The sequence reads were mapped to the zebrafish reference genome. Reads mapping to individual genes were counted and counts were normalised to the size of the transcript and the size of the library (RPKM -reads per kilobase of transcript and per million reads within the library) (Mortazavi et al., 2008) (Figure 1.9). Differential gene expression analysis was performed to identify genes that are more highly expressed in GFP+ than in GFP- cells. Among them, there were a large number of genes that are known to be expressed in primitive red blood cells (Table 1.1), reflecting the fact that the vast majority of the qmc551:GFP+ cells are prRBCs at 30 hpf. In addition, the analysis also detected

genes that are associated with definitive haematopoiesis like *gfi1aa*, *runx1* and *ikfz1*. Their expression level was much lower, consistent with the relative under-representation of these cells in the GFP+ cell population. When the list of differentially expressed genes was searched for those with high expression but unknown function two long-noncoding RNA genes attracted our attention. They are known under the accession numbers si:ch211-214p16.1 (NR_110_414) and si:ch211-214p16.2 (NR_131_174) but will henceforth be referred to as *lncRNA1* and *lncRNA2*, respectively. The RNA-Seq data showed that both were expressed at decent levels in GFP+ cells, with lncRNA2 levels exceeding those of lncRNA1 by threefold (Table 1.1). Their very low RPKM counts in GFP- cells suggested that their expression was strongly associated with haematopoiesis. This idea was validated by whole mount in situ hybridisation experiments that were performed on 26 hpf embryos. These WISH analysis showed that the two genes were indeed expressed in two sites of zebrafish haematopoiesis, the ventral wall of the dorsal aorta and the posterior blood island, but not in any non-haematopoietic tissue (Figure 1.10).

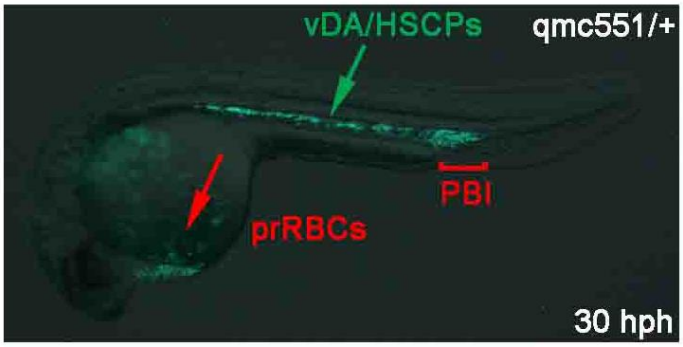


Figure 1.8 30 hpf zebrafish gene trap line qmc551

Lateral view of a fixed qmc551 embryo. The embryo gene trap line qmc551/+ expresses GFP in primitive red blood cells (red arrow) in the yolk sac and in haemogenic endothelial cells of the ventral wall of the dorsal aorta (green arrow) (obtained from Dr Martin Gering).

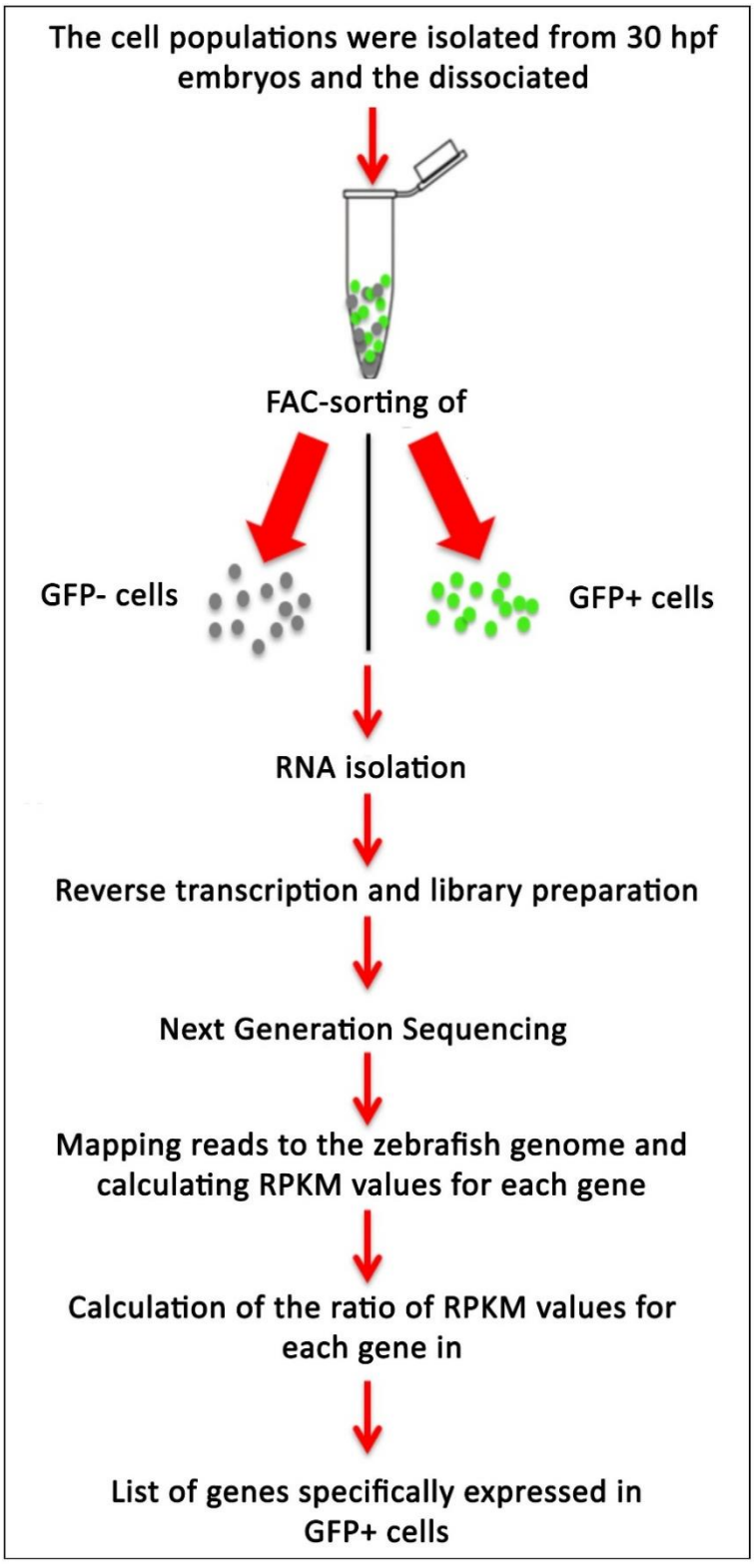


Figure 1.9 Flowchart summarizing the main steps in the RNA-seq procedure for data analysis of differentially expressed genes

Table 1.1 List of the some interesting genes expression in qmc551:GFP+ cells relative to qmc551:GFP- and an RPKM value in qmc551:GFP+ cells.

Gene	GFP-neg	GFP-pos	Ratio
ela2	0.02	65.39	3661.84
hif1al2	0.11	262.05	2457.37
nt5c2l1	0.37	852.21	2317.90
zgc:163057	0.19	413.04	2204.33
slc4a1a	1.70	3587.48	2115.22
alas2	1.44	2745.26	1903.50
klf1	0.05	98.65	1872.52
nmt1b	0.44	764.08	1726.74
gata1a	0.12	178.23	1528.94
rhd	0.13	189.57	1470.86
epb41b	0.29	416.93	1434.66
rfesd	0.15	204.76	1364.48
gfi1b	0.28	341.18	1208.34
rell2	0.12	144.74	1182.89
tfr2	0.17	198.07	1170.70
hbbe3	93.55	102077.58	1091.19
hbbe2	5.60	5900.24	1054.04
hbae3	125.02	128181.63	1025.28
hbae1	249.97	241775.38	967.21
cahz	2.68	2497.53	932.60
epor	0.15	135.12	895.48
slc43a3a	0.07	55.77	781.00
hbbe1.1	67.36	49343.10	732.58
fam65c	0.12	73.61	592.33
si:dkey-261j4.3	0.24	143.49	590.41
si:ch211-5k11.12	38.64	19851.53	513.80
MGC173646	33.28	12985.50	390.25
klf17	0.93	340.92	367.71
ucp3	0.26	90.89	351.08
cyp46a1.1	0.18	57.22	325.18
tfr1a	2.78	888.05	318.99
si:ch211-214p16.2	0.55	169.26	309.65
ikzf1	0.17	52.24	308.16
urod	7.74	2205.85	285.17
blvrb	5.13	1317.95	257.03
blf	2.55	600.90	235.84
si:ch211-214p16.1	0.23	48.45	209.02
ppox	1.15	223.58	194.22
mir144	0.69	121.26	176.69
fech	1.64	266.16	162.36

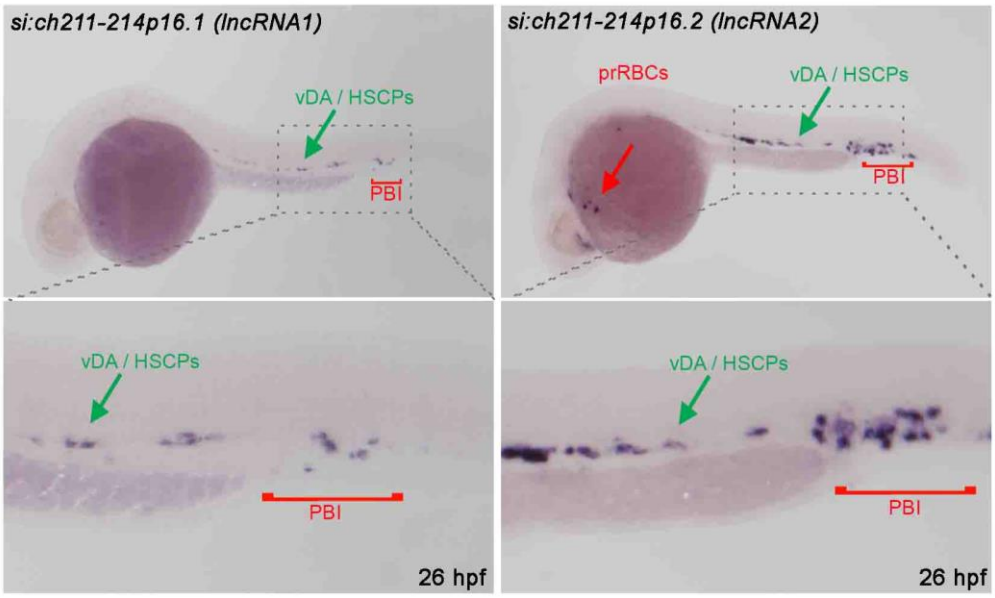


Figure 1.10 lncRNA1 and lncRNA2 are expressed in sites associated with primitive and definitive haematopoiesis of zebrafish

Images of wt embryos stained in a whole in situ hybridisation experiment using probes for lncRNA1 and lncRNA2. All embryos are 26 hpf and presented in lateral views with the anterior to the left and the dorsal side up and close up of the tail region. Red bracket: the PBI, green arrow: endothelial cells of the vDA and HSCPs (Moore, unpublished data).

1.5.1 lncRNA genes play essential regulatory roles in numerous cells types, including haematopoietic cells

1.5.1.1 long non-coding RNAs

The functional understanding of the genome has changed dramatically recently. The discovery of non-coding RNAs (ncRNAs) have change the traditional definition of genes and a functional region of the genome. High-throughput transcriptomic analyses have revealed that only 1-2% of the transcripts encode for proteins, whereas the vast majority are transcribed as ncRNAs (Consortium, 2004, Consortium et al., 2007). NcRNAs are functional molecules that are transcribed but not translated into proteins that expressed at low levels, often in a cell/tissue-specific manner or in a developmental stage-specific manner (Ward et al., 2015). NcRNAs are divided into two major groups based on their size namely either short ncRNAs (sncRNA) or long ncRNAs (lncRNAs). This category includes either housekeeping or regulatory ncRNAs. Constitutively expressed housekeeping ncRNAs are ribosomal (rRNA), transfer (tRNA), small nuclear (snRNA), and small nucleolar RNAs (snoRNAs). Whereas, the regulatory ncRNAs are microRNAs (miRNAs), small interfering RNAs (siRNAs), Piwi-associated RNAs (piRNAs) long noncoding RNAs (lncRNAs) promoter-associated RNAs (PARs) and enhancer RNAs (eRNAs) (Cech and Steitz, 2014). This section focuses on long ncRNAs (lncRNAs) and how they involved in processes such as development and pathologies.

The ncRNAs longer than 200 nucleotides are called long ncRNAs (lncRNAs) (Pauli et al., 2012). They have accomplished a remarkable of various functions in biological processes. There are several subtypes of lncRNAs, and these can be classified according to their different four major features, including location and context, effects exerted on DNA sequences, functioning mechanism and targeting

mechanism (Ma et al., 2013, Pauli et al., 2012, Ma et al., 2012, Hu et al., 2011, Wilusz et al., 2009). These long non-coding RNAs have been discovered to play essential roles in regulating developmental processes. They regulate gene expression at the levels of modifying the epigenetic status of chromatin, transcription, RNA processing, and translation besides apoptosis (Figure 1.11) (Ma et al., 2013, Wilusz et al., 2009). Despite numerous instances illustrating the functional roles of lncRNAs, great majority, the mechanisms of action remain unknown. However, the biochemical and molecular characterization details of lncRNAs will be fundamental for the understanding of main cell biology, evolution, health, and disease.

In the 1990s, some of the first examples of gene-specific regulatory roles for long noncoding RNAs were uncovered, include the discovery of lncRNAs involved in epigenetic regulation, such as H19 (Brannan et al., 1990). H19 is an imprinted gene that was identified as being maternally expressed in mice (Bartolomei et al., 1991). H19 resembles mRNAs because it is transcribed by RNA polymerase II, spliced, 3’ polyadenylated, and localized to the cytoplasm. However, H19 is not translated, despite containing small open reading frames. The H19 gene sequence is highly conserved across mammals (Brannan et al., 1990). The H19 promoter is methylated on the paternal chromosome, inhibiting the transcription of paternal H19 (Srivastava et al., 2000).

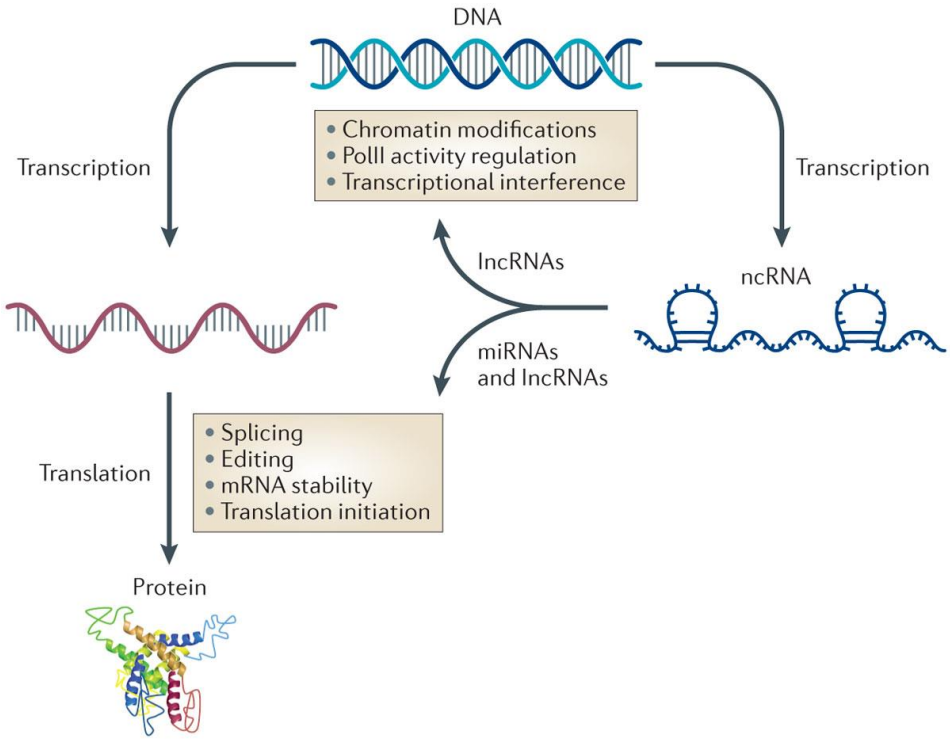


Figure 1.11 Long non-coding RNAs have the capacity to regulate gene expression at variant levels (Wahlestedt, 2013)

1.5.1.2 lncRNA genes play important roles during the regulation of blood cells development

LncRNAs play significant roles in the control of gene expression, by regulating transcriptional and translational processes. Moreover, many studies have demonstrated that lncRNAs are involved in a variety of biological processes, such as stem cell pluripotency and reprogramming, X-chromosome inactivation, and DNA methylation (Pauli et al., 2012, Ma et al., 2012, Hu et al., 2011, Ma et al., 2013, Wilusz et al., 2009). In 2008, Ebralidze et al. demonstrated that antisense lncRNA controls the expression of the transcription factor PU.1, which is required for the development of B lymphocytes and macrophages during haematopoiesis. They demonstrated that antisense lncRNA controls the level of PU.1 protein by interfering with mRNA translation (Ebralidze et al., 2008). Other researchers have presented evidence that lncRNAs are relevant during haematopoiesis (Hu et al., 2011). They analysed the expression of long intergenic non-coding RNA (lincRNA) during mouse erythropoiesis, and identified a lincRNA involved in erythroid differentiation that promotes apoptosis, which they named lincRNA-EPS (erythroid prosurvival) (Hu et al., 2011). In this study, they knocked down lincRNA-EPS genes in mouse fetal liver progenitors, which blocked the differentiation of erythroid progenitors (Hu et al., 2011). Moreover, researchers have demonstrated that lncRNAs are expressed during red blood cell maturation (Alvarez-Dominguez et al., 2014). They isolated lncRNAs from mouse fetal liver $TER119^{+}$ cells and uncovered a novel erythroid-specific lncRNA in the nucleus that showed striking developmental-stage expression specificity and is conserved in humans. In addition, the analysis showed the lncRNAs are required for red blood cells maturation in the fetal liver but are missing in adult bone marrow. In 2014, Paralkar et al. found that lncRNAs were able to regulate erythrocytosis. They

demonstrated a novel, cell-specific lncRNA that was expressed during mouse erythro-megakaryopoiesis, transcribed from conventional gene promoters, regulated by critical haematopoietic transcription factors, and conserved between different mouse strains (Paralkar et al., 2014).

A recent study by Luo et al. (2015) suggested that lncRNAs may be involved in differentiation and self-renewal in the hematopoietic system. Some lncRNAs are specifically expressed in differentiated HSC-derived lineages, due to the DNA methylation-mediated regulation of gene expression (Luo et al., 2015). They discussed two lncRNAs, named LncHSC-1 and LncHSC-2, identified in mice. LncHSC-1 regulates the differentiation of myeloid cells, whereas LncHSC-2 is involved in T cell differentiation and HSC self-renewal. In addition, LncHSC-2 enrichment was identified at key hematopoietic transcription factor binding sites, including those for E2A, which plays an important role in haematopoietic differentiation. To examine the roles played by these two genes, Luo et al. generated LncHSC-1 and LncHSC-2 knockdown mice and analysed their distinct effects on HSC self-renewal and differentiation (Luo et al., 2015). Studies suggest that lncRNAs are more lineage-restricted than coding mRNAs. Although they are expressed at lower levels than mRNAs, lncRNAs are expressed in a more cell-type-specific manner (Alvarez-Dominguez et al., 2014, Paralkar et al., 2014).

1.5.1.3 lncRNA in diseases and disorders

Because lncRNAs are expressed in precisely regulated developmental patterns, the misregulation or mutation of lncRNAs may be associated with the development of diseases or disorders. Interestingly, genome-wide association studies have revealed that most disorders are located outside of protein-coding genes (Batista and Chang, 2013). Runt-related transcription factor (RUNX1), also known as acute myelogenous leukaemia 1 (AML1) (add them on abbreviations list), regulates haematopoiesis by modulating the expression of other hematopoietic genes. The t(8;21) chromosomal translocation, which involves *RUNX1,* causes acute myeloid leukaemia. A novel intragenic lncRNA, which is transcribed by an upstream promoter and overlaps with *RUNX1,* called RUNX1 overlapping RNA *(RUNXOR),* has been found to be upregulated in AML. Quantitative real-time PCR analysis performed on blood samples from AML patients found that *RUNXOR* was upregulated in AML patients compared with normal control patients (Figure 1.12) (Wang et al., 2014).

Moreover, specific interactions between the 3'-end of *RUNXOR* and several regions in the *RUNX1* locus have been identified, particularly those sites near the *RUNX1* promoters, P1 and P2. In addition, an interaction was detected between chromatin DNA and *RUNX1* enhancer1 and enhancer2. Together, these data suggest that *RUNXOR* may be physically involved in the chromosomal translocations that occur during malignancies (Wang et al., 2014).

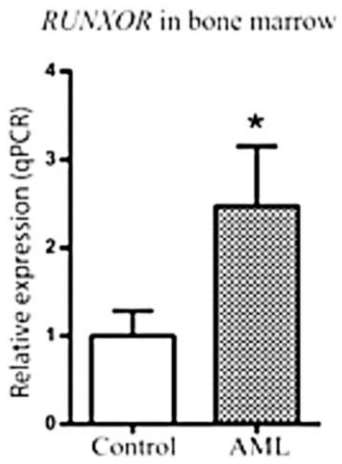


Figure 1.12 Expression of RUNXOR in AML

" Upregulated *RUNXOR* lncRNA in AML patients' bone marrow. The mRNA levels were quantitated by qPCR. For comparison, the normal control was arbitrarily set as 1. All data shown are mean±SD from three independent experiments. *$p < 0.05$ as compared with the normal controls'' (Wang et al., 2014).

1.6 The CRISPR/Cas9 system as a tool for lncRNA genes deletion

To evaluate the role and to learn more about the function of *lncRNA* genes during haematopoiesis, we decided to generate a both *lncRNA1* and *lncRNA2* mutants using the CRISPR/Cas9 system.

The **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats /**C**RISPR **a**ssociated **s**ystem (CRISPR/Cas) constitutes a prokaryotic, adaptable immune system that protects against foreign nucleic acids, such as viruses. As part of the adaptation process, short repetitions of base sequences are derived from the DNA fragments of invading foreign organisms, called protospacers, which are incorporated into the CRISPR locus to build a CRISPR array. This repeat spacer array serves as a type of immune memory. During the re-expression process, the CRISPR array is transcribed and processed into small interfering CRISPR RNAs (crRNAs). These transcripts direct Cas endonucleases to a specific site of an exogenous genome near the species-specific protospacer adjacent motif (PAM). The CRISPR/Cas complex cleaves and degrades the complementary targeted DNA, resulting in a double-strand break (DSB) (Figure 1.13) (Barrangou, 2015, Wiedenheft et al., 2012, Sander and Joung, 2014).

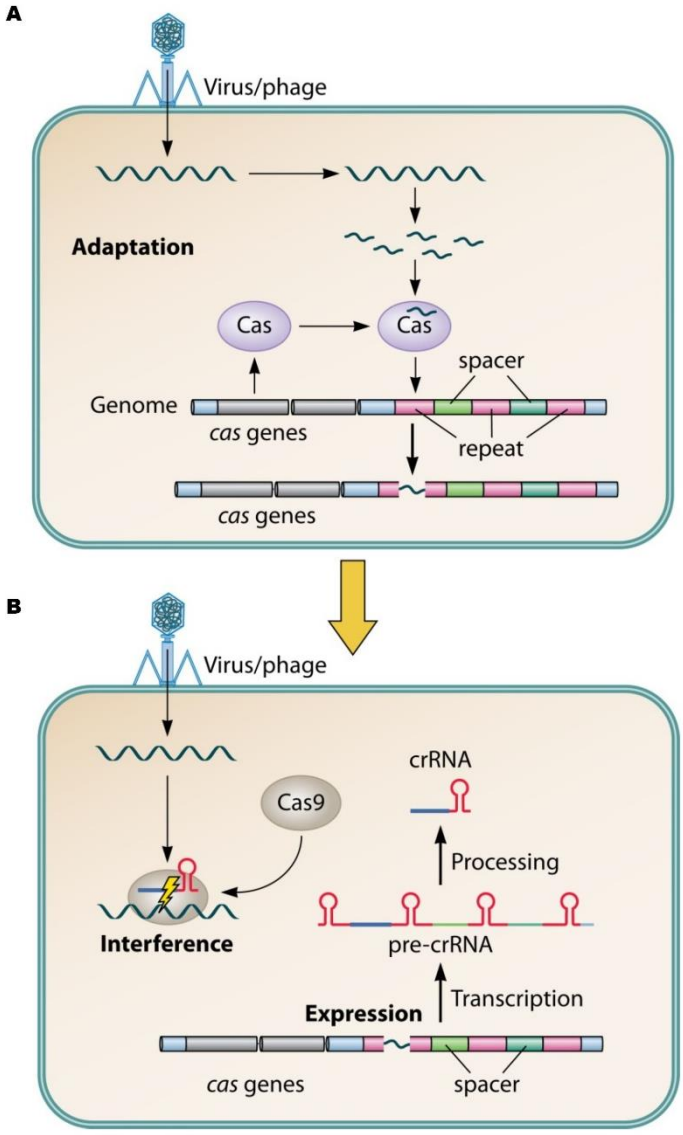


Figure 1.13 The process used by CRISPR/Cas to cleave foreign DNA as part of the acquired immune system

(A) The prokaryotic adaptive immune system. Cas protein recognises fragments of the invading DNA. The small fragment is incorporated into the spacer region of CRISPR. **(B)** Expression and interference. The CRISPR region is transcribed to generate pre-crRNA that are then processed into smaller units of RNA, named crRNA. By taking advantage of the homology of the spacer sequence present in crRNA, foreign DNA is captured, and the Cas9 protein cleaves DNA (Ishino et al., 2018).

The genome-editing tool used in this project was a type II CRISPR from *Streptococcus pyogenes*. The two primary components required to perform genome editing are the Cas9 nuclease and a chimeric single guide RNA (gRNA). The gRNA consists of a crRNA that defines the genomic target site for Cas9 and a constant trans-activating crRNA (tracrRNA) that acts as a scaffold that links the crRNA to Cas9 and facilitates the processing of mature crRNAs derived from CRISPR arrays. The target site is a 20-bp sequence immediately 5’ of a PAM sequence (5’-NGG). The Cas9-induced DSB can then be repaired through non-homologous end joining (NHEJ). The repaired process can introduce insertions/deletions (indels), disrupting the target gene. However, if a repair template containing a homology sequence is used, and the mutation will be repaired according to the supplied sequence, known as homology-directed repair (HDR) (Sander and Joung, 2014, Cong et al., 2013, Bertrand et al., 2010, Barrangou, 2015).

1.7 Project aims and Objectives

Preliminary data on 30 hpf zebrafish embryos suggest that the lncRNA1 and lncRNA2 genes are specifically expressed in cells of the haematopoietic lineage during embryonic development. Here, we want to extend the gene expression analyses and study the role of the lncRNA genes during zebrafish haematopoiesis.

The aims of this project are:

1. To investigate the expression pattern of lncRNA1 and lncRNA2 genes in wt embryos throughout embryogenesis to develop a full picture of the genes' expression patterns.
2. To verify and further define the cells types that lncRNA genes are expressed in by performing RNA whole mount in situ hybridisation experiments on various mutant zebrafish lines.
3. To evaluate the role and to learn more about the function of lncRNA genes in haematopoiesis, the two genes shall be deleted from the zebrafish genome individually and in combination using the CRISPR/Cas9 system.

2. Materials and methods

2.1 Materials

2.1.1 Buffers and solutions

Table 2.1 List of buffers and solutions. The list provides a description of their ingredients and gives the temperature they were stored at

Buffers/ Solutions	Ingredients	Store
0.03% PTU	0.3 g U-Phenylthiourea (PTU) (Sigma, P7629) in 1 L dH_2O & filter sterilized	4°C
20x Saline Sodium Citrate (SSC) pH7	3 M NaCl; 0.3 M Sodium Citrate	RT
4% paraformaldehyde (PFA)	2 g paraformaldehyde (Sigma, 15812-7); 1 M NaOH; 1xPBS	4°C
60x E3 Buffer	34.4 g NaCl; 1.52 g KCl; 5.8 g $CaCl_2.2H_2O$; 9.8 g $MgSO_4.7H_2O$ made up to 2 L with dH_2O	RT
E3 Buffer (for a working solution	15 µl of 60x E3 Buffer & 500 µl of methylene blue in 900 ml of dH2O	RT
BCL3	0.1 mM Tris-HCL (pH9.5); 0.1 mM NaCl; 50 mM MgCl2; 0.1% Tween20	RT (up to 2 weeks only)
Hybe -/- buffer (pH6.0)	50% Formamide (Sigma, F9037), 5X SSC (pH7), 9.2 mM Citric acid, 0.1% Tween20, sterile dH_2O	-20°C
Hybe +/+ buffer (pH6.0)	50% Formamide (Sigma, F9037), 5X SSC (pH7), tRNA (50mg/ml) (torula yeast, Type VI, Sigma, R6625) 50 mg/ml Heparin, 9.2 mM Citric acid (1 M), 0.1% Tween20, sterile dH_2O	-20°C
Embryo Lysis Buffer	10 ml 1 M Tris-HCL pH8.5, 0.4 ml 0.5 M EDTA pH8.0, 10 ml 10% Tween 20 (0.5%) in 200 ml sterile dH_2O	RT
Maleic Acid Buffer (MAB) pH7.5	0.1 M Maleic Acid, 0.15 M NaCl, sterile distilled water, adjust pH to 7.5 using NaOH pellets, autoclave	4°C

MAB Block	2% Roche-Boehringer Blocking reagent TM in MAB	-20°C
1X Phosphate buffered saline (PBS)	10X PBS (GIBCO, 14190250), sterile dH_2O	RT
PBSTw	1xPBS; 0.1% Tween20	RT
10 mg/ml Proteinase K	Proteinase K from Tritirachium album (Sigma, P2308 100MG, 30 units/mg) in glycerol storage buffer	-20°C
LB-Medium	10 g Bacto-Tryptone; 5 g Bacto-yest extract; 10 g NACL; dH2O to 1 liter	RT
Ampicillin-containing LB-Medium (100 µg/ml)	300 µl of Ampicillin (50 µg) in 300 ml of LB-Medium	4°C
Kanamycin - containing LB-Medium	150 µl of Kanamycin (25 µg) in 300 ml of LB-Medium	4°C
2x SSC Tween20 (SSCTw)	5 ml 20xSSC; 45 ml dH2O; 225 µl 20%Tween20	RT
MABTw	50 ml MAB; 250 µl 20%Tween20	4°C
0.2xSSCTw	5 ml 2xSSC, 45ml dH2O; 225 µl 20%Tween20	RT
Tween20	10 ml Tween 20; 40 ml dH_2O	RT
1x PBS	100 ml 10xPBS and add 900 ml dH_2O; get solution autoclaved	RT
PBSTw	50 ml of 1x PBS and add 250 µl Tween20 (20%)	RT
PBSTw 20mM EDTA	PBSTw, 0.5 M EDTA pH8	RT
5X base buffer	0.5 g NaOH, 2 ml 0.5M EDTA (pH8) bring up the volume to 100 ml with sterile dH_2O	RT
5X Neutralisation Solution	3.15 g 1 M Tris-HCl (pH8) in 100 ml of Sterile dH_2O	RT
50X TAE buffer	Tris Base, glacial acetic acid, 0.5 M EDTA (pH8), dH_2O	RT
MS222 4 g/l pH7-7.5	4 g Ethyl 3-aminobenzoate (Tricaine) methanesulfonate (MS222) (Sigma, A5040-100G) adjust pH with Tris- HCl pH9.5	4°C
50mM NaOH	0.1 ml of 1 M NaOH in 2 ml of dH_2O	RT
DNA loading dye	6X gel loading dye (NEB, B7024S)	RT
Soc Medium	2% w/v tryptone, 0.5% w/v Yeast extract, 8.56 mM NaCl, 2.5 mM KCl, dH_2O, 10 mM MgCl2, 20 mM Glucose	RT

2.1.2 Zebrafish maintenance and breeding

Adult wild-type zebrafish (*Danio Rerio*), transgenic and mutant lines were kept at 28.5°C. The fish were kept on a 14-hour light and 10-hour dark cycle. Eggs were laid and fertilised soon after the light came on. Fertilised eggs were collected with a tea strainer and raised in 1x E3 buffer with methylene blue, to suppress the growth of mould (Westerfield, 1993). Morphological features were used to stage embryos under a light microscope as described by Kimmel and colleagues (Kimmel et al., 1995). To prevent pigmentation, embryos were grown from 18 hpf in the presence of PTU (1.5 ml 0.3% PTU in 25 ml E3 buffer). PTU inhibits tyrosinase (Blagoeva and Stoichev, 1979) and maintains the embryo's optical transparency (Westerfield, 1995) for in situ hybridisation. Larvae that were allowed to grow beyond the free-feeding stage were taken to the fish aquarium. All experiments on zebrafish were performed under UK Home Office regulations and were approved by the local ethical review board under project license numbers; PPL 40/2457, PPL 30/3378 and PPL 30/3356 and personal license number; I4EA16886.

2.1.3 Mutant and transgenic zebrafish lines used for whole amount in situ hybridisation in this study

Carriers of the mutant alleles *gfi1aa*^{GTqmc551} *and gfi1ab*^{sh320} are viable and can be maintained as heterozygous and homozygous carriers (Thambyrajah et al., 2016)(*gfi1ab*^{sh320} provided by Dr Robert Wilkinson's from Sheffield). Homozygous carriers of *cloche*^{s5} (Reischauer et al., 2016), *mind bomb*^{ta52b} (Itoh et al., 2003)*,* *runx1*^{W84X} (Sood et al., 2006) and *cmyb*^{t25127} (Soza-Ried et al., 2010) alleles are lethal and were maintained as heterozygous carriers.

2.1.4 Gene information, Synteny and Alignment

The Ensembl website was used to retrieve information about genes, genomes and their conservation. The browser can be accessed at URL: http://www.ensembl.org/index.html.

2.1.5 Primer design

Primers for all PCR reactions were designed using Primer3 input web (version 0.4.0.) that can be accessed at http://bioinfo.ut.ee/primer3/

2.2 Molecular biology methods

2.2.1 Embryo lysis or fin clip biopsies for DNA extraction and genotype

Anaesthetised embryos (0.7 ml 4 g/l MS222 in 25 ml E3 buffer) or fin clip biopsies were transferred into a microfuge tube. Excess liquid was removed, and lysis buffer was added (20 µl per embryo or 50 µl per fin clip biopsy). The samples were incubated overnight at 55°C. The next day the samples were vortexed briefly and centrifuged. Then Proteinase K was inactivated at 100°C for 15 min. After incubation dH_2O was added to a final volume of 100 µl for each sample. The samples were stored at - 20°C.

2.2.2 Genotyping embryos after *in situ* hybridisation

After staining in an in-situ experiment, embryos are stored in 80% glycerol (in PBSTw) in the cold room at 4°C. For genotyping, individual embryos were transferred into a fresh 1.5 ml Eppendorf tubes and then gradually washed out of 80% glycerol. They were taken through 50% and 30% of glycerol in PBSTw (5 min in each) and finally washed three times for 5 min in 1X PBSTw. Then the PBSTw was removed and the embryos were incubated in 100 µl of 300 mM NaCl at 65°C for 4 hours to reverse crosslinking (Gansner et al., 2008). After the incubation, the NaCl solution was discarded and 25 µl 1X Base buffer was added to each embryo which was then incubated at 95°C for 20 min. Next, the tubes were cooled and 25 µl 1X neutralisation solution was added. The tubes were then vortexed and centrifuged at 3000 rpm for 5 min. The genomic DNA is in the supernatant.

2.2.3 PCR amplification

Sequences of interest were amplified on genomic DNA or cDNA using Polymerase Chain Reaction (PCR). The essential reaction components and standard conditions are as shown in Table 2.2 and Table 2.3, respectively; annealing temperatures and elongation times were adjusted in every experiment. The amplification was carried out in a TECHNE TC-512 thermocycler (Bibby Scientific). The oligonucleotides used to amplify the regions are listed in Table 3.1

Table 2.2 Reaction components for a standard PCR experiment using NEB standard Taq polymerase

Components	Amount
dH2O	16.5 µl (depends on DNA added)
10x Standard Taq reaction buffer	2.5 µl
200 µM dNTPs	2.5 µl
Forward primer 10 µM	1 µl
Reverse primer 10 µM	1 µl
gDNA 100 ng/µl	1 µl (could be variable)
Taq DNA polymerase (NEB cat. #MO273S)	0.5 µl
Total	25 µl

Table 2.3 General PCR thermocycler steps and conditions using NEB Standard Taq Polymerase

Thermocycler Steps		Temperature	Time
Initial Denaturing		94°C	5 min
35 cycles	Denaturing	94°C	30 sec
	Annealing (optimized)	55-65°C	30 sec
	Extension	68°C	1 m/kb
Final extension		68°C	5 min
Hold		4°C	

2.2.4 Agarose gel electrophoresis

PCR products were examined by gel agarose electrophoresis. The gel was prepared by dissolving agarose powder (0.5-2 % based on the size of the expected PCR fragment) in 0.5x Tris-acetate-EDTA (TAE) buffer in a microwave. Once cooled down to around 50°C, 0.5 mg/ml ethidium bromide was added, and the gel solution was poured into a gel tray. DNA loading dye (0.1 x volume) was added to each sample before the samples were loaded into the wells of the gel. One kb or 100 bp NEB DNA Ladder were used as size markers. The

voltage was set to 100 V and gels were run for 30 min in an electrophoretic chamber (Advance, Mupid®-ex). The DNA was visualised on a UV transilluminator. The results of the electrophoresis experiment were documented on the Bio-Rad Chemi XRS Gel Documentation system.

2.2.5 DNA extraction from agarose gel & measurement of nucleic acid concentrations

The band of interest was cut out of the agarose gel on a high wavelength UV transilluminator. The DNA was extracted from the agarose using the GenElute PCR Clean-Up Kit (Sigma cat. #NA1020) following the manufacturer's instructions. The concentration of the eluted DNA was determined using a NanoDrop ND-1000 (Thermo Scientific, MA, USA). An A260/280 wavelength ratio of ≈1.8 and ≈2.0 was considered to indicate high purity.

2.2.6 DNA sequencing

For sequencing, the DNA concentration was adjusted to 1 ng/µl per 100 bp of product. 5 µl of sequencing primers were required at a concentration of 3.2 pmol/ul. Samples were sent to Source BioScience (Nottingham, UK) for Sanger sequencing. Then, the sequencing results were analysed on Bioedit (version 7.2) and in Clone Manager.

2.3 Whole mount *in situ* RNA Hybridization (WISH)

2.3.1 Making probe

2.3.1.1 cDNA ligation, transformation and extraction

Taq-generated PCR fragments with 3' adenosine overhangs were ligated to the linear pGEM-Teasy plasmid with 3' thymidine overhangs using T4 DNA ligase (Promega, cat. #A1360). The ligation reactions were set up, as shown in Table 2.4. The reaction mixture was incubated at RT for 1 hr.

Table 2.4 The reaction set up to ligate the PCR fragment with pGEM-Teasy

Component	Amount
PCR product	3 µl
T4 DNA Ligase Buffer (2X)	5 µl
pGEM-Teasy vector DNA (50 ng)	1 µl
T4 DNA Ligase	1 µl
Total	10 µl

5 µl of the ligation mix was used to transform 50 µl of competent E-coli (DH5α, NEB, cat. #C2987H) following the provider's instructions. The mix was incubated on ice for 30 minutes, heat-shocked at 42°C for 40 seconds and placed back on the ice for 5 min. Then, 500 µl of SOC media was added to the transformed bacteria. The sample was placed in the 37°C shaking incubator for 1 hr to allow bacteria growth. Next, 200 µl of culture were spread out onto LB agar plates containing 100 µg/ml ampicillin. The plates were incubated at 37°C overnight. The next day, single colonies were picked and used to inoculate 3 ml of LB media containing 100 µg/ml ampicillin for a DNA

mini-prep. Bacterial cultures were incubated overnight with shaking at 37°C. Plasmid DNA was isolated from overnight bacterial cultures using a mini-prep kit (Sigma, cat. #PLN10). The experiment was performed as recommended in the manufacturer's instructions. To verify that the bacterial clone mini-prep carries a plasmid with the correct insert, plasmid DNA was isolated from the transformants and digested with *EcoRI* (NEB, cat. #R0114S) at 37°C. The digests were then examined by agarose gel electrophoresis. DNA samples were also sent for sequencing.

2.3.1.2 *In vitro* transcription to make antisense RNA probe

To be able to isolate larger quantities of the plasmid DNA, 100 ml bacterial cultures were grown in LB medium containing 100 µg/ml ampicillin. The plasmid DNA was isolated using the midi-prep kits (NucleoBond Xtra Midi Plus, cat. #740412.10) according to the manufacturer's instructions. To verify that the bacterial clone midi-prep carries a plasmid with the correct insert, plasmid DNA was isolated from the transformants and digested with *EcoRI* at 37°C. The digests were then examined by agarose gel electrophoresis. DNA samples were also sent for sequencing. As soon as the sequencing had been confirmed, the DNA solution was used to generate a Digoxigenin-labelled RNA probe. 10 µg of DNA was linearised using *ApaI* at 25°C. After successful linearization, distilled H_2O was added to the DNA solution to a final volume of 200 µl. An equal volume of phenol:chloroform:isoamyl alcohol mix (25:24:1)(Sigma, cat #P2069-400ML) was added. The samples were mixed by vortex and centrifuged at 13.000 rpm for 5 min. Then, the upper aqueous phase was collected and transferred to a fresh tube and an equal volume of chloroform:isoamyl mix (24:1)(Sigma, cat. #C0549) was added. Again, the samples were vortexed and centrifuged for 5 min at

13.000 rpm. The upper phase was transferred to a fresh tube, and the DNA was precipitated with 1:10 volume KAc (3M, pH 5.2) and 2.5x volumes of 100% ethanol. The mix was kept in -80°C for 30 min and centrifuged at 13.000 rpm for 15 min before the pellet was washed with 70% ethanol and resuspended in 50 µl of sterile dH_2O. The linearised DNA was then used in an *in vitro* transcription reaction that was set up, as shown in Table 2.5. The reaction mixture was incubated with the appropriate RNA polymerase *SP6* at 37° for 2 hrs.

Table 2.5 The reaction set up for in vitro transcription

Component (Roche, cat. #11175025910)	**Amount**
Linearised template DNA (1-2 µg)	X µl (depends on concentration)
dH_2O	12.5 -x µl (depends on DNA amount)
DIG-UTP NTP mix	2 µl
5x transcription buffer	4 µl
RNase inhibitor	0.5 µl
RNA polymerase SP6	1 µl
Total	20 µl

Following 2 hrs of incubation, 1 µl DNaseI (Roche, cat. #11175025910) solution was added to each sample. The samples were incubated for 15 min at 37°C. The reaction was stopped by adding 1 µl 0.5 M EDTA pH 8 and 80 µl dH_2O. The RNA was precipitated by adding 1/3 volume 10 M NH_4Ac and 2.5 volumes of 100% ethanol. The samples were incubated on ice for 10 min and then centrifuged for 10 min at 13.000 rpm. The pellet was washed with 70% ethanol, then dried briefly and eventually resuspended in 100 µl of sterile dH_2O. The probe solution was stored at -80°C (Appendix E).

2.3.1.3 Embryo collection, chorion removal and fixation

Embryos were collected at the stage of interest. Embryos older than 18 hpf were firstly dechorionated manually, then fixed in 4% paraformaldehyde reagent (PFA) overnight on a rocker in the cold room at 4°C. Embryos younger than 18 hpf were fixed in their chorions. Fixed embryos were washed in PBSTw (PBS, Sigma cat. #P5119) 3 times for 5 min. Embryos younger than 18 hpf were washed and then dechorionated. All embryos were gradually dehydrated by performing 5 min washes in increasing concentrations (25%, 50%, 75% and 100%) of methanol in PBSTw. Finally, the embryos were washed once more in 100% methanol and then stored at -20°C.

2.3.1.4 Whole mount RNA *in situ* hybridisation (WISH)

The process was performed over three consecutive days.

On day one, the dehydrated and permeabilised embryos were rehydrated. For this purpose, embryos were washed in 75%, 50% and 25% methanol before being briefly washed 3x in PBSTw. Embryos older than 24 hpf were permeabilised further by treating them with 10 µg/ml of proteinase K in PBSTw. Exact treatment times depend on the age of the embryos (Table 2.6) and may need to be adjusted based on the activity of the proteinase K solution.

Table 2.6 Proteinase K treatment durations according to the age of embryos

Embryo age	Duration of Proteinase K treatment
26-31hpf	5 min
2 dpf	15 min
3 dpf	30-35 min
4 dpf	45 min
Over 4 dpf	1 hr

The embryos were rinsed in 100% PBSTw to get rid of the proteinase K. Next, the embryos were re-fixed in 4% PFA for 20 min at RT. Then, the embryos were washed 5 times in PBSTw for 5 min to dispose of the PFA. Following a wash in 50% PBSTw and 50% Hybe+/+ for 5 min, the embryos were incubated in 100% Hybe +/+ solution at least for 1 hr at 65°C. The Hybe +/+ was then replaced by Hybe+/+ containing the probe at a dilution of 1:200. Depending on the probe concentration after *in vitro* transcription, the probe dilution may need to be adjusted. The embryos were incubated in the probe solution overnight at 65°C for probe hybridisation.

On day two, the probe was recovered and the embryos were per washed for 10 minutes each in 100% Hype -/-, 75% Hype -/- 25% SSCTw, 50% Hype -/- 50% SSCTw, 25% Hybe-/- 75% SSCTw, then followed by a wash in 100% 2x SSCTw. All these incubation steps are performed at 65°C. The embryos were subsequently washed 4 times in 0.2x SSCTw for 15 min each at 65°C. These four washes remove a non-specifically bound probe. Subsequently, the embryos were gradually washed into a maleic acid buffer. For this purpose, embryos were incubated at RT for 5 min each in mixtures of 75% 0.2x SSCTw 25% MABTw, 50% 0.2x SSCTw 50% MABTw, 25% 0.2x SSC 75% MABTw, followed by a wash in 100% MABTw. The embryos were then blocked in MAB Block (non-specific protein binding was blocked by incubating embryos in MAB block) for at least 1 hr at RT followed by incubation in anti-DIG antibody (Roche, cat. #11093274910) at a concentration of 1:5000 overnight on a rocker at 4°C in the cold room.

On day three, the embryos were washed in MABTw 8 times for 15 min at RT to remove any excess antibody. The embryos were then washed 3x for 5 min in BCL3 buffer before being placed in the BM Purple:BCL3 (1:1) mix (Sigma, cat. #11442074001). Embryos were left on a rocking shaker in the dark at RT until they developed colour

and were deemed to have been stained sufficiently. Washing the embryos in PBSTw containing 20 mM EDTA 3 times for 5 min stopped the reaction. The embryos were then fixed in 4% PFA at RT for 20 min after that washed three times for 5 min in PBSTw. They were then taken through increasing concentrations 30%, 50% and 80% of glycerol in PBSTw. Embryos were stored at 4°C in 80% glycerol indefinitely.

2.3.1.5 Imaging

Images of stained embryos were taken with a digital camera attached to a dissecting microscope (Nikon DS-5MC, Nikon, model SMZ1500, Japan, respectively). Adobe Photoshop software was used to assemble and edit images.

2.4 Targeting genes using the CRISPR/Cas9 technique

2.4.1 Single-guide RNA (sgRNA) design

The CRISPRscan web tool (https://www.crisprscan.org/) was used to look for the suitable sgRNA target sites. Suitable target sites are 20bp sequences upstream of a Protospacer Adjacent Motif (PAM) sequence (NGG). To generate the complementary sgRNA, this TS sequence is cloned into the sgRNA expression vector pDR274 that was developed by the Young lab (Hwang et al., 2013b). To the presence of a T7 promoter, this vector allows the in vitro transcription of sgRNAs by T7 RNA polymerase. This enzyme prefers a GG dinucleotide at the transcription start, thus, restricting the sgRNA design to 5'GG-N18-3' sequences. It has been shown that GN and NG dinucleotides are also acceptable at the 5' end of the target site. Their conversion to a GG sequence in the sgRNA optimises transcription and does not markedly interfere with TS recognition by the sgRNA. Thus, suitable TS have the sequence 5'-GN/NG-(18)-PAM-3' (Moreno-Mateos et al., 2015). To clone the sgRNA-coding sequence in this vector, pairs of primers are designed (a) with complementary sequences that can be hybridised to form a linker and (b) with 5' overhangs to allow the integration into the *BsaI* digested pDR274 vector fragment. Table 2.7 shows all TSs sequences and oligos designed.

Table 2.7 The sequences and oligonucleotides used for targets site

TS	Strand	Sequence 5'→3'	Tools	Score value	Off-target site	Complementary oligos	Oligos sequence (F)5'→3'/ (R)3'→5'
TS 14	-	GTCTATATGCACAATTAAAGTGG	CRISPR scan	16	0	DC1330 (F)	TAgGCTATATGCACAATTAAAG
						DC1331 (R)	AAACCTTTAATTGTGCATATAG
TS 65	+	TGTGTTGTGTTCTTCCAGTAAGG	CRISPR scan	65	0	DB909 (F)	TAgGTGTTGTGTTCTTCCAGTA
						DB910 (R)	AAACTACTGGAAGAACACAACA
TS 81	+	GAGAGGAACTCTAACGGGTTTGG	CRISPR scan	55	0	DC1442 (F)	TAgGGAGGAACTCTAACGGGTT
						DC1443 (R)	AAACAACCCGTTAGAGTTCCTC
TS 67	+	GTCGATGATGGGTGACGGGCGGG	CRISPR scan	67	0	DC1394 (F)	TAgGCGATGATGGGTGACGGGC
						DC1395 (R)	AAACGCCCGTCACCCATCATCG
TS 63	+	GTTTATTGATCAAACATCTCAGG	CRISPR scan	63	0	DC1332 (F)	TAGgTTATTGATCAAACATCTC
						DC1333 (R)	AAACGAGATGTTTGATCAATAA

2.4.2 Cloning of the sgRNA-coding linker in pDR274

The pDR274 plasmid is digested with *BsaI* (NEB, cat #R3535L) at 37°C for overnight. The next day, the sample run on 1% agarose gel to check the plasmid had been digested and to isolate the vector fragment from the small *BsaI* fragment. 10 µl of 100 µM each oligo was mixed and hybridised to form linkers. For this purpose, samples were heated at 95°C for 5 mins and then left to cool gradually to RT. The linkers were ligated to the *BsaI* vector fragment and incubated at 16°C overnight. The reaction mix is shown in Table 2.8.

Table 2.8 The reaction mix to insert the linker into the expression vector

Components	Amounts
Linker	2 µl
pDR274	5 µl
10x ligation buffer	1 µl
dH_2O	1.5 µl
T4 DNA ligase	0.5 µl
Total	10 µl

The following day the mixture was used to transform subcloning efficiency E.coli (DH5α, NEB, Cat. #C2987H). At the end of the transformation procedures, the cells were spread on LB plate (kanamycin 50 µg/ml) and incubated at 37 °C overnight. Colonies grown on the plate were used to inoculate 3 ml LB medium liquid cultures that were incubated in a shaking incubator overnight. From cells pelleted from these cultures, plasmid DNA was extracted using the mini-prep and midi-prep kits described previously. The samples

were sent for sequencing using M13F primer. This was to verify the sequence of the sgRNA coding sequence.

2.4.3 sgRNA synthesis for microinjection

Single gRNAs were generated as run-off transcripts with defined 5' and 3' ends. The 5' end corresponds to the transcription start site of T7 RNA polymerase. The 3' end is defined by the restriction site for *DraI* at which the plasmid is cut. All midi-prep sample was digested with *DraI* (NEB, cat. #R0129L). This enzyme has two restriction sites on either side of the sgRNA coding sequence on the vector. The digest, thus, releases the sgRNA-coding sequence with the upstream T7 promoter from the expression vector. To verify the digestion, reaction mixture run on 1% agarose gel as described above. The digested DNA was purified and then used in the *in vitro* transcription. The mixture reaction was made up as shown in Table 2.9. The sample was incubated at 37°C overnight. At the end of the incubation, 1 µL of TURBO DNase (mMESSAGE mMACHINE, Cat no. AM1344) was added and incubated at 37°C for 15 min to digest the DNA template. Then, 15 µL of ammonium acetate 5 M ammonium acetate, 100 mM EDTA (mMESSAGE mMACHINE, Cat no. AM1344) was added to stop the reaction and then add 35 µL of dH_2O. To remove the enzymes, the RNA was mixed with phenol:chloroform:isoamyl alcohol (25:24:1) mix, vortexed and centrifuged at 13.000 rpm for 10 min. Then, the upper aqueous phase was collected and transferred to a fresh tube and an equal volume of a chloroform:isoamyl alcohol (24:1) mix was added, vortexed and centrifuged for 10 min at 13.000 rpm. The upper phase was transferred to a fresh tube and 100 µL 100% isopropanol were added. The sample was vortexed and chilled for 30 min at -80°C. Then, the sample was centrifuged at 13.000 rpm for 10 min to pellet

RNA. After the centrifugation, the supernatant was removed and the RNA pellet was dried. The pellet was resuspended in 15 µL of dH_2O. The concentration of the RNA was measured on the Nanodrop. To make sure that the sgRNA is not degraded, a mix of 1 µL of the sgRNA, 5 µL of loading dye and 4 µL of dH_2O was incubated at 65°C for 5 min then placed on ice for few seconds. The mixture was run on 1.5% agarose gel and the product kept at -80°C.

Table 2.9 The mixture reaction to make *in vitro* transcription

Components	Amount
5x transcription buffer	20 µl
DTT	10 µl
RNasin (Promega, cat. #N251A)	1 µl
Unlabeled rNTPs (25mM) (NEB, cat. #N0446S)	2 µl
RNA polymerase (T7) (NEB, cat. #M0251L)	2 µl
DNA fragment (1-2 µg)	X µl
dH_2O	xµl
Total	100µl

2.4.4 Microinjection needle preparation

The microinjection needles were pulled with a puller (Sutter Instrument, Novato, CA). The glass capillaries (HARVARD, GC 100F-15, cat. #30-0020) with an internal filament with the current puller filament the following parameters were used: heat (305), pull (200), time (115) and velocity (115). Injection needles were back loaded using micro loader pipette tips (Eppendorf, cat. #E270445J). Needle tips were broken (just before use) with clean forceps to enable the

delivery of the injected solution. The drop size was adjusted with the help of a graticule in an eyepiece where 6.5 graticule unite = 1 nl .

2.4.5 Microinjection

5 µl of the injection solution was prepared by mixing of 1.5 µl Cas9 protein (BioLabs Kit, cat. M0646T), 0.5 µl buffer (1x Cas9 Nuclease Reaction Buffer) and 1.7 µl red ruby 33 mg/ml (diluted 1:6) then 80-100 ng of gRNAs added then incubated for 5 min at 37°C as described by Burger and colleagues (Burger et al., 2016). Injections were performed into one-cell stage embryos. For this purpose, the fertilised eggs 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 injection needle. The needle was then backfilled with 3 µl of the injection solution using a micro loader pipette tip (Eppendorf, Hamburg, Germany). Volumes of 1 nl were injected per embryo.

2.5 Adult zebrafish fin clipping

To collect fin clip biopsies, fish were anaesthetised in a solution containing 200 ml of facility water to which approximately 6 ml of 4 g/L MS222. The fish was observed for a while until it turned belly up, indicating it was anaesthetised. The anaesthetised fish was then transferred immediately onto a clean surface using a plastic spoon. The fin was clipped with sterilised surgical scissors at a point not higher than halfway between the tip of the fin and the point where the scales end. The fin clip was carefully transformed into a 1.5 ml Eppendorf tube and the tube was left on ice for further processing.

2.6 Caudal fin clipping at three days post-fertilization

Zebrafish larvae were anaesthetised in a petri dish containing 29 ml E3 buffer (1x), 1 ml MS222 (4 g/l) and 3 µl Tween 20 (20%). Sedated larvae were transferred onto a microscope slide. Under a stereomicroscope, the tail was amputated with a sterile scalpel distal to the end of the arterial circulation. The amputated tail was transferred into a PCR tube with 15-20 µl of 50 mM NaOH. The tube was incubated at 98°C for 2 min. Next, after the incubation, 1/10th volume of 1 M Tris-HCL (pH8.0) was added. The tube was left in -20°C for further processing. One µl of the gDNA used in a 25 µl PCR reaction for genotyping done (Table 2.2).

3. Results

3.1 The *lncRNA1* and *lncRNA2* genes are expressed in primitive and definitive haematopoietic cells in zebrafish embryos

3.1.1 Introduction

A transcriptome analysis performed on GFP-positive and GFP-negative cells isolated from 30 hpf *qmc551* transgenic zebrafish embryos identified the genes for lncRNA1 and lncRNA2 as being specifically expressed in GFP+ cells (Moore, unpublished data). To determine the cells types that expressed lncRNA1 and lncRNA2 and to define the timing of their expression, lncRNA1 and lncRNA2 in-house probes were generated to perform whole mount in situ hybridisation experiments on wt zebrafish embryos. The expressions of *lncRNA1* and *lncRNA2* were first examined at 26 hpf in wt embryos. Using a probe directly transcribed from a PCR product, the *lncRNAs* have previously been shown to be expressed in the vDA and the PBI in wt zebrafish embryos (Moore, unpublished data). Here, a detailed analysis of expressions during ontogeny were conducted by WISH to identify where and at what points *lncRNAs* are expressed.

3.1.2 *lncRNA1* is expressed in primitive and definitive haematopoietic cells in zebrafish embryos

During early ontogeny *lncRNA1* expression was not detected which may be due to low expression levels that are difficult to detect by WISH (Figure 3.1A). *lncRNA1* expression was first detected at 26 hpf in the position of the trunk region (vDA) and the PBI which is

consistent with Dr Moore's results. The cells in the vDA may represent the earliest blood-primed endothelial cells which give rise to the haemogenic endothelium. The PBI cells may be immature blood progenitors of the primitive or pre-definitive wave (Figure 3.1B). By 30 hpf, *lncRNA1* is still expressed in the vDA and the PBI and was expressed at higher levels than during the earlier stage (Figure 3.1C). At 2 dpf, *lncRNA1* expressing cells can be observed in the CHT. The expression level appears to be higher than at earlier stages. These cells may represent HSPCs. There was no expression remaining in the vDA (Figure 3.1D). By 3 dpf, *lncRNA1* expression is still observed in the CHT and begins to increase steadily in the thymus over the next days (Figure 3.1E-F).

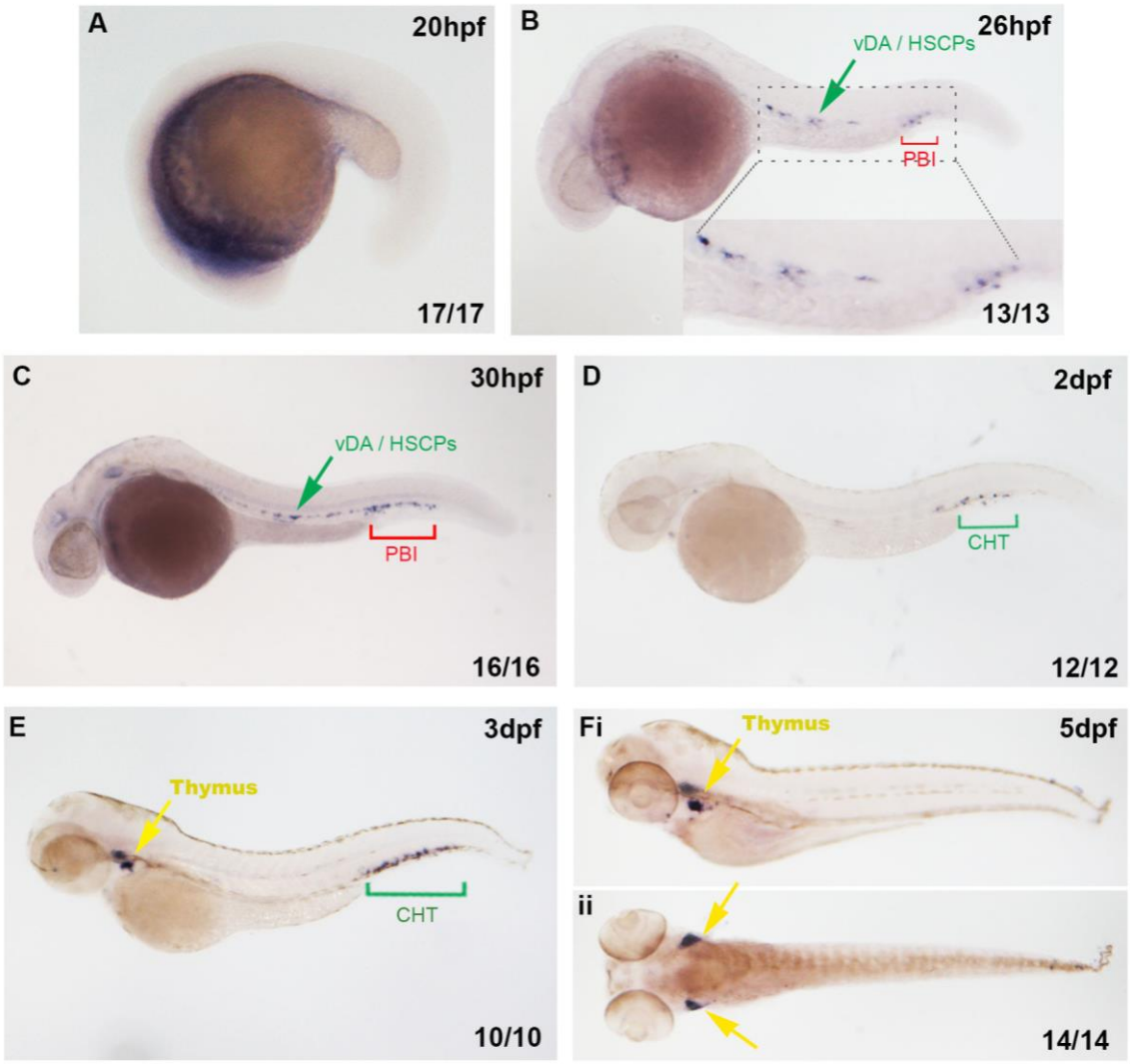


Figure 3.1 *lncRNA1* is expressed in sites associated with primitive and definitive haematopoiesis during zebrafish development

Images of wt embryos stained in a whole in situ hybridisation experiment using a probe for lncRNA1. All embryos are presented in lateral views with the anterior to the left and the dorsal side up. An additional dorsal view is presented in (Fii). (B) Close-up image of the trunk and tail regions. Red bracket: the PBI, green arrow: endothelial cells of the vDA and HSCPs, green bracket: the CHT, yellow arrow: thymus. The numbers of stained embryos over the total number of embryos analysed are provided on the panels.

3.1.3 *lncRNA2* is expressed in primitive and definitive haematopoietic cells in zebrafish embryos

lncRNA2 expression was first detected at around 18 ss in a pair of stripes of cells (Figure 3.2A). At 20 hpf, this bilateral expression had fused at the midline, in the position of the ICM and extended towards the posterior end of the zebrafish embryo. These subsets of cells could represent prRBC progenitors or endothelial progenitors (Figure 3.2B). At 23 hpf, *lncRNA2* expression could be observed in the YS and appeared to enter the circulation. The cells pooled around the heart and in the PBI. These cells could represent erythroid cells (Figure 3.2C). By 26 hpf, *lncRNA2* expressing cells became restricted to the vDA, with residual expression in the yolk sac (YS) and PBI, which is consistent with Dr Moore's results. These cells may represent haemogenic endothelial cells and immature prRBC progenitors, respectively (Figure 3.2D). The lncRNA2 gene expression was restricted to the vDA and the CHT at 2 dpf embryo (Figure 3.2E). Starting at 3 dpf, the expression levels of *lncRNA2* began to be reduced in the CHT before disappearing completely at 4 dpf (Figure 3.2F-G).

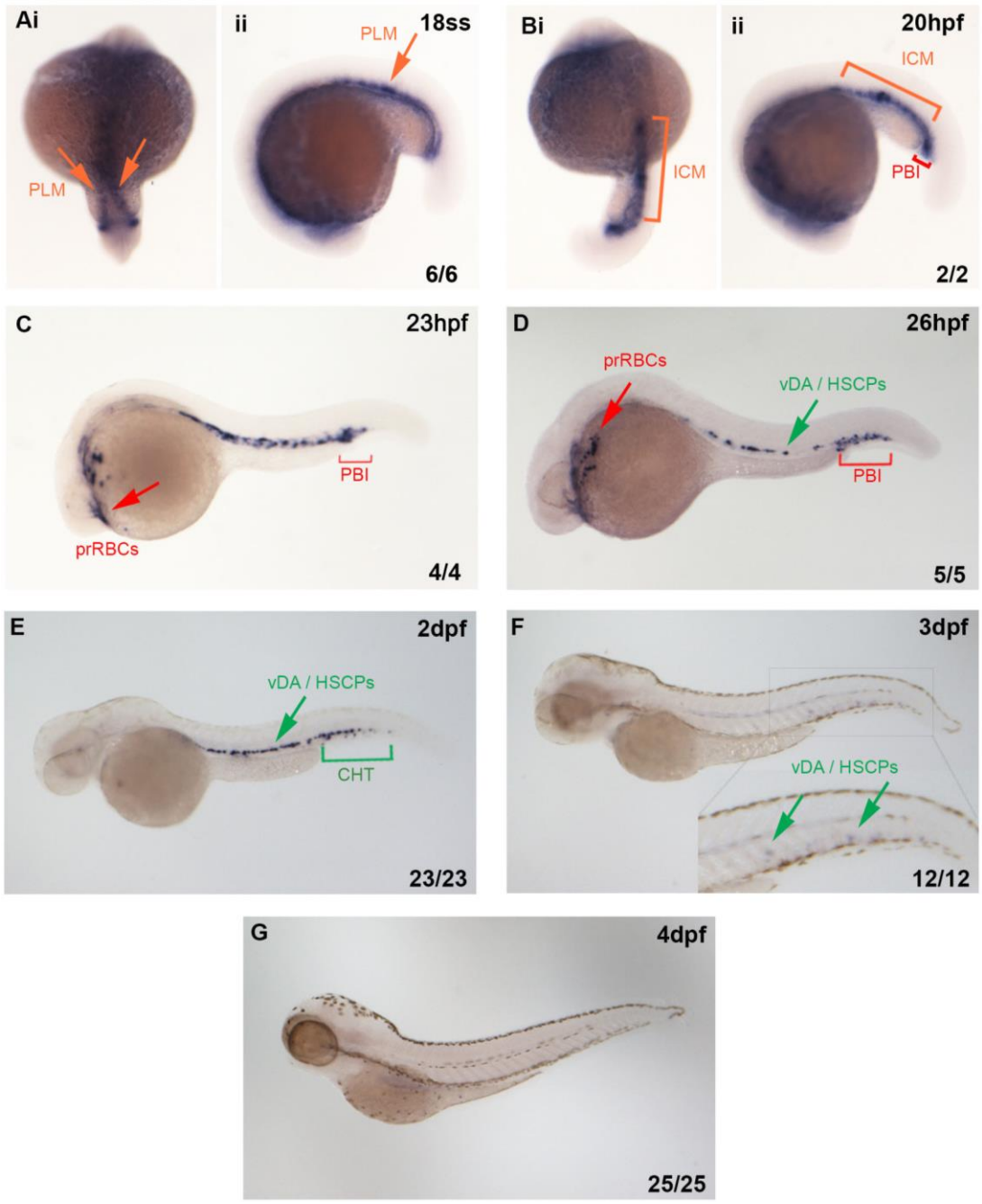


Figure 3.2 *lncRNA2* expression is detected at sites associated with primitive and definitive haematopoiesis during zebrafish development

Images of wt embryos stained in a WISH experiment using a probe for lncRNA2. (Aii and Bii) Views of embryos are dorsal with anterior up. (Ai and Bi) and (C-G) Embryos are presented in lateral views with the anterior side to the left and the dorsal side up. (F) Close-up image of the tail region. Orange arrow: the PLM, orange bracket: the ICM, Red bracket: the PBI, Red arrow: the YS/prRBCs green arrow: endothelial cells of the vDA and HSCPs, green bracket: the CHT. The numbers of stained embryos over the total number of embryos analysed are provided on the panels.

3.2 Loss of *lncRNA1* and *lncRNA2* expression in cloche mutants confirms that the *lncRNA*s are expressed in blood and endothelial cells in the wt zebrafish embryo

To confirm that the lncRNAs 1 and 2 were indeed expressed in haematopoietic and endothelial cells of wt embryos, their expression was studied in *cloche (clo)* mutant zebrafish embryos that lack most haematopoietic, endothelial and endocardial cells (Reischauer et al., 2016). In zebrafish, *cloche* encodes a PAS-domain-containing bHLH transcription factor (Npas4l) on chromosome 13, which is expressed with a highly specific spatiotemporal pattern, starting during late gastrulation. The clo^{s5} mutant allele is a single-base mutation in exon 3 converts the tyrosine 136 codon into a premature stop codon. The clo^{s5} mutation leads to the loss of most endothelial and haematopoietic cells in zebrafish (Figure 3.3A).

lncRNA1 and *lncRNA2* expression patterns were investigated by WISH. For this analysis, a mixed batch of wt, clo^{s5} heterozygous, and clo^{s5} homozygous embryos were obtained from a clo^{s5} heterozygous incross. At 30 hpf, a subset of the embryos that were stained in the lncRNA1 and lncRNA2 WISH experiments completely lacked all vDA or PBI staining (Figure 3.3B). To see whether these embryos were homozygous clo^{s5} mutants, genotyping was carried out on WISH-stained embryos. Three consecutive PCR amplifications were applied. The first set of primers are (DB1434 and DB1435) designed to anneal to sequences upstream from the second set of primers and are used in an initial PCR reaction. Amplicons resulting from the first PCR reaction are used as a template for the second set of primers (DB1410 and DB1413) and a second amplification step. Finally, amplicons resulting from the second PCR reaction are used as a template for the third set of primers (DB1411 and DB1412) and a third amplification step. The gDNA extracted from stained and

unstained WISH embryos. The expected size of the nested PCR fragment is ~340 bp. Sanger DNA sequencing of the PCR fragments using primer DB1411 showed of the embryos with no *lncRNA1* and *lncRNA2* expression were homozygous for the point mutation that converts the tyrosine 136 codon into a premature stop codon (Figure 3.3C). These findings support the view that *lncRNA1* and *lncRNA2* are expressed in blood and/or endothelial cells of the wt embryo.

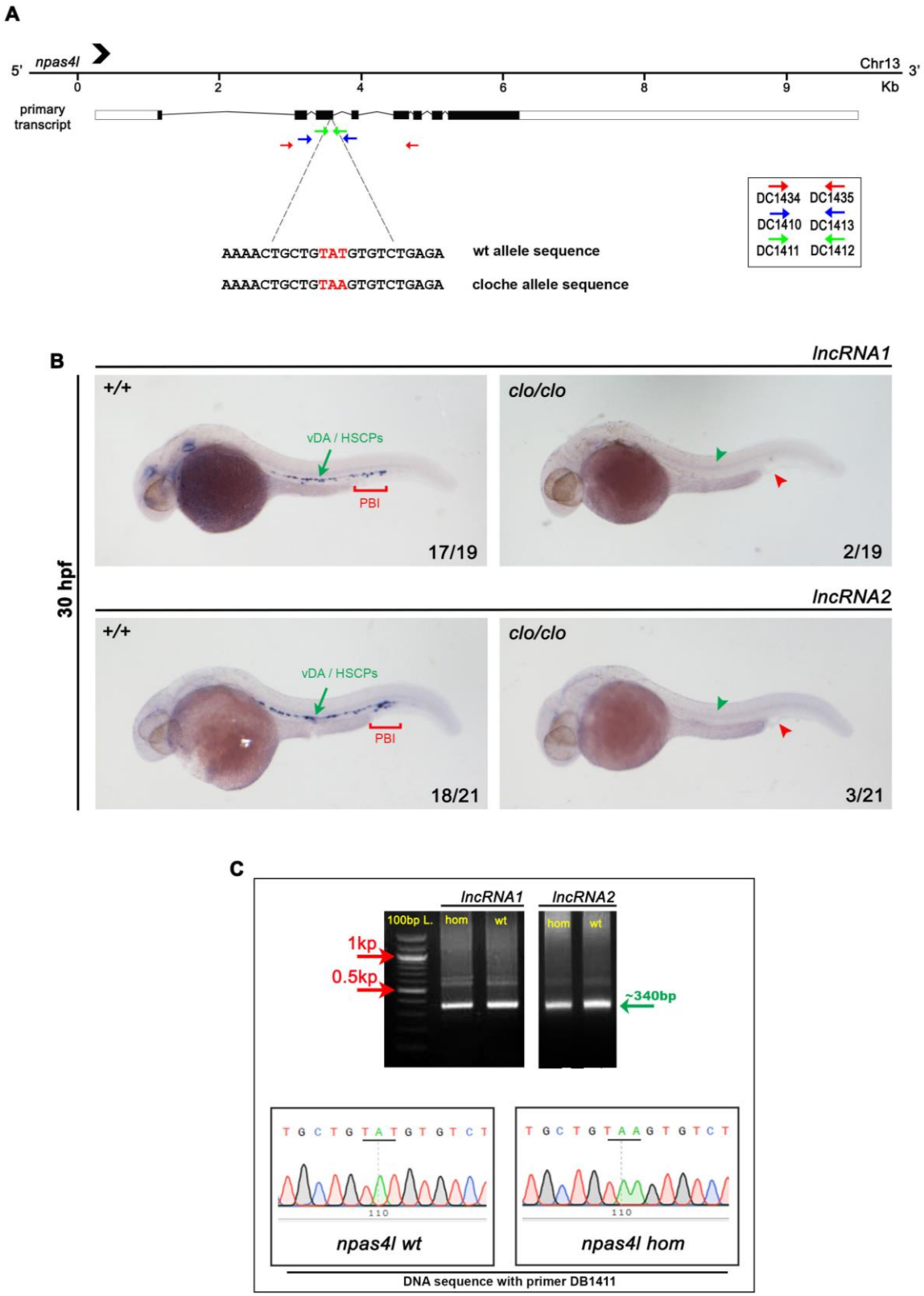


Figure 3.3 Loss of *lncRNA1* and *lncRNA2* expression in *cloche* mutant embryos confirms that the lncRNAs are expressed in blood and/or endothelial cells in the wt zebrafish embryo

(A) Schematic representation of the *npas4l*^{s5} allele at the *npas4l* locus on zebrafish chromosome 13 (Reischauer et al., 2016). The oligos used to PCR amplify the sequence around the mutation site for genotyping are shown in the box. **(B)** Lateral view of 30 hpf embryos obtained from an incross of *cloche* heterozygous zebrafish subjected to WISH using *lncRNA1* and *lncRNA2* probes. All embryos are presented in a lateral view with the anterior side facing to the left and the dorsal side up. Green arrow: endothelial cells of the vDA and HSPCs, red bracket: the PBI, green arrowhead: missing expression in the endothelial cells of the vDA and HSPCs, red arrowhead: missing expression in the primitive haematopoietic progenitors of the PBI. The numbers of embryos with phenotypes over the total numbers of embryos analysed are provided in each panel. **(C)** PCR amplification was performed on gDNA extracted from WISH embryos to distinguish *wt* and *clo*^{s5} alleles. The PCR fragment was sequenced using primer DB1411. The panels show Sanger DNA sequencing reads of the coding strands of wt and *clo*^{s5} alleles. The sequence of the mutated codon is underlined.

3.3 The loss of *lncRNA1* and *lncRNA2* expressions in a subset of cells in *mib* mutant reveals that these cells are definitive blood cell progenitors

The results obtained from the WISH analysis of the *clo* embryos indicated that lncRNA genes were expressed in blood and/or endothelial cells. It remained unclear whether *lncRNA1* and *lncRNA2* were expressed in primitive or definitive haematopoietic cells or both. To address this issue, *lncRNAs* expressions were examined in *mind bomb (mib)* mutant zebrafish embryos using WISH. The zebrafish *mib*^{ta52b} allele mutant carries a point mutation in exon 22 of the *mib* gene on chromosome 2 that is predicted to result in an amino acid substitution, a change from methionine to arginine M>R (ATG>AGG) (Figure 3.4A). This mutation interferes with the function of the ubiquitin E3 ligase encoded by the *mib* gene. This E3 ligase targets the Notch ligand Delta. The failure to ubiquitinate Delta leads to deficits in the Notch signalling pathway (Itoh et al., 2003). The *mib* zebrafish mutant fails to specify arterial blood vessels (Lawson et al., 2001) and display lacks in definitive haematopoiesis, without affecting primitive haematopoiesis (Burns et al., 2005, Gering and Patient, 2005).

Homozygous *mib*^{ta52b} mutants were obtained for WISH analysis from an incross of heterozygous carriers to investigate the *lncRNA1* and *lncRNA2* expressions patterns in these mutants and to compare these patterns with those in wt siblings. By approximately 26 hpf, homozygous *mib*^{ta52b} mutant embryos can easily be identified by their characteristic trunk and tail curvature. In these mutant embryos, the *lncRNAs* expression in the PBI was similar to that of the wt siblings. In these mutant embryos, *lncRNA2* expression was observed at higher levels than *lncRNA1*. *LncRNA2*-expressing cells were detected in the trunk region anterior to the PBI while no

expression was observed over the YS, suggesting that *lncRNA2*-expressing cells could be prRBCs that had failed to enter circulation and were trapped in the posterior ICM. In contrast to the prRBCs, expression of both *lncRNA* genes was no longer detected in the vDA in homozygous *mib*^{ta52b} mutant embryos, suggesting that these cells are associated with the definitive wave of haematopoiesis (Figure 3.4B). To confirm that the embryos with the curved body axis were indeed homozygous *mib*^{ta52b} mutant embryos, gDNA was extracted from stained embryos and PCR reactions were performed to amplify the mutated sequence. The region of interest was amplified using external primers DB1414 and DB1417. The PCR product was used as a template for the subsequent nested PCR using the nested internal primers DB1415 and DB1416. The expected nested PCR fragment is ~240 bp. Sanger DNA sequencing using primer DB1415 confirmed that embryos with *mib* phenotypes were indeed homozygous carriers of *mib*^{ta52b} alleles. The *mib* allele sequence revealed a point mutation resulting in a predicted amino acid substitution (ATG>AGG) (Figure 3.4C). These findings suggest that the two *lncRNA* genes are expressed in both primitive and definitive blood cells in zebrafish.

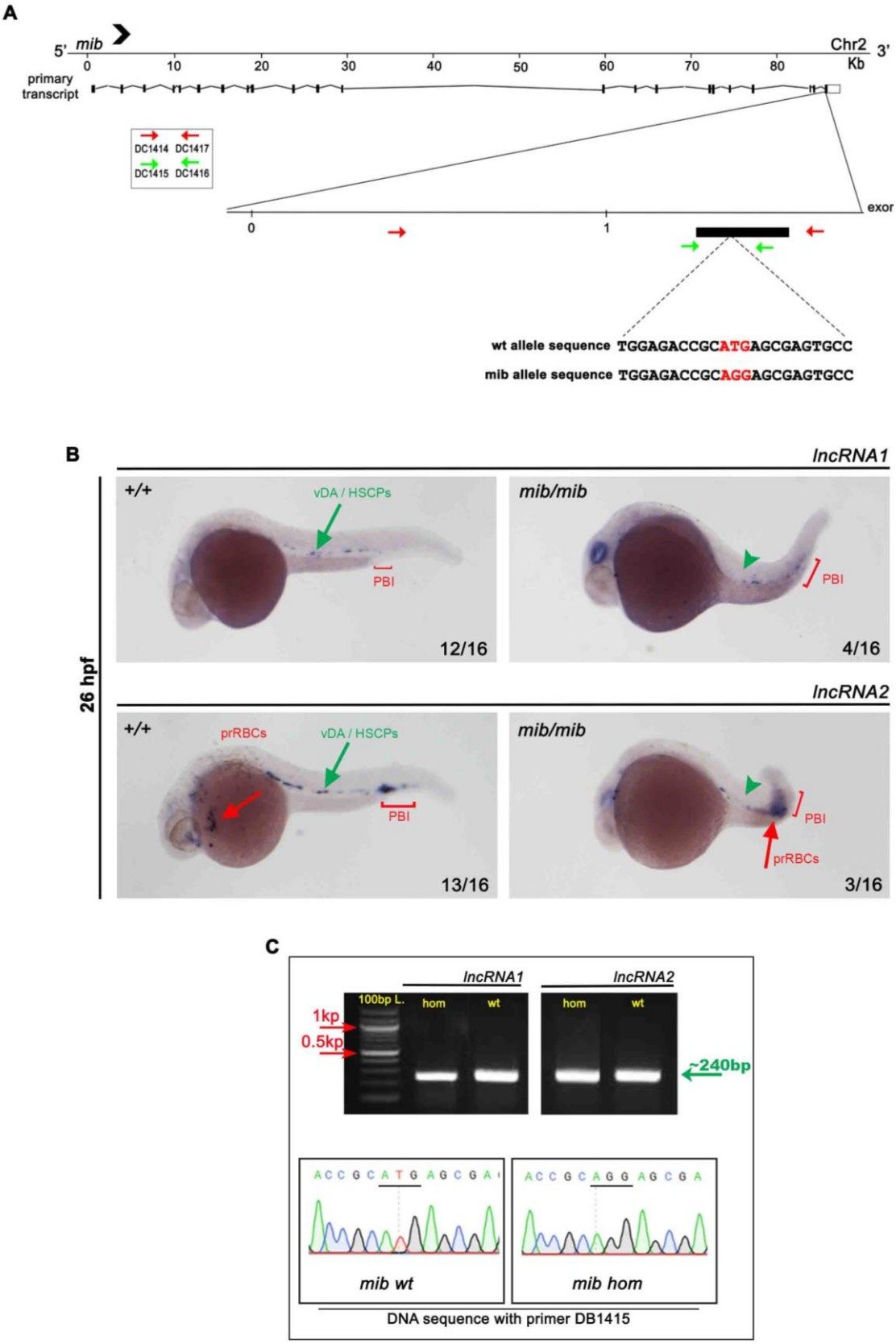


Figure 3.4 The loss of *lncRNA1* and *lncRNA2* expressions in a subset of cells in *mib* mutant embryos reveals that these cells are definitive blood cell progenitors

(A) Schematic representation of the *mib*^{ta52b} allele at the *mib* locus on zebrafish chromosome 2. The oligos used to PCR amplify the sequences around the mutation site for genotyping are shown in the box. **(B)** Lateral view of 26 hpf embryos obtained from an incross of *mib* heterozygous zebrafish and subjected to WISH using *lncRNA1* and *lncRNA2* probes. All embryos are presented in the lateral view with the anterior side facing left and the dorsal side up. Green arrow: endothelial cells of the vDA and HSPCs, red arrow: prRBCs, red bracket: the PBI, green arrowhead: missing expression in the endothelial cells of the vDA and HSPCs. The numbers of embryos with phenotypes over the total numbers of embryos analysed are provided in each panel. **(C)** PCR amplification was performed on gDNA extracted to distinguish wt and *mib*^{ta52b} alleles. The PCR fragments were sequenced using oligo DB1415. The panels show Sanger DNA sequencing reads of the template strand of DNA isolated from wt and *mib*^{ta52b} homozygous mutant embryos. The coding sequence has been underlined.

3.4 *lncRNA2* is upregulated in the vDA and PBI of qmc551 homozygous mutants

RNA-Seq data on both 20 hpf qmc551 heterozygous and homozygous of prRBCs reveals low *lncRNA1* and higher *lncRNA2* expression (Moore et al, unpublished data) (Figure 3.5C). The expression of *lncRNA2* was investigated to find out whether functional redundancy is the reason for the lack of haematopoietic cells defects in *qmc551* homozygous mutant embryos. However, the *lncRNA1* in *qmc551* homozygous mutant embryos was not investigating due to the low level. For this analysis, a mix of wt, heterozygous, homozygous *qmc551* zebrafish embryos were obtained from in-crossing *qmc551* heterozygous zebrafish which were sorting regarding the intensity of the GFP expression. WISH was performed on embryos at ~30 hpf. *lncRNA2* expression was investigated by WISH which revealed it is expressed in the vDA and the PBI of 30 hpf wt embryos (Figure 3.5Bi). In *qmc551* heterozygous embryos, this expression pattern is slightly higher than in the wt embryos (Figure 3.5Bii). In contrast, *qmc551* homozygous embryos display a significant increase in *lncRNA2* expression in the embryonic midline and the tail region (Figure 3.5Biii). A close-up image of the trunk reveals that *lncRNA2* expression is associated with the ventral aspect of the DA and PBI (Figure 3.5Biii). This is suggested that Gfi1aa repressed *lncRNA2* expression in the vDA and PBI of wt embryos.

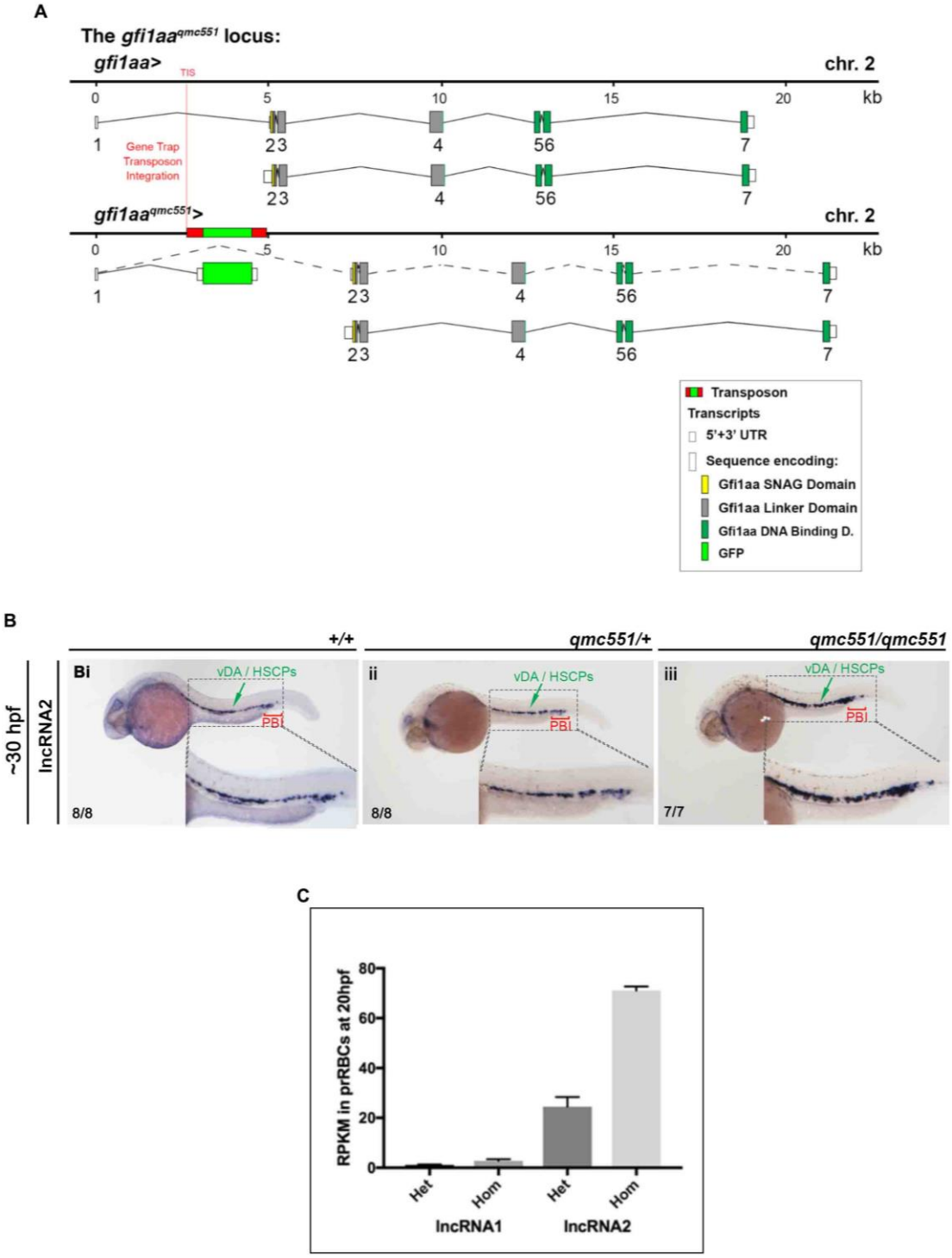


Figure 3.5 *lncRNA2* is upregulated in the vDA and PBI of *qmc551* homozygous mutants

(A) Genomic maps of wt and mutant of *gfi1aa* alleles, including information about transcripts and their splice products. **(B)** Expression of *lncRNA2* in ~30 hpf **(i)** wt, **(ii)** *qmc551* heterozygous, **(iii)** qmc551 homozygous and embryos detected by WISH. All embryos are presented in the lateral view with the anterior side facing left and the dorsal side up, and close up of the trunk to the tail region. Green arrow: endothelial cells of the vDA and

HSPCs, red bracket: the PBI. **(C)** The levels of *lncRNA1* and *lncRNA2* expression in prRBCs at 20 hpf in qmc551 heterozygous and homozygous. The expression levels are given in RPKMs (Moore, unpublished data). (Genomic map scheme obtained from Dr Martin Gering).

3.5 *lncRNA2* expression is lost in the onset of definitive haematopoietic cells in Gfi1aa/1ab double mutants

Gfi1aa expression is lost in primitive haematopoietic cells of homozygous *qmc551* fish, but definitive haematopoiesis is normal due to *gfi1ab* which is upregulated and substituted for Gfi1aa at the onset of definitive haematopoiesis (Thambyrajah et al., 2016). The expression of *lncRNA2* was investigated to find out whether combined loss of Gfi1aa/1ab has an impact on *lncRNA2* expression in definitive blood progenitors. *lncRNA2* expression was examined at 30 hpf by WISH. For this analysis, *qmc551;sh320* double homozygous embryos were obtained from in-crossing *qmc551hom;sh320het* zebrafish. The *sh320* line is a *gfi1ab* mutant allele, created by Dr Robert Wilkinson's lab in Sheffield, consisting of a 4 bp sequence insertion into exon 2 of the *gfi1ab* locus on chromosome 6 (Figure 3.6A). This insertion results in a frameshift in the *gfi1ab* coding sequence. The *lncRNA2* expression is lost in the vDA. A close up of the trunk of a *qmc551;sh320* double homozygous embryo showed that *lncRNA2* expression is not found in the vDA (Figure 3.6B). The gDNA was extracted from the double mutants embryos to identify the *sh320* allele. In the PCR reaction, oligos of DB1219 and DB1220 were used to amplify a ~260 bp fragment. Sanger DNA sequencing using primer DB393 confirmed that embryos are homozygous carriers of $gfi1ab^{sh320}$ alleles. The $gfi1ab^{sh320}$ allele sequence revealed a 4 bp insertion (Figure 3.6C). These suggested that combined loss of Gfi1aa/1ab has a defect on *lncRNA2* expression in definitive blood progenitors.

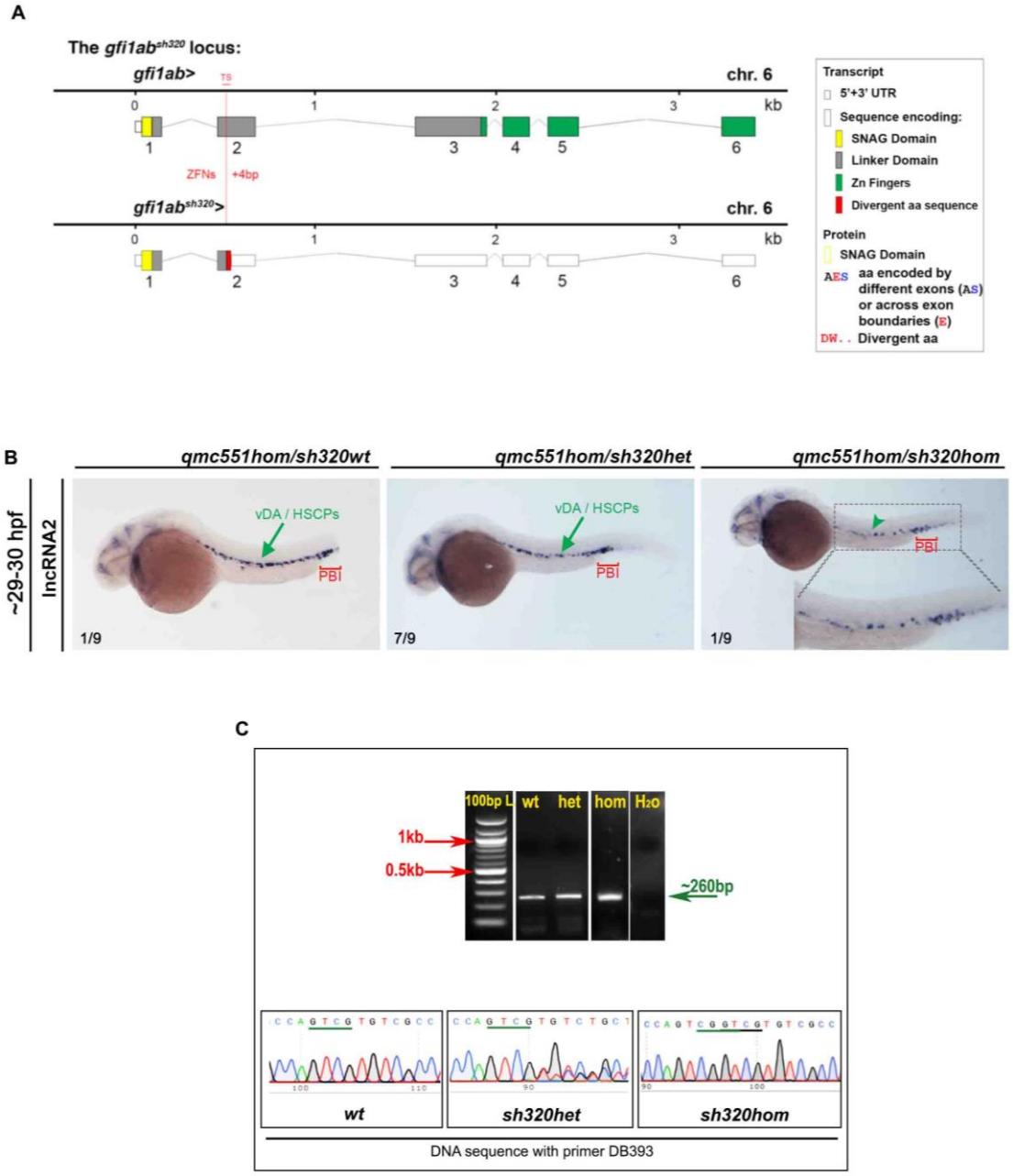


Figure 3.6 *lncRNA2* gene expression was lost in the definitive blood cells of gfi*1aa/gfi1ab* double mutants

(A) Genomic maps of wt and mutant of *gfi1ab* alleles, including information about transcripts and their splice products. **(B)** Expression of *lncRNA2* in ~30 hpf *qmc551hom/sh320hom* embryos detected by WISH. The embryo is presented in the lateral view with the anterior side facing left and the dorsal side up, and close up of the trunk to the tail region. Red bracket: the PBI and green arrowhead: missing expression in the vDA and HSPCs. **(C)** PCR amplification of gDNA extracted from *qmc551hom/sh320hom* WISH embryos to distinguish *wt* and *sh320* alleles. The panels show Sanger DNA sequencing reads of the template strands for *wt* and *sh320* alleles. The coding sequence has been underlined. (Genomic map scheme obtained from Dr Martin Gering).

3.6 *lncRNA1* expression is depleted in the vDA and adult haematopoietic cells of *runx1* homozygous zebrafish embryos

The WISH data described above showed that *lncRNA1* expression increases during the definitive wave of zebrafish haematopoiesis. To verify lncRNA expression in definitive blood cells, WISH analysis was performed on *runx1* zebrafish mutants. The *runx1*^{W84x} mutant allele harbours a G to A nucleotide substitution in exon 2 on chromosome 1 of the zebrafish genome that leads to a premature stop codon and the production of a non-functional truncated Runx1 protein (Figure 3.7A) (Sood et al., 2006). In the wt embryo, *runx1* is first expressed in the prRBCs at 12 hpf. At 26 hpf, runx1 expression is strong in the vDA (Kalev-Zylinska et al., 2002). The *runx1* mutant embryos display defects in the formation of definitive HSPC and its derivatives but no observable effects on primitive haematopoiesis (Jin et al., 2009, Gering and Patient, 2005, Burns et al., 2005, Sood et al., 2006, Kalev-Zylinska et al., 2002).

lncRNA1 expression was examined at around 28 hpf using WISH analysis in *runx1*^{W84x} mutant embryos. For this analysis, *runx1* heterozygous and homozygous embryos were obtained from a *runx1*^{W84x} heterozygous incross. Analysis of the stained embryos revealed that 6 out of 23 embryos displayed severely reduced *lncRNA1* expression in the vDA but retained some expression in the PBI as well as in cells in circulation (Figure 3.7B). The gDNA was extracted from these embryos to examine whether they were homozygous carriers of the *runx1*^{W84x} allele. For the PCR reaction, oligos DB1457 and DB1458 were used to amplify a ~230 bp fragment that was subsequently sequenced using primer DB1459 (Figure 3.7C). The sequencing revealed that the six embryos were

all homozygous carriers. These suggested that lncRNA1 are expressed in dHSCs of zebrafish.

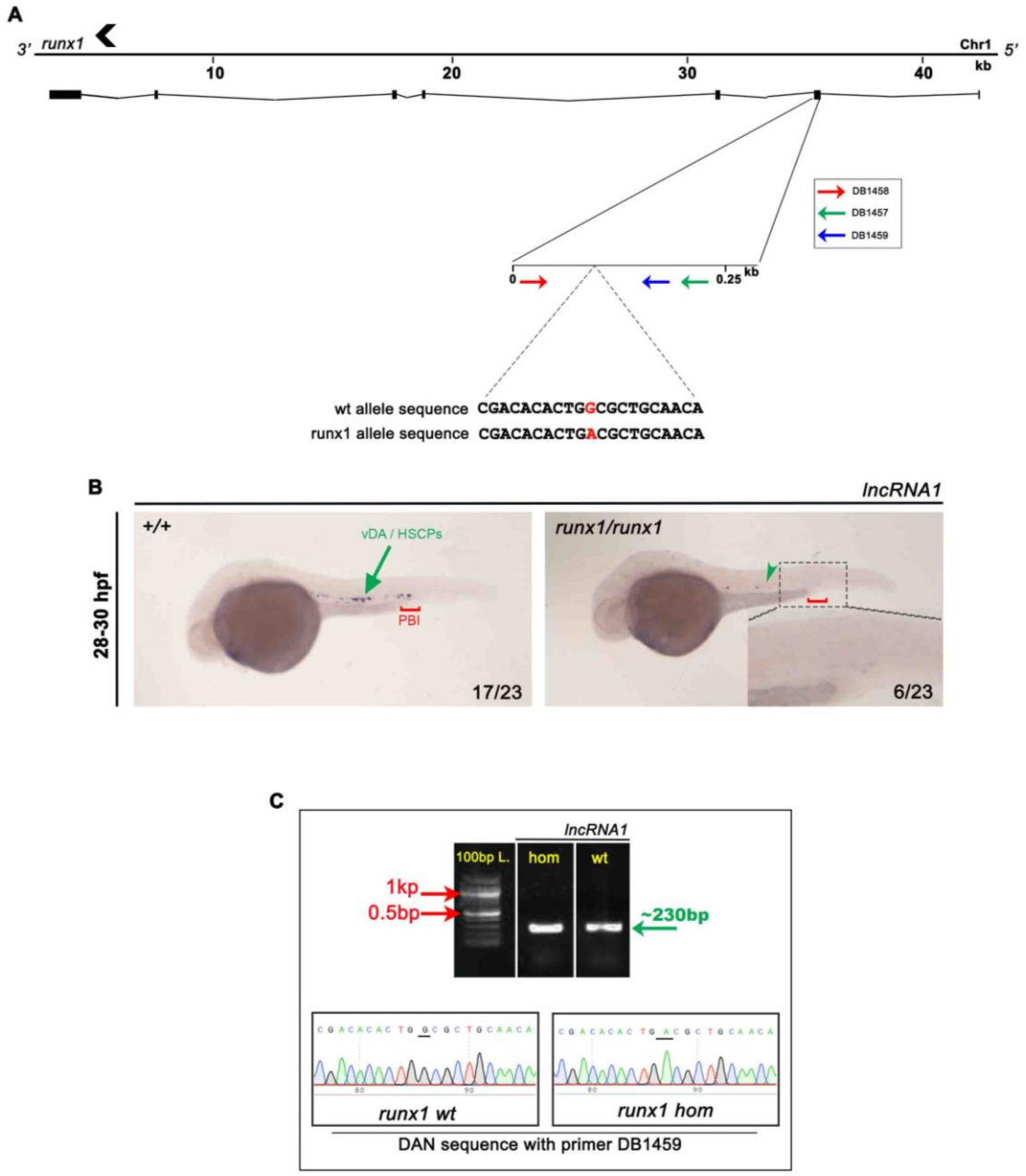


Figure 3.7 *lncRNA1* expression is strongly reduced in the vDA of *runx1* homozygous zebrafish embryos

(A) Schematic representation of the *runx1*^{w84x} allele at the *runx1* locus on zebrafish chromosome 1. The oligos used to PCR amplify the sequence around the mutation site for genotyping are shown in the box. **(B)** Lateral view of ~28-30 hpf embryos obtained from an incross of *runx1* heterozygous zebrafish and stained in a WISH using lncRNA1 probes. All embryos are presented in a lateral view with the anterior side facing left and the dorsal side up. Green arrow: the vDA and HSCPs, red bracket: the PBI, green arrowhead: strongly reduced in the vDA and HSCPs, red arrowhead: missing expression in the PBI. The numbers of genotyped embryos over the total numbers of embryos analysed are provided in each panel. **(C)** PCR amplifications were performed on gDNA extracted from WISH embryos to distinguish *wt* and *runx1*^{w84x} alleles. The panels show

Sanger DNA sequencing reads of the template strand of *wt* and *runx1*^{w84x} alleles. The coding sequence has been underlined.

3.7 lncRNA1 gene expression was observed in T cells of zebrafish embryos

To investigate in what types of HSCs express the *lncRNA1* transcript, WISH analysis was performed in the progeny of *cmyb*^{t25127} obtained from *cmyb*^{t25127} heterozygous incross. The *cmyb*^{t25127} allele is a T to A point mutation in exon 6 at the *cmyb* locus on zebrafish chromosome 23 (Figure 3.8A). It represents an isoleucine to asparagine missense mutation in the DNA binding domain of the Cmyb proteins. Mutant proteins have been shown to fail to bind its DNA target sequence. *cmyb* mutant embryos form normal numbers of prRBCs but have a severe defect in definitive haematopoiesis, including a loss of thymocytes (Soza-Ried et al., 2010).

WISH was performed on five day-old embryos derived from an incross of *cmyb*^{t25127} heterozygous carriers. *lncRNA1* expression was depleted in the thymus in 5 out of 20 embryos (Figure 3.8B). The gDNA was extracted from these embryos to examine their *cmyb* genotype. Nested PCR reactions that involved an external set of oligos (DB827 and DB828) and then a nested pair of oligos (DB980 and DB981) were used to amplify a ~280 bp fragments. These fragments were subsequently sequenced using primer DB1446 (Figure 3.8D). The DNA sequences revealed that embryos that lacked thymic lncRNA1 expression were homozygous for the *cmyb*^{t25127} allele.

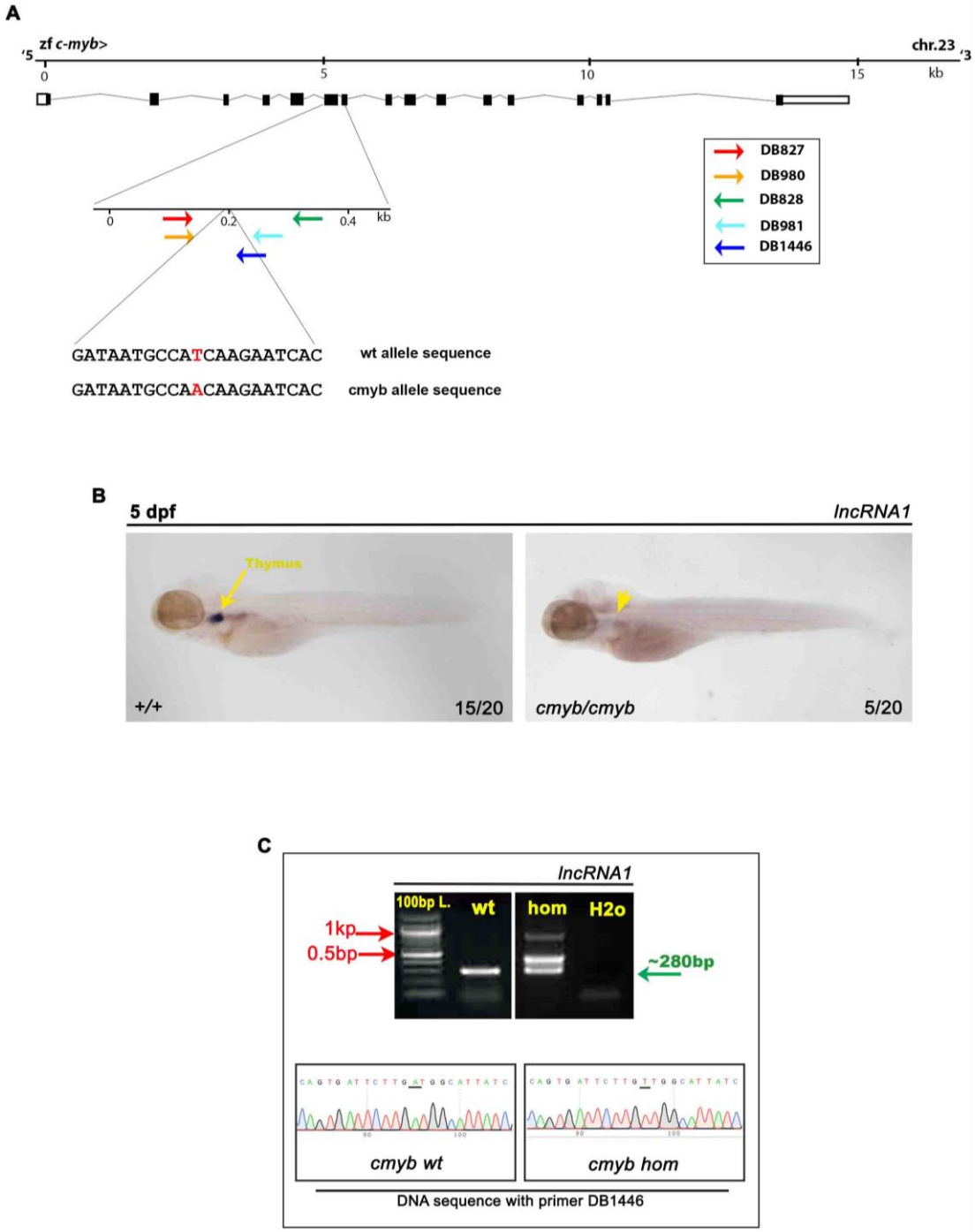


Figure 3.8 *lncRNA1* gene expression is lost due to loss of the T cells of cmyb mutant embryos

(A) Schematic representation of the *cmyb*^{t25127} allele at the *cmyb* locus on zebrafish chromosome 23. The oligos used to PCR amplify the sequences around the mutation site for genotyping are shown in the box. **(B)** Lateral view of 5 dpf embryos obtained from an incross of *cmyb* heterozygous zebrafish and stained in a WISH using lncRNA1 probes. All embryos are presented in a lateral view with the anterior side facing left and the dorsal side up. Yellow arrow: thymus. Yellow arrowhead: missing expression in T cells in the thymus. The numbers of embryos with phenotypes over the

total numbers of embryos analysed are provided in each panel. **(C)** PCR amplification on gDNA extracted from WISH embryos to distinguish wt and *cmyb*^{t25127} alleles. The panels show Sanger DNA sequencing reads of the template strand of the wt and *cmyb*^{t25127} alleles. The coding sequence has been underlined. (Genomic map scheme was provided by PhD student Jawaher Almulhim).

Summary

- The close inspection of lncRNA genes' expression patterns by RNA in situ hybridisation in wt embryos showed that both *lncRNA1* and *lncRNA2* are expressed in primitive and definitive blood cells and that the level of *lncRNA2* expression is considerably higher than that of *lncRNA1*.
- *Cloche* mutant embryos that lack blood, endothelial and endocardial cells showed no expression of either *lncRNA*1 or lncRNA2, confirming that both genes are indeed expressed in haematopoietic cells.
- *Mib* mutant embryos generate primitive but not definitive blood cells. Consequently, the specific loss of *lncRNA1* and *lncRNA2* expression in definitive, but not primitive blood cells confirmed lncRNAs expression during both stages of haematopoiesis.
- Analysis of *lncRNA1* and *lncRNA2* expressions patterns in different mutant zebrafish embryos confirmed that *lncRNA1* and *lncRNA2* are expressed in both primitive and definitive blood cells and suggested that lncRNA2 expression is highest at immature stages while the lncRNA1 is highest at later stages.
- *lncRNA2* is upregulated in the vDA and PBI of *qmc551* homozygous embryos, suggesting that *lncRNA2* could rescue primitive and definitive haematopoiesis in the absence of Gfi1aa
- Loss of expression of *lncRNA2* in the combined loss of Gfi1aa/1ab suggests that *lncRNA2* expression in the onset of definitive haematopoietic cells.
- *lncRNA1* expression in *runx1* homozygous embryos is significantly reduced in the vDA and slightly retained some

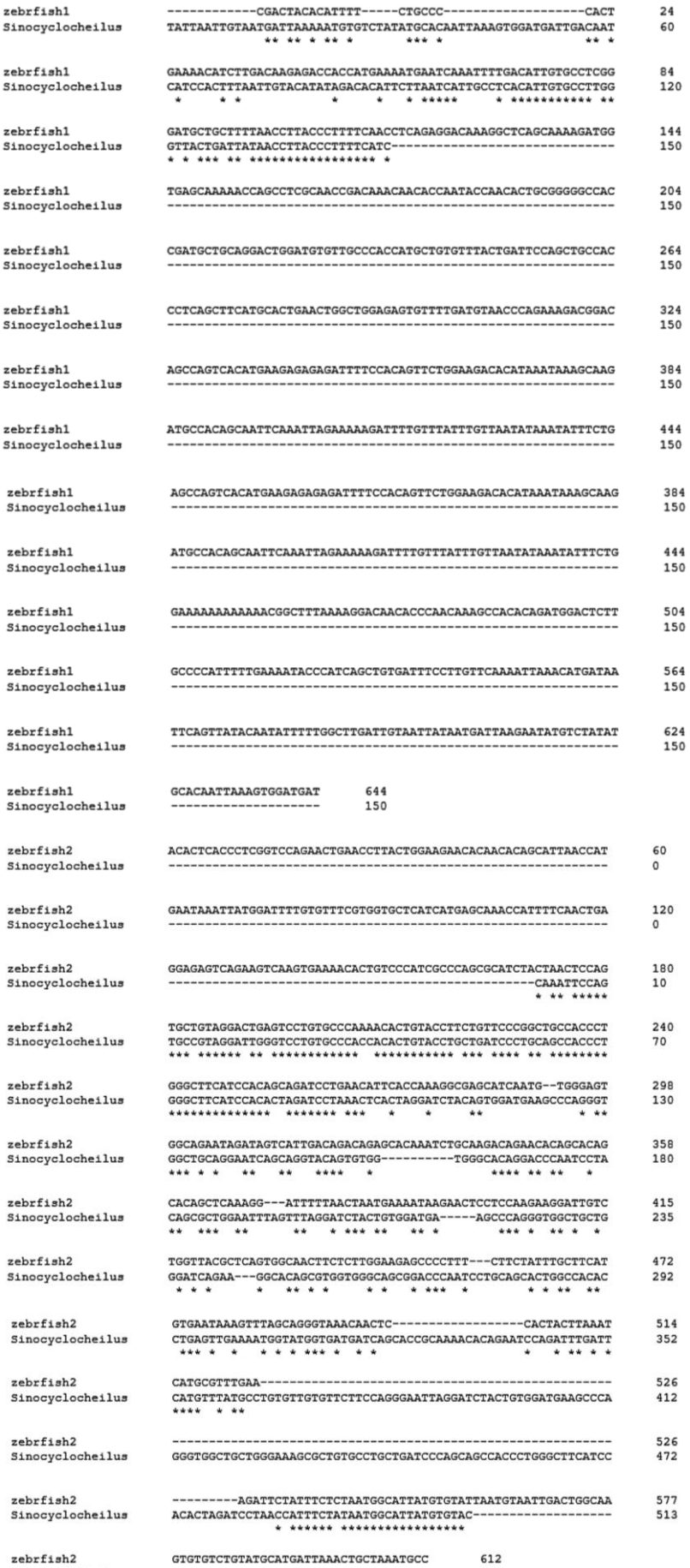
expression in the PBI as well as in cells in circulation this suggested that lncRNA1 is expressed in definitive haematopoietic cells.

- *lncRNA1* expression is missing in *cmyb* mutant zebrafish embryos suggested that the *lncRNA1* is expressed in T-cell progenitors in the thymus.

3.8 lncRNA1 and lncRNA2 have homologues in two other members of the carp family of freshwater fish but not in other vertebrate species

To see whether other vertebrate species carry sequences homologous the zebrafish lncRNAs (si:ch211-214p16.1-201 (lncRNA1) and si:ch211-214p16.2-201 (lncRNA2)), blastn searches of the non-redundant DNA sequences database were performed. The lncRNAs genes are not conserved in all vertebrates but they do have homologues in a two genus of carp (Cyprinidae) freshwater fish family; sinocyclocheilus and cyprinus carpio (NCBI September/2019)). The comparative alignment analysis identifies some shared sequences between zebrafish and carp (

Figure 3.9). However, the vertebrates do not contain *lncRNA1* and *lncRNA2* genes such as human and mouse, the comparative alignment analysis identifies shared alignment regions surround lncRNA genes between zebrafish and other organisms; (Figure 3.10 and Figure 3.11, respectively).


zebrfish1 ----------CGACTACACATTTT-----CTGCCC----------CACT 24
Sinocyclocheilus TATTAATTGTAATGATTAAAAATGTGTCTATATGCACAATTAAAGTGGATGATTGACAAT 60
zebrfish1 GAAAACATCTTGACAAGAGACCACCATGAAAATGAATCAAATTTTGACATTGTGCCTCGG 84
Sinocyclocheilus CATCCACTTTAATTGTACATATAGACACATTCTTAATCATTGCCTCACATTGTGCCTTGG 120
zebrfish1 GATGCTGCTTTTAACCTTACCCTTTTCAACCTCAGAGGACAAAGGCTCAGCAAAAGATGG 144
Sinocyclocheilus GTTACTGATTATAACCTTACCCTTTTCATC---------- 150
zebrfish1 TGAGCAAAAACCAGCCTCGCAACCGACAAACAACACCAATACCAACACTGCGGGGGCCAC 204
Sinocyclocheilus ---------- 150
zebrfish1 CGATGCTGCAGGACTGGATGTGTTGCCCACCATGCTGTGTTTACTGATTCCAGCTGCCAC 264
Sinocyclocheilus ---------- 150
zebrfish1 CCTCAGCTTCATGCACTGAACTGGCTGGAGAGTGTTTTGATGTAACCCAGAAAGACGGAC 324
Sinocyclocheilus ---------- 150
zebrfish1 AGCCAGTCACATGAAGAGAGAGATTTTCCACAGTTCTGGAAGACACATAAATAAAGCAAG 384
Sinocyclocheilus ---------- 150
zebrfish1 ATGCCACAGCAATTCAAATTAGAAAAAGATTTTGTTTATTTGTTAATATAAATATTTCTG 444
Sinocyclocheilus ---------- 150
zebrfish1 AGCCAGTCACATGAAGAGAGAGATTTTCCACAGTTCTGGAAGACACATAAATAAAGCAAG 384
Sinocyclocheilus ---------- 150
zebrfish1 ATGCCACAGCAATTCAAATTAGAAAAAGATTTTGTTTATTTGTTAATATAAATATTTCTG 444
Sinocyclocheilus ---------- 150
zebrfish1 GAAAAAAAAAAAACGGCTTTAAAAGGACAACACCCAACAAAGCCACACAGATGGACTCTT 504
Sinocyclocheilus ---------- 150
zebrfish1 GCCCCATTTTTGAAAATACCCATCAGCTGTGATTTCCTTGTTCAAAATTAAACATGATAA 564
Sinocyclocheilus ---------- 150
zebrfish1 TTCAGTTATACAATATTTTTGGCTTGATTGTAATTATAATGATTAAGAATATGTCTATAT 624
Sinocyclocheilus ---------- 150
zebrfish1 GCACAATTAAAGTGGATGAT 644
Sinocyclocheilus ---------- 150
zebrfish2 ACACTCACCCTCGGTCCAGAACTGAACCTTACTGGAAGAACACAACACAGCATTAACCAT 60
Sinocyclocheilus ---------- 0
zebrfish2 GAATAAATTATGGATTTTGTGTTTCGTGGTGCTCATCATGAGCAAACCATTTTCAACTGA 120
Sinocyclocheilus ---------- 0
zebrfish2 GGAGAGTCAGAAGTCAAGTGAAAACACTGTCCCATCGCCCAGCGCATCTACTAACTCCAG 180
Sinocyclocheilus ----------CAAATTCCAG 10
zebrfish2 TGCTGTAGGACTGAGTCCTGTGCCCAAAACACTGTACCTTCTGTTCCCGGCTGCCACCCT 240
Sinocyclocheilus TGCCGTAGGATTGGGTCCTGTGCCCACCACACTGTACCTGCTGATCCCTGCAGCCACCCT 70
zebrfish2 GGGCTTCATCCACAGCAGATCCTGAACATTCACCAAAGGCGAGCATCAATG--TGGGAGT 298
Sinocyclocheilus GGGCTTCATCCACACTAGATCCTAAACTCACTAGGATCTACAGTGGATGAAGCCCAGGGT 130
zebrfish2 GGCAGAATAGATAGTCATTGACAGACAGAGCACAAATCTGCAAGACAGAACACAGCACAG 358
Sinocyclocheilus GGCTGCAGGAATCAGCAGGTACAGTGTGG----------TGGGCACAGGACCCAATCCTA 180
zebrfish2 CACAGCTCAAAGG---ATTTTTAACTAATGAAAATAAGAACTCCTCCAAGAAGGATTGTC 415
Sinocyclocheilus CAGCGCTGGAATTTAGTTTAGGATCTACTGTGGATGA-----AGCCCAGGGTGGCTGCTG 235
zebrfish2 TGGTTACGCTCAGTGGCAACTTCTCTTGGAAGAGCCCCTTT---CTTCTATTTGCTTCAT 472
Sinocyclocheilus GGATCAGAA---GGCACAGCGTGGTGGGCAGCGGACCCAATCCTGCAGCACTGGCCACAC 292
zebrfish2 GTGAATAAAGTTTAGCAGGGTAAACAACTC----------CACTACTTAAAT 514
Sinocyclocheilus CTGAGTTGAAAATGGTATGGTGATGATCAGCACCGCAAAACACAGAATCCAGATTTGATT 352
zebrfish2 CATGCGTTTGAA---------- 526
Sinocyclocheilus CATGTTTATGCCTGTGTTGTGTTCTTCCAGGGAATTAGGATCTACTGTGGATGAAGCCCA 412
zebrfish2 ---------- 526
Sinocyclocheilus GGGTGGCTGCTGGGAAAGCGCTGTGCCTGCTGATCCCAGCAGCCACCCTGGGCTTCATCC 472
zebrfish2 ---------AGATTCTATTTCTCTAATGGCATTATGTGTATTAATGTAATTGACTGGCAA 577
Sinocyclocheilus ACACTAGATCCTAACCATTTCTATAATGGCATTATGTGTAC---------- 513
zebrfish2 GTGTGTCTGTATGCATGATTAAACTGCTAAATGCC 612
Sinocyclocheilus ---------- 513

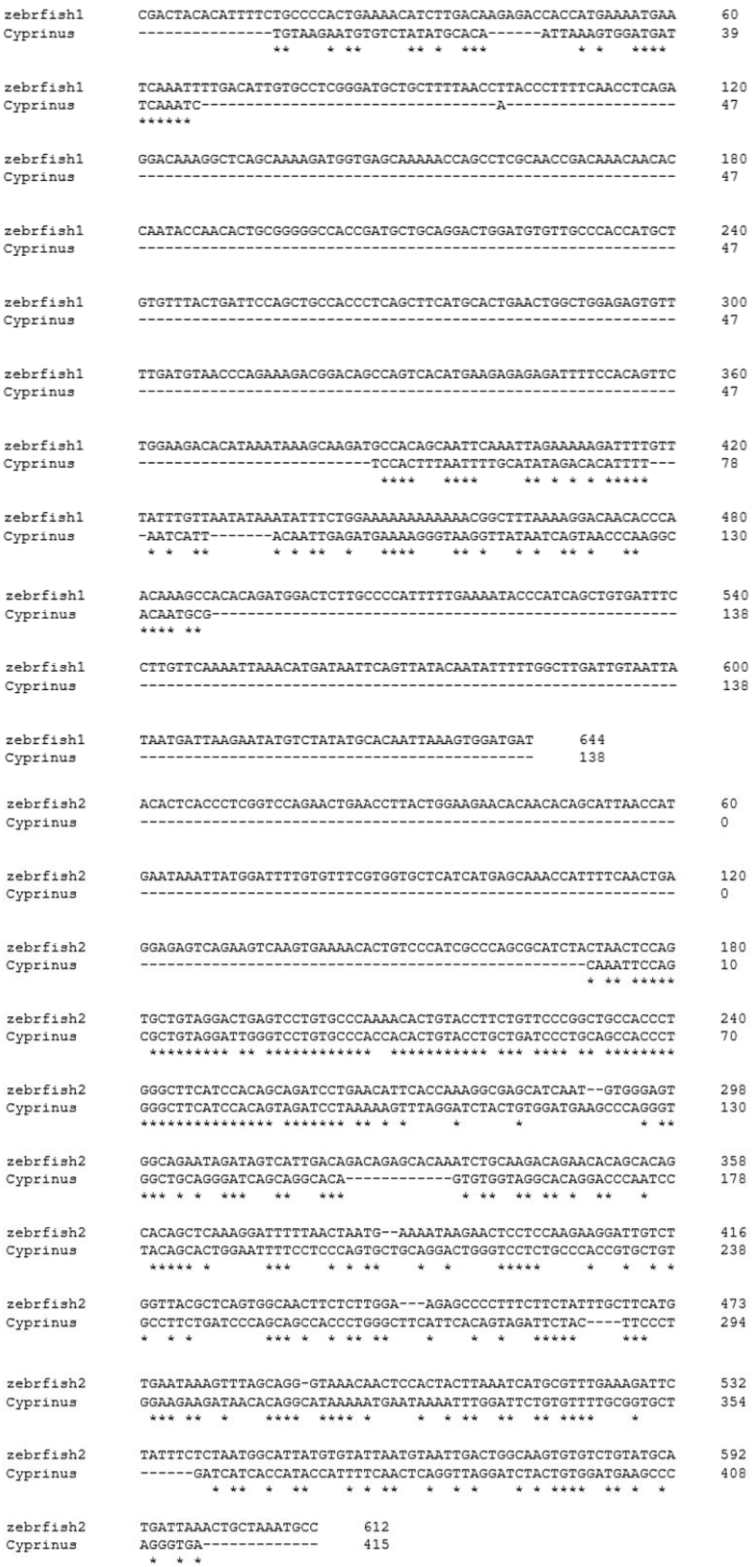


Figure 3.9 The alignment of lncRNA1 and lncRNA2 transcripts between zebrafish and members of the carp family of freshwater fish

The sequence of the lncRNAs zebrafish on the top and the other fish are on the bottom

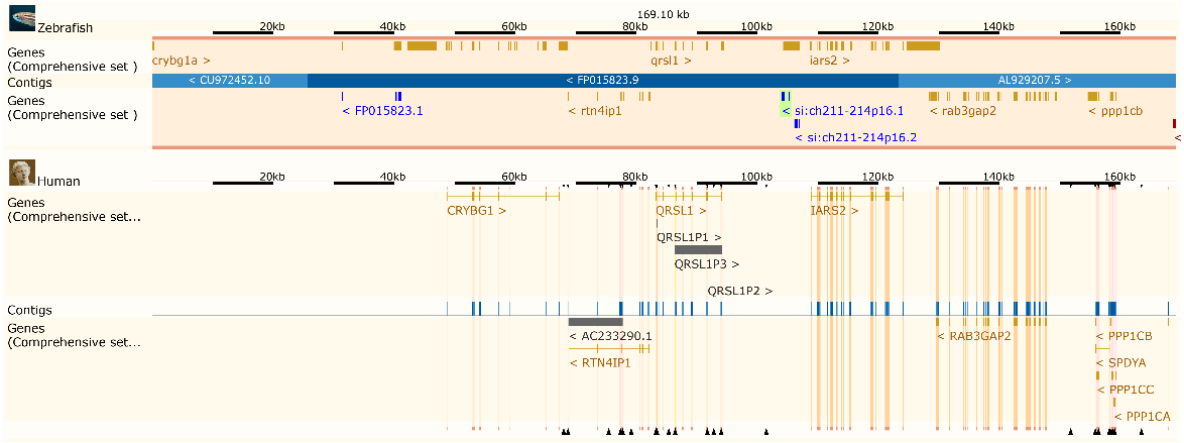


Figure 3.10 Conservation shared synteny genes human and zebrafish

The diagram illustrates the synteny of zebrafish to human and genomic region. On the left side they shared QRSL1, RTN4IP1 & CRYBG1 genes. On the right side they shared IARS2, RAB3GAP2, & PPP1CB genes (GRCz11 September/2019).

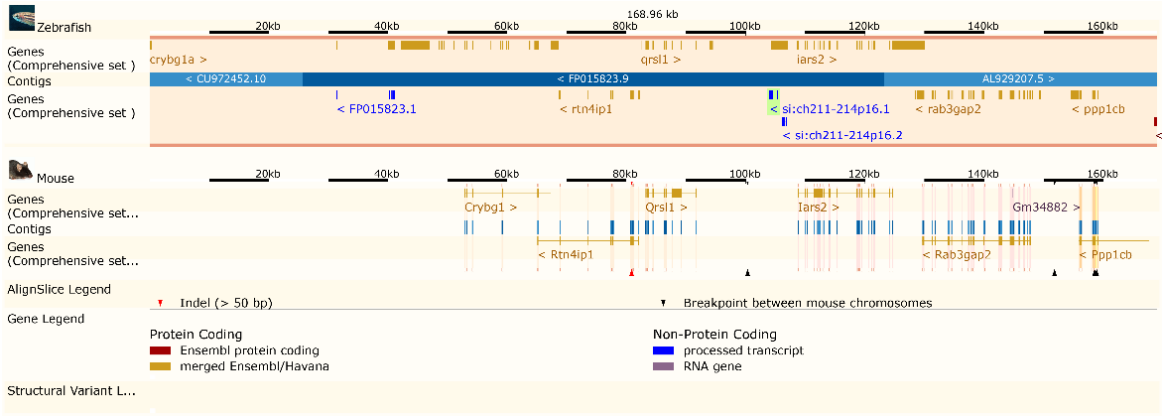


Figure 3.11 Conservation shared synteny genes mouse and zebrafish

The diagram illustrates the synteny of zebrafish to mouse and genomic region. On the left side they shared *QRSL1, RTN4IP1 & CRYBG1* genes. On the right side they shared *IARS2, RAB3GAP2, & PPP1CB* genes (GRCz11 September/2019).

3.9 The genomic map of *lncRNA1* and *lncRNA2* of wt zebrafish

The *lncRNA1* and *lncRNA2* are neighbouring genes, located on chromosome 17 of the *Danio rerio* genome. Each gene is made up of 2 exons. The two genes are transcribed in the same direction. Given the genes' interesting expression patterns we wanted to know the role of the genes. To identify their role, the genes were meant to be deleted from the genome, individually and in combination. For this purpose, target sites were defined and guide RNAs were designed and generated. The success of the experiment relied on having an accurate DNA sequence of the region of interest. To verify that the genomic DNA sequence of the lncRNA genes of our wt fish corresponded to that in the reference genome. Sections of the region of interest were amplified and sequenced to confirm the reference sequence and to identify polymorphisms (Figure 3.12). All the primers that were used to amplify sections of the whole region are listed in Table 3.1. The resulting sequence was blasted against the zebrafish genome (GRCz11 July/2019). Some single nucleotide polymorphisms (SNPs) were identified in the region of interest. It was 77 presenting SNPs with the reference zebrafish genome has been found around the amplified region (exons and introns) (Appendix A).

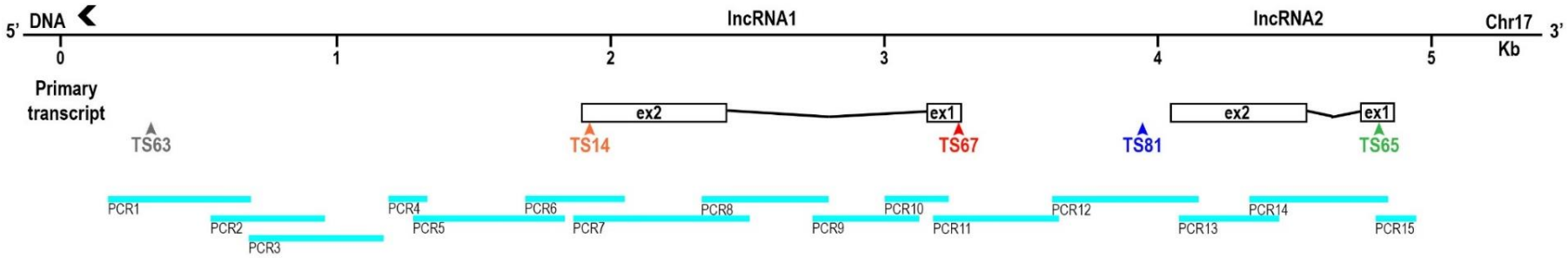


Figure 3.12 Mapping of *lncRNA1* and *lncRNA2* wt zebrafish genes

Schematic shows the structure the lncRNAs locus on chromosome 17 of the zebrafish genome. The position of the target site for the CRISPR/Cas9 system is shown in different colours. The PCR products are depicted in light blue.

Table 3.1 illustrates sets of primers used in PCR amplification (GRCz11)

Gering Lab Database (DB) Number	Sequence 5'→3'	PCR Number on Map
DB1388 F	GGGGTCAAACACACTGCAAA	PCR1
DB1327 R	CGGAACTCCTTATGTCATGCA	
DB1388 F	GGGGTCAAACACACTGCAAA	PCR2
DB1403 R	AAAAAGGGGGAGGAGCTG	
DB834 F	CAGTTCATTTCAGTGCACCA	PCR3
DB833 R	CGAGTCACCACAGAACCAAA	
DB779 F	CATGATCGCACTTTGTTTGC	PCR4
DB884 R	ACTAAAGACAGAGACAAACCGA	
DB829 F	TGTCAGTTATTCCTGGATTAGGC	PCR5
DB830 R	TCAAACATGCAGCAGGAA	
DB831 F	TTTGCGCAAGGATTTAAATATG	PCR6
DB883 R	CCCAACAAAGCCACACAGAT	
DB682 F	TCATCCACTTTAATTGTGCATATAGA	PCR7
DB879 R	TGTAGGACAAAGGCTCAGCA	
DB882 F	CAACACATCCAGTCCTGCAG	PCR8
DB681 R	TTTTCTGCCCCACTGAAAAC	
DB1469 F	TCACCCAAATATGTCCTCACC	PCR9
DB924 R	GGGCAGAAAATGTGTAGTCG	
DB1396 F	GGAAACCTGCAGCCATTG	PCR10
DB924 R	CGACTACACATTTTCTGCCC	
DB878 F	CTCTGAGGTTGAAAAGGGTAAGG	PCR11
DB1393 R	AGAACCAGCTTTCGAGTCCA	
DB876 F	TGGACTCGAAAGCTGGTTCT	PCR12
DB665 R	GGGCTTCATCCACAGCAGAT	
DB750 F	GGCATTTAGCAGTTTAATCATGC	PCR13
DB665 R	GGGCTTCATCCACAGCAGAT	
DB907 F	TCTGCCACTCCCACATTGAT	PCR14
DB908 R	CACTCACCCTCGGTCCAGAAC	
DB832 F	ATGATGAGCACCACGAAACA	PCR15
DB874 R	GACCAATGGAGGAGGGGATT	

3.10 Deletion of *lncRNA* genes from the wild-type zebrafish genome using CRISPR/Cas9

Here, the CRISPR/Cas9 system was used to delete the *lncRNA* genes from the zebrafish genome in an attempt to uncover the functions of the *lncRNAs* genes during zebrafish haematopoiesis. Suitable target sites within the *lncRNA1/2* locus were identified using the CRISPRscan UCSC track tool (Moreno-Mateos et al., 2015) (Figure 3.12). Target sequences (TSs) are 20-bp sequences that lie adjacent to a PAM sequence. The sequence follows the pattern 5'-GG-N18-PAM-3'. The GG dinucleotide at the 5' end is needed for T7 enzyme which prefers a GG dinucleotide at the transcription start and transcription yields have been shown to be reduced if these two Guanines are absent. It has been shown that sgRNAs could tolerate 2-nt mismatches (denoted as GGN18 and GGN20) at their 5' ends with still conferring their efficiency (Hwang et al., 2013a, Moreno-Mateos et al., 2015). The two TSs were TS14 and TS65, located downstream of *lncRNA1* and upstream of *lncRNA2*, respectively (Figure 3.13A), with score values of 16 and 65, respectively. The target sequence is introduced into the BsaI digested the guide RNA expression vector pDR274 in the form of a DNA linker. The linker is generated by hybridising complementary DNA oligos that possess additional nucleotides at their 5' ends. These give the linker 5' overhangs that can be ligated with the sticky ends of the linearised guide RNA expression vector pDR274. Within the vector, a T7 promoter is located upstream of the guide RNA sequence that can be used for *in vitro* transcription of the guide RNA (Figure 3.13B & C). To ensure that a single size RNA product is generated, the vector is digested with DraI prior to *in vitro* transcription. This digest generates two linear DNA molecules, one of which carries that T7 promoter upstream of the guide RNA coding sequence. The ligation of annealed linkers into pDR74 transform into subcloning *E.coli*.

Successful cloning of the correct sequences was confirmed by Sanger sequencing (Figure 3.13D) (at Source Biosciences, Nottingham). All gRNAs were designed and generated using the same process in this project.

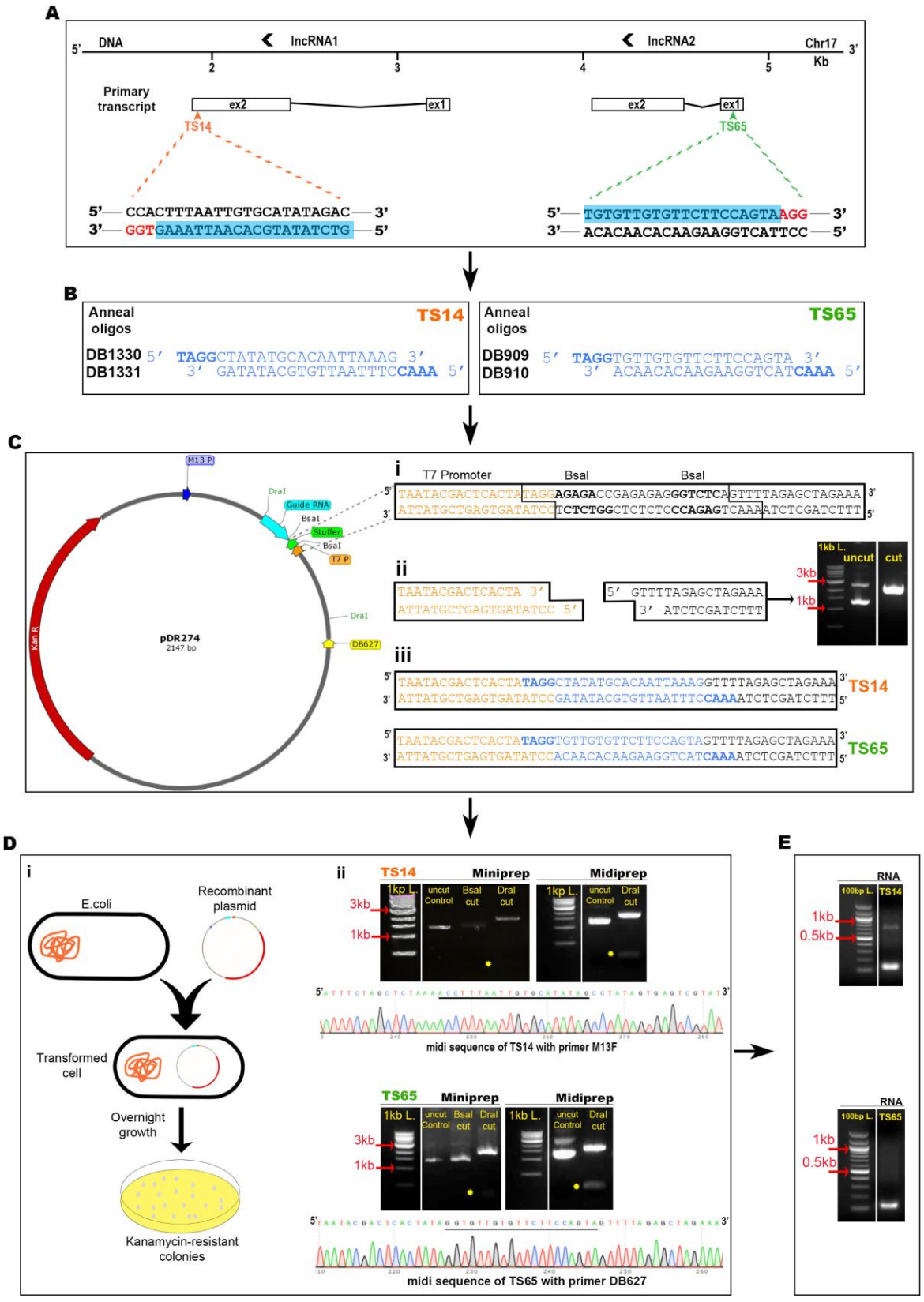


Figure 3.13 Generation of gRNAs to knockout *lncRNA1* and *lncRNA2* genes of zebrafish using the CRISPR/Cas9 system

(A) Schematics showing the location of targeted genomic sequences; the PAM sequence is highlighted in red and the TS-specific section of the gRNA

is highlighted in blue. **(B)** Complementary DNA oligos were designed that represent the target site sequences harbouring a T7 polymerase binding site and the GG dinucleotide which represents the start of the transcribed sequence (Bold) and annealed to make a double-stranded linker fragment. **(C)** Linker integration into guide RNA expression vector pDR274. (i) Plasmid pDR274 possesses two BsaI recognition sites (Bold) downstream of the T7 promoter (orange). (ii) Plasmid pDR274 was digested with BsaI, generating two 5' overhangs. The digest was confirmed by gel electrophoresis. (iii) The 5' overhangs of the linker were hybridised to the corresponding sequences of the BsaI digested pDR274 expression vector. The sequences were ligated by T4 DNA ligase. **(D)** Propagation of the recombinant plasmid in *E. coli*. (i) The ligation mix was used to transform *E. coli*. Kanamycin-resistant colonies were identified on Kanamycin-containing agar plates. (ii) Mini scale preparation of plasmid DNA was followed by BsaI and DraI restriction digests. Recombinant plasmids were prepared at a medium scale, digested with BsaI and DraI to confirm the presence of the linker. The plasmid DNAs were then sequenced using either primer M13F or DB627. Exemplary chromatograms are shown that highlight the presence of the target site sequence. **(E)** *In vitro* transcription. The digested DNA was used as a template in the *in vitro* transcription. RNA product was analyzed by gel electrophoresis.

To test both gRNAs in a single experiment, the co-injection of gRNAs and Cas9 was performed, and the efficiency of genomic deletions was evaluated using PCR amplification. The gRNAs (TS14 and TS65) were mixed with Cas9 protein and the fluorescent dye mini ruby. A volume of 1 nl of the mix was injected into 1-cell-stage wild-type (wt) zebrafish embryos. To investigate the success of the injection, the embryos were examined with fluorescent microscopy at approximately 4 hpf (Figure 3.14A). To investigate whether the deletion was successful, single embryos were genotyped at 24 hpf. First, gDNA was extracted from one uninjected embryo, which used as the control, and from ten injected embryos (Figure 3.14B). Their gDNA was used a template in nested PCR reactions with primers hybridised to sequences flanking the DNA fragment that was meant to be deleted from the genome. The first PCR involved the external oligos DB831 and DB778, the second the nested oligos DB1400 and DB874. Given the large distance between the primer binding sites in the wt sequence (> 3 kb), successful PCR amplification was expected only on mutant DNA templates that lacked the DNA sequence between the two target sites. The expected mutant PCR fragment was ~340 bp in size. Five out of ten of the injected embryos displayed a PCR product of that size. Similar products were not detected in samples of uninjected embryos. The DNA sequence of the embryo 7 PCR fragment confirmed the deletion of the DNA sequence between TS14 and TS65. The sequencing data revealed a 2901 bp deletion and a 2 bp insertion between the two TSs just downstream of both TS PAM regions (Figure 3.14B). These results indicated the successful targeting of the *lncRNA1* and *lncRNA2* locus in these embryos, and their siblings were raised. When the F0 injected embryos were ~3-4 months old, founders were to be identified that carried the lncRNA genes mutation in the germline. The potential founders were either incrossed or crossed with wt fish, and their progeny were analysed for the presence of a deletion allele.

The F0 injected fish did not appear to transmit the mutation to their offspring, and no founders have been found (Figure 3.14C). After 4 unsuccessful attempts (summarised in Table 3.2) other pairs of guides RNAs were tested.

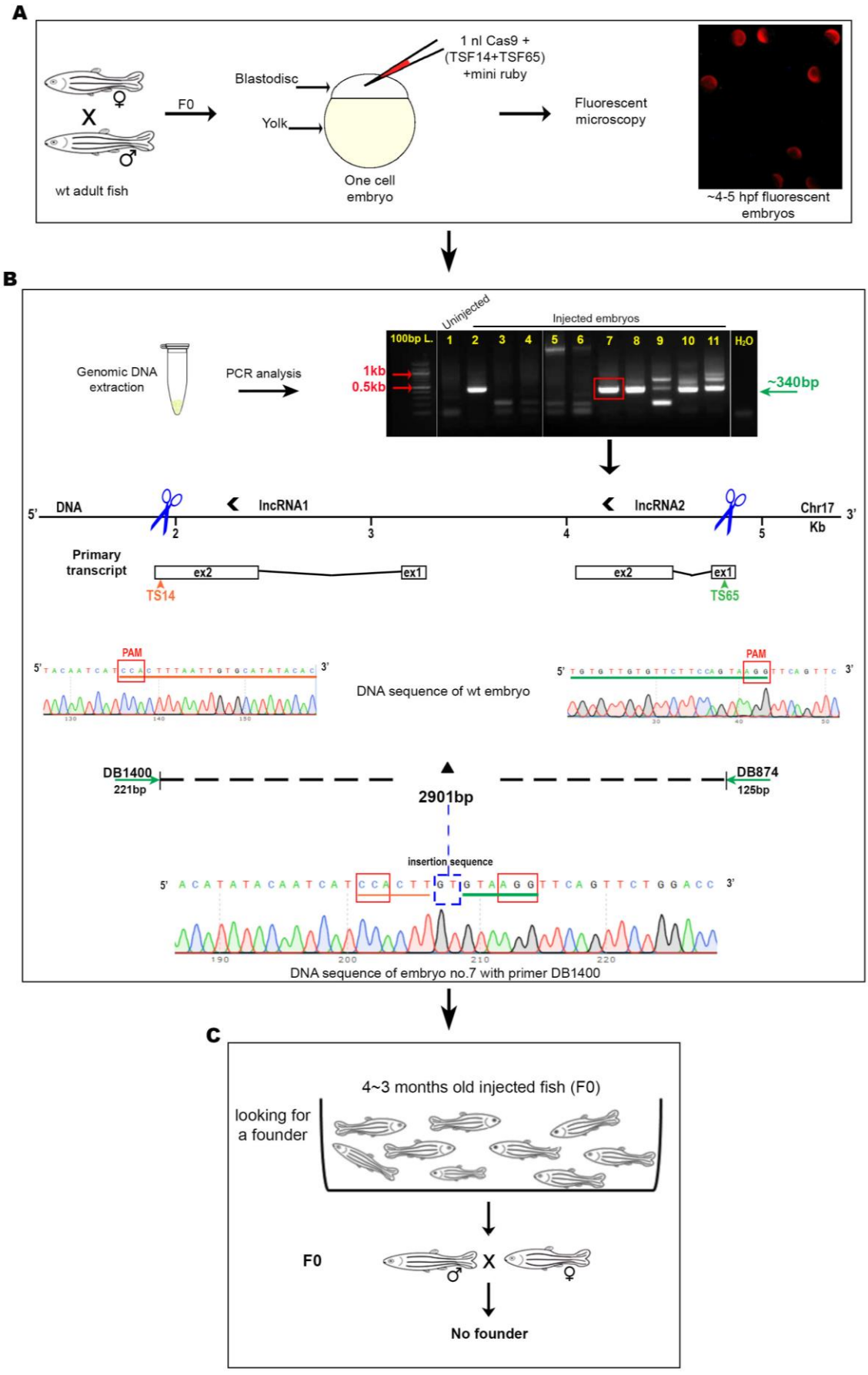

A
wt adult fish
F0
Blastodisc
Yolk
One cell embryo
1 nl Cas9 + (TSF14+TSF65) +mini ruby
Fluorescent microscopy
~4-5 hpf fluorescent embryos
B
Genomic DNA extraction
PCR analysis
Uninjected
Injected embryos
100bp L.
1kb
0.5kb
~340bp
DNA
lncRNA1
lncRNA2
Chr17
Kb
Primary transcript
ex2
ex1
TS14
TS65
PAM
DNA sequence of wt embryo
DB1400
221bp
DB874
125bp
2901bp
insertion sequence
DNA sequence of embryo no.7 with primer DB1400
C
4~3 months old injected fish (F0)
looking for a founder
F0
No founder

Figure 3.14 The CRISPR/Cas9 system was used to delete the *lncRNA1* and *lncRNA2* genes from zebrafish genome

(A) Embryos derived from a cross of wt fish were injected with Cas9, guide RNAs 65 and 14 and the fluorescent dye mini ruby. Embryos that displayed red fluorescence at ~4-5 hpf were grown up o/n. **(B)** At 24 hpf, genomic DNA was extracted from a subset of embryos. The region of interest was PCR amplified using primers that flank the deleted region. The primers used were DB1400 and DB874. The PCR product was then separated on an agarose gel to determine whether the gene editing was successful. The mutant PCR product was expected to have a size of ~340bp. The PCR product of embryo 7 was extracted from the gel and sequenced with oligo DB1400. The sequence confirmed the amplification of the expected mutant allele (embryo 7, highlighted in red). As this indicated successful targeting of the lncRNA1 and lncRNA2 locus in these embryos, their siblings were raised. **(C)** Check offspring for germ transmission. Once F0 injected were ~3-4 months old, the fish were incrossed or crossed to wt fish and their progeny were analysed for the presence of a deletion allele using the same nested PCR protocol. The injected F0 fish did not transmit the mutations to their offspring.

Table 3.2 The attempts to knockout *lncRNA1* and *lncRNa2* from wt zebrafish genome by injected with TS14 and TS65

Injection	No.1	No.2**.**	No.3	No.4
Stock No	SN743	SN798	SN838	SN843
TS concentrations	TS14 ~78 pg TS65 ~91 pg	TS14 ~100 pg TS65 ~100.6 pg	TS14 ~98 pg TS65 ~110 pg	TS14 ~98 pg TS65 ~110 pg
Total injected	126	68	63	114
Genotyped embryos number	10	10	18	5
Embryos displayed the mutant PCR band	7/10	5/10	9/18	4/5
Surviving injected embryos number	89	43	29	56
Fish examined number	38	29	7	27
Embryos were looked at for each of the founders	20	20	20	20
Founders	0	0	0	0

3.11 Using CRISPR/Cas9 to delete lncRNA genes from *qmc551/qmc554* double mutant zebrafish

The same pair of guide RNAs was also injected with Cas9 into embryos that were heterozygous for mutant alleles of *gfi1aa* and *gfi1b*. The rationale for this experiment was based on the finding that (a) *lncRNA1* and *lncRNA2* are expressed in prRBCs and (b) that their expression was subject to direct or indirect regulation by Gfi1aa and Gfi1b. This suggested that the lncRNAs are members of the gene regulatory network that controls the differentiation of prRBCs. It was therefore of interest to know whether there was a genetic interaction between the Gfi1aa/1b repressors and the two *lncRNA*s. To speed up the generation of quadruple mutants, the CRISPR/Cas9 experiment was performed in *qmc551*;*qmc554* double heterozygous embryos. The *qmc554* line is a *gfi1b* mutant allele, created in our lab by Dr Chris Moore. The *gfi1b*^{qmc554} allele a CRISPR/Cas9-induced deletion, and a frameshift mutation in exon 4. Moore and his colleagues (2018) showed that Gfi1aa and Gfi1b in zebrafish suppress endothelial gene expression and promote late erythroid gene expression in immature primitive erythroblasts. Therefore, mutations in both genes result in zebrafish with delayed differentiation of primitive erythroblast from mesenchymal haemangioblasts, although differentiation is not completely blocked (Moore et al., 2018).

A 1 nl mix containing gRNAs (TS14 and TS65), Cas9 and the fluorescent dye mini ruby was injected into the 1-cell stage of *qmc551;qmc554* heterozygous embryos, obtained from an outcross of wt male with *qmc551;qmc554* homozygous female fish. At approximately 4 hpf, the embryos were examined by fluorescent microscopy to confirm the success of the injection (Figure 3.15A). To determine whether the region of interest was deleted

successfully, single embryos were genotyped. gDNA was extracted from 1 uninjected embryo and four injected embryos (Figure 3.15B). Then, the region of interest, which indicated the presence of a genomic deletion between the two target sites, was PCR amplified on genomic DNA using the same oligos that were described above. The expected deleted PCR fragment is ~340 bp in size. Uninjected embryos displayed the expected wt fragment, which was 2901 bp in size and another unexpected fragment; however, they did not contain the deleted fragment. Three of the four injected embryos displayed the deleted fragment, which is ~340 bp. To confirm the mutation, the amplified bands from embryo 5 were sequenced. The sequencing data showed that the fragment represented an approximately 2900 bp deletion just downstream of both TSs and an 18 bp insertion between the two TSs (Figure 3.15B). This result indicated the successful targeting of the *lncRNA1* and *lncRNA2* locus in these *qmc551;qmc554* double heterozygous embryos. Their siblings were raised to adulthood.

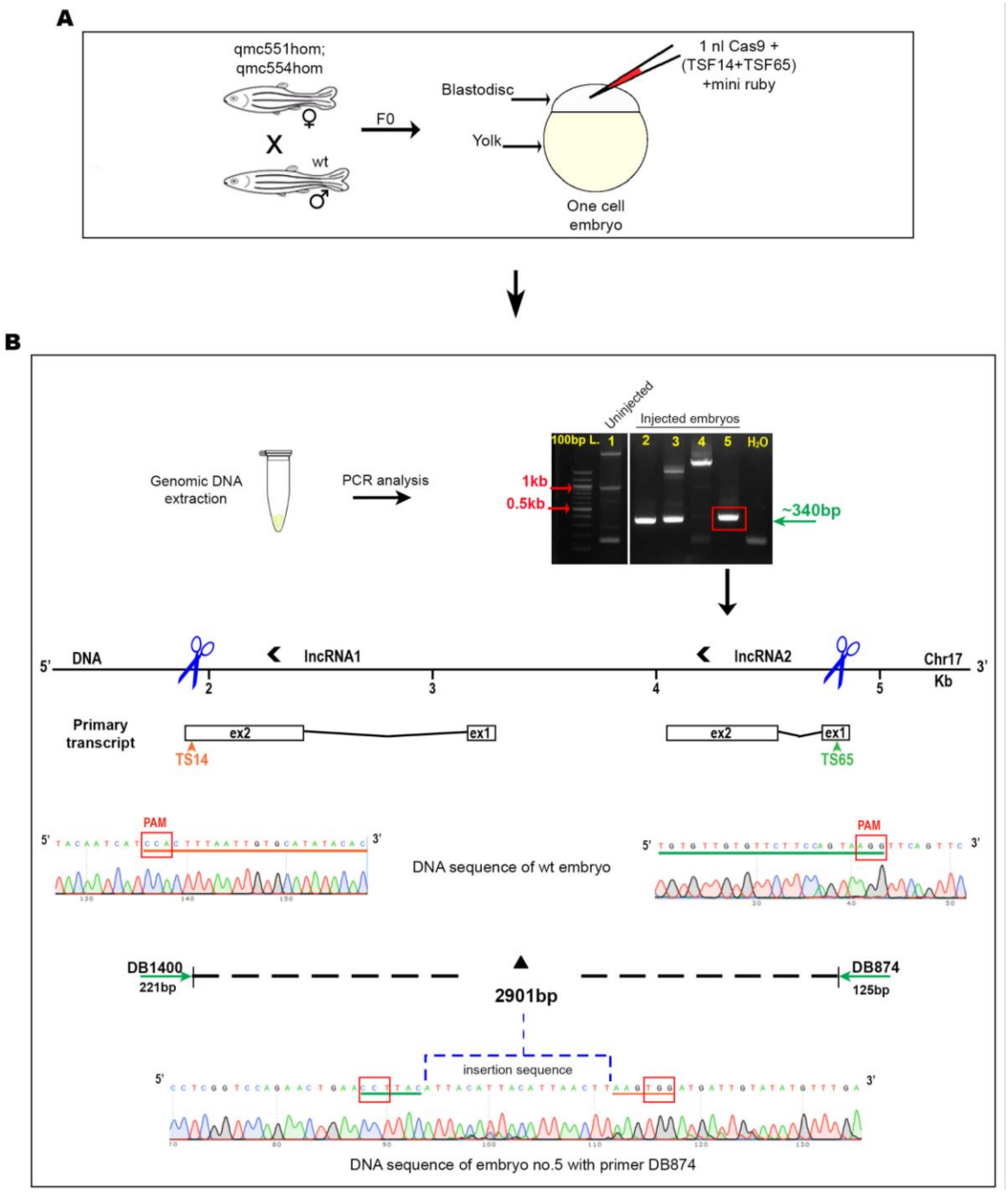


Figure 3.15 CRISPR/Cas9 system was used to delete the *lncRNA* genes from qmc551 +/-; qmc554 +/- zebrafish embryos

(A) Embryos derived from a cross of wt male with a qmc551 hom; qmc554 hom female fish were injected with Cas9, guide RNAs 65 and 14 and the fluorescent dye mini ruby. Embryos that displayed red fluorescence at ~4-5 hpf were grown up o/n. **(B)** At 24 hpf, genomic DNA was extracted from a subset of embryos. The region of interest was PCR amplified using primers that flank the deleted region. The primers used were DB1400 and DB874. The PCR product was then run on an agarose gel to determine whether the genome editing was successful. The PCR product amplified on the mutant allele was expected to be ~340bp in size. The PCR product

generated on the gDNA of embryo 5 (highlighted in red) was sequenced using primer DB874. The sequence confirmed the expected deletion between TS14 and TS65 and revealed the insertion of an unrelated 18 bp sequence. Siblings were raised.

Once F0 injected embryos were ~3-4 months old, founders were to be identified that carried the *lncRNA* genes mutations in the germline. For this purpose, potential founders were crossed to wt fish, and their progeny were analysed for the presence of the deletion allele. The embryos were collected, and gDNA was extracted for genotyping. The gDNA was used as a template to amplify the region including the TSs (using the same oligos used to amplify the sequences injected into the embryos). Based on the previous genotyping of injected F0 embryos, the presence of a ~340 bp genomic deletion was expected. Of the thirty-three fish that were screened, only one male carried the *lncRNA1* and *lncRNA2* locus mutation were identified. The PCR analysis of the gDNA extracted from embryo no.1 revealed the presence of a 340 bp DNA fragment (Figure 3.16Ai). This fragment was isolated from the gel and sequenced to (a) verify whether the fragment represented the *lncRNA* mutations and (b) determine the exact size of deletion/insertion. The detected mutant was confirmed by plasmid sequencing of embryo no.1. The plasmid sequence showed that the cleavage occurred within TS14 (Figure 3.16Ai (A)), the sequence reversed in the middle the expected deleted region (Figure 3.16Ai (B)) and then continued upstream of TS65 (Figure 3.16Ai (C)). This new mutant *lncRNAs* allele was identified as *qmc22*1. The line was established, and once sexually mature, more than 47 fish were fin-clipped and genotyped to identify carriers, using PCR amplification under the same conditions as described previously. Unfortunately, the mutation was not transmitted to any F1 offspring. However, re-checked the embryonic transmission rate of the founder, and it has been found that 1 out of 20 embryos that has the mutant allele.

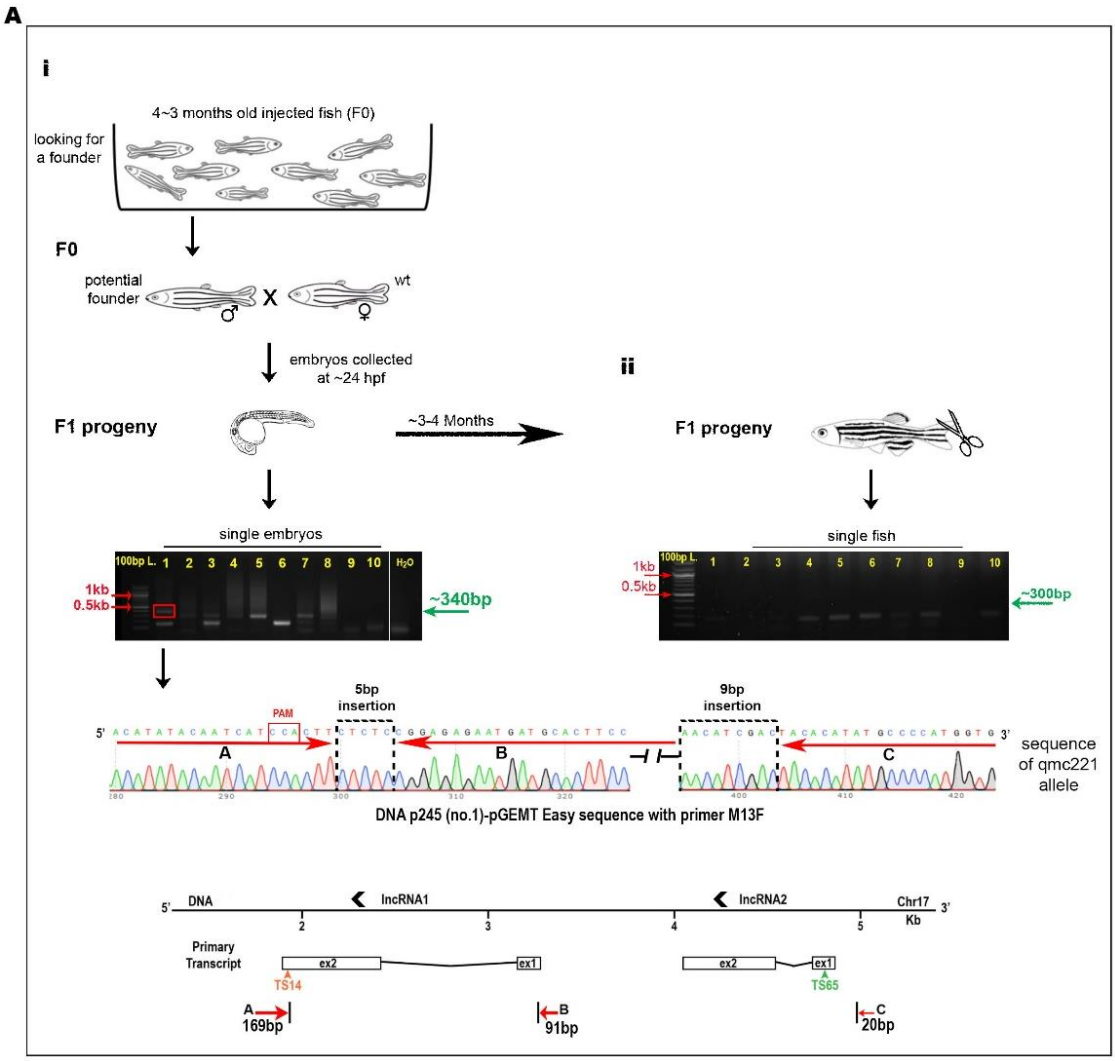


Figure 3.16 Identification of qmc221 heterozygous larvae

(A) Check offspring for germ transmission. **(i)** Once F0 injected were ~3-4 months old, the fish were crossed to wt fish and their progeny were analysed for the presence of a deletion allele. The PCR was performed as described in Figure 3.15. The PCR product was then run on an agarose gel, the predictable deleted band is ~340bp. The detected mutation was confirmed by plasmid sequencing on embryo No.1 (highlighted with red). The plasmid sequence reading represented the cleavage was within TS14 (A), flipped sequence at the mid of expected deleted region (B) then reading upstream of the TS65 (C). This way the qmc221 allele was identified that carried ~2900bp deletion. **(ii)** F1 progeny of the qmc221 founder were raised to adulthood. Adult F1 fish were fin-clipped and genotyped. For this purpose, genomic DNA was isolated from the fin clips. PCR was performed as described in Figure 3.15 to identify the presence of the *qmc221* allele of the lncRNA1 and lncRNA2 locus. The mutation was not transmitted to the F1 offspring.

3.12 CRISPR/Cas9 was used to delete lncRNA genes from the wt zebrafish genome

Due to the failure of two previous deletion attempts, a new TS (TS63) was generated (following the same previously described procedures). The TS63 is downstream of the *lncRNA1* gene with score values of 63 (Figure 3.17A). For this experiment, the TSs (TS63 and TS65) were used to delete *lncRNA1* and *lncRNA2* from the wt zebrafish genome using CRISPR/Cas9 technology (Figure 3.17B-C). The injection was performed as described previously, and the expected deleted was than 4453 bp. To investigate whether the region of interest was deleted, single embryos were genotype. gDNA was extracted from 1 uninjected embryo and ten injected embryos (Figure 3.17D). Then, the region of interest was PCR amplified from the gDNA extracted from injected embryos using external oligos (DB1407-DB778) and internal oligos (DB1388-DB874) that reflected the deleted region of interest. The PCR fragment amplified on the deletion allele is expected to be ~300 bp in size. The uninjected embryo displayed no mutant fragments, whereas nine out of ten of the injected embryos displayed the mutant fragment. Therefore, the top band of embryo 9 was extracted, purified and sequenced. The Sanger sequencing analysis revealed that the fragment contained a 4459 bp deletion, starting downstream of TS63 and extending beyond TS65, with no insertions (Figure 3.17D). Because this indicated the successful targeting of the *lncRNA* locus in these embryos, their siblings were raised.

The ~3 months injected fish were screened for the germ-line transmission of the deletion alleles. For this purpose, potential founders were crossed with wt fish. Randomly selected group of 10 embryos were collected and genotyped as before. PCR amplification with using external oligos (DB1407-DB778) and internal oligos

(DB1388-DB874) revealed the presence of the ~300 bp fragment in group of 10 embryos derived from one of the potential founders (Figure 3.18). The PCR fragment was isolated and sequenced. The sequencing data showed that 4456 bp of sequence between the two TSs had been deleted and that no extra nucleotides had been added. This mutant allele of the lncRNA1/2 locus was called *qmc224*.

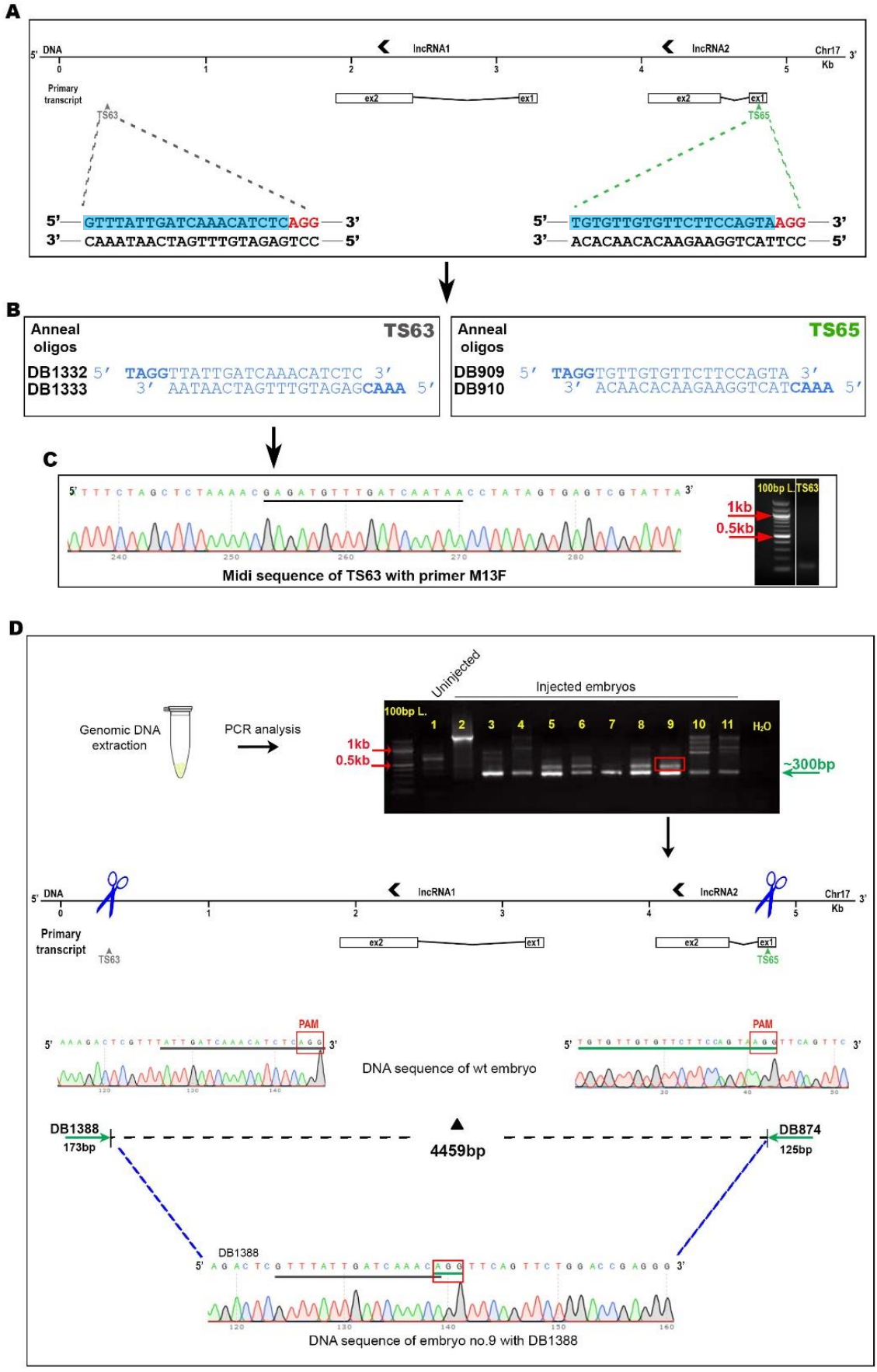

A
DNA
lncRNA1
lncRNA2
Chr17
Kb
Primary transcript
ex2
ex1
TS63
TS65
5' GTTTATTGATCAAACATCTCAGG 3'
3' CAAATAACTAGTTTGTAGAGTCC 5'
5' TGTGTTGTGTTCTTCCAGTAAGG 3'
3' ACACAACACAAGAAGGTCATTCC 5'
B
Anneal oligos
TS63
DB1332 5' TAGGTTATTGATCAAACATCTC 3'
DB1333 3' AATAACTAGTTTGTAGAGCAAA 5'
Anneal oligos
TS65
DB909 5' TAGGTGTTGTGTTCTTCCAGTA 3'
DB910 3' ACAACACAAGAAGGTCATCAAA 5'
C
Midi sequence of TS63 with primer M13F
100bp L.
TS63
1kb
0.5kb
D
Genomic DNA extraction
PCR analysis
Uninjected
Injected embryos
100bp L.
1 2 3 4 5 6 7 8 9 10 11 H2O
1kb
0.5kb
~300bp
Primary transcript
PAM
DNA sequence of wt embryo
DB1388
173bp
4459bp
DB874
125bp
DB1388
DNA sequence of embryo no.9 with DB1388

Figure 3.17 Knockout *lncRNA1* and *lncRNA2* genes of zebrafish using the CRISPR/Cas9 system

(A) Schematics showing the location of targeted genomic sequences; the PAM sequence is highlighted in red and the TS-specific section of the gRNA is highlighted in blue. **(B)** Complementary DNA oligos designed representing the target site sequences harbouring a T7 polymerase binding site and the GG dinucleotide which represents the start of the transcribed sequence (Bold) and annealed to make double-stranded linker fragment. **(C)** The presence of the linker (TS63) sequenced using primer M13F. Exemplary chromatograms are shown that underlined the presence of the target site sequence. The digested fragment been used as a template and was subjected to IVT and analysed by gel electrophoresis. **(D)** At 24 hpf, genomic DNA was extracted from a subset of embryos. The region of interest was PCR amplified using primers that flank the deleted region, the primers used were DB1388 and DB874. The PCR product was then run on an agarose gel to determine whether the edit was successful, the predictable deleted band should be ~300bp. The successful mutation was confirmed by DNA sequencing (embryo 9, highlighted in red) and was related to the wt sequence. As this indicated successful targeting of the lncRNAs locus in these embryos, their siblings were raised.

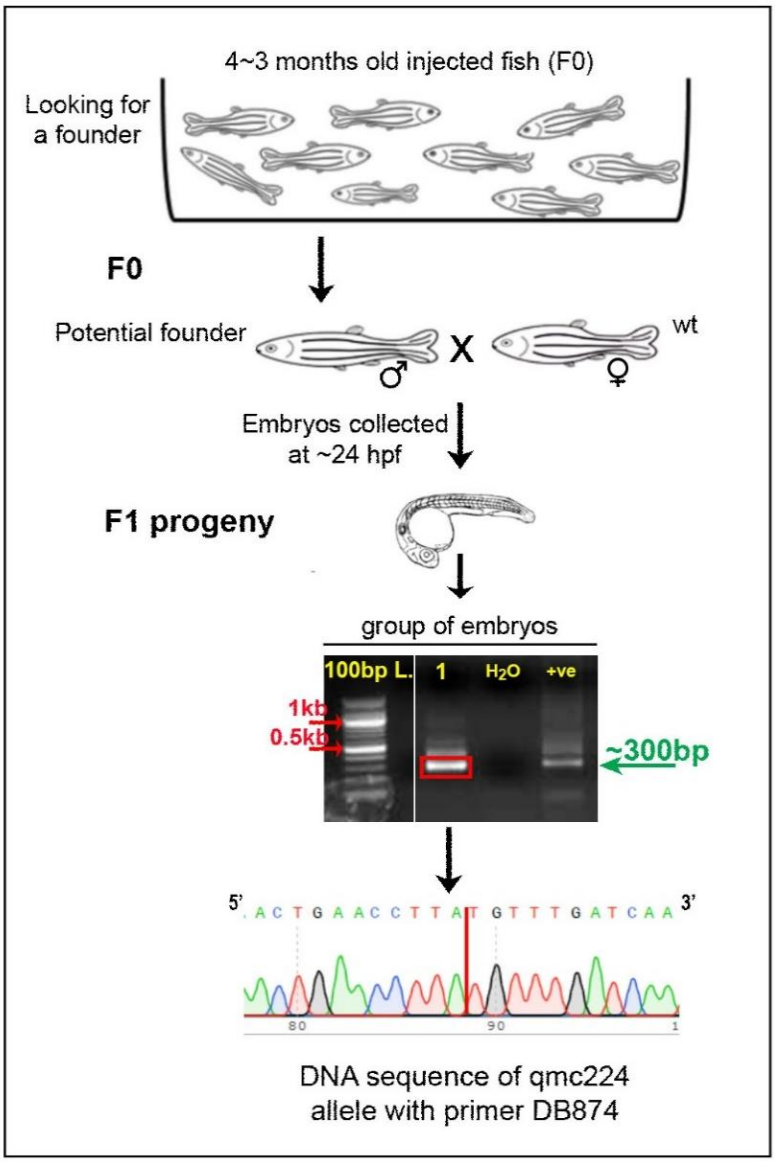


Figure 3.18 Generation and identification of *qmc224* carrier

Check offspring for germ transmission. Once F0 injected were ~3-4 months old, the fish were crossed to wt fish and their progeny were analysed for the presence of a deletion allele. The PCR was performed as described in Figure 3.17. The PCR product was then run on an agarose gel, the predictable deleted band is ~300 bp. The detected mutant was confirmed by DNA sequencing on embryo No.1 (highlighted with red). This way the *qmc224* allele was identified that carried a 4456 bp deletion.

Summary

- To investigate the function/s of *lncRNA1* and *lncRNA2* during zebrafish haematopoiesis, the CRISPR/Cas9 system was used to delete the *lncRNA1* and *2* genes from the genome.
- WT zebrafish embryos were injected with lncRNAs gRNAs and Cas9 and were raised to adulthood. Unfortunately, the mutant alleles detected in injected embryos failed to be transmitted to the next generation.
- The guide RNAs were also injected with Cas9 protein in zebrafish embryos that were double heterozygous for the mutant gfi1aa and gfi1b alleles *qmc551* and *qmc554*. Among the injected fish, one male founder was identified which carried the *lncRNAs*^{qmc221} mutant allele. However, none of its progeny eventually carried this allele in its germline.
- Due to the failure of previous injections, a new TS was generated (TS63). The injection was successful when co-injection TS63 and TS65 and Cas9 were injected wt zebrafish. Among the injected fish, one male founder was identified which carried the *lncRNAs*^{qmc224} mutant allele.

3.13 CRISPR/Cas9 system was used to delete *lncRNA1* from the zebrafish genome

3.13.1 Introduction

In addition to deleting both genes in combination, we attempted to eliminate them individually. Using the CRISPRscan track on the UCSC web page, the sequence between the two *lncRNA* genes was searched for suitable target sites. The one with the highest score that was also unique in the zebrafish genome was TS67, a sequence located right at the beginning of exon 1 of the *lncRNA* gene (Table 2.7). Deletion of the sequences between this site and TS14 was expected to remove the *lncRNA1* gene but leave the *lncRNA2* gene.

The TS14 and TS67 are downstream of *lncRNA1* and upstream of the same gene, respectively (Figure 3.19). TS14 and TS67, were designed and generated, using the previously described protocol (Figure 3.19A-B). Next WT embryos were collected and injected at the 1-cell stage with a mixture containing both TSs, Cas9 and mini ruby. To determine whether the CRISPR/Cas9 system successfully deleted the region of interest, 7 injected embryos and 1 uninjected embryo were randomly collected for genotyping after approximately 24 hpf. gDNA was extracted from each embryo and was used as a template for PCR reactions. The aim was to detect the presence of a genomic deletion between the two target sites. For this purpose, a nested PCR experiment was set up that used a pairs of external and internal oligos (DB877-DB831 and DB1393-DB1400, respectively). Amplification of the deletion allele was expected to yield a PCR fragment that was approximately 540 bp in size. The results showed that gDNA of four out of the 7 injected embryos allowed the amplification of a PCR fragment of that size. The PCR fragment of the embryo 4 sample was isolated, purified, and sequenced. The sequence analysis demonstrated the presence of a *lncRNA*1 deletion

allele. The sequence showed that the break occurred inside the TSs and only a few base pairs upstream of the PAM sequence. It also revealed the presence of a 6 bp insertion (Figure 3.19C). This result confirmed that the CRISPR/Cas9 system successfully deleted the *lncRNA1* gene from the locus and justified the raising of the siblings.

Once the F0 injected fish were sexually mature, the fish were outcrossed with wt fish to obtain a founder (Figure 3.19D). gDNA was extracted and genotyped from the embryos from successful crosses. Unfortunately, although more than 62 potential founders were screened, no founder has been identified.

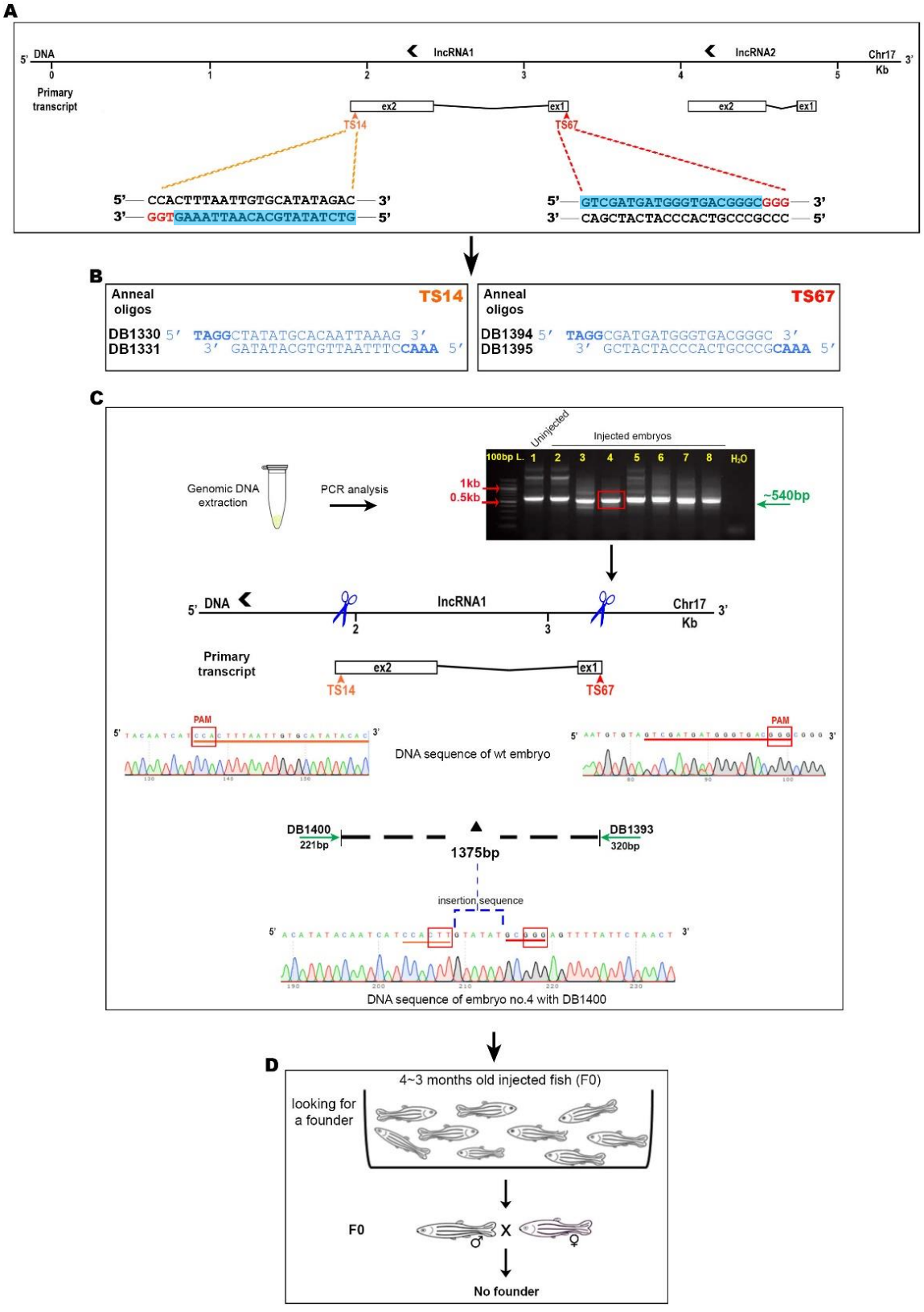


Figure 3.19 Genomic deletion of the zebrafish *lncRNA1* gene using the CRISPR/Cas9 system

(A) Schematics showing the location of targeted genomic sequences; the PAM sequence is highlighted in red and the TS-specific section of the gRNA

is highlighted in blue. **(B)** Complementary DNA oligos were designed that represented the target site sequences harbouring a T7 polymerase binding site and the GG dinucleotide which represented the start of the transcribed sequence (Bold) and annealed to make a double-stranded linker fragment with 5' overhangs. **(C)** At 24 hpf, genomic DNA was extracted from a subset of those embryos that had been injected with the gRNAs and the Cas9. The gDNA served as the template in PCR experiments with the primers DB1400 and DB1393. The PCR products were separated on an agarose gel to determine whether the genome editing was successful. The PCR reaction was expected to generate ~540bp fragment on the deletion allele. The successful mutagenesis was confirmed by DNA sequencing (embryo 4, highlighted in red) and was related to the wt sequence. As this indicated successful targeting of the *lncRNA1* locus in these embryos, their siblings were raised. **(D)** Check offspring for germ transmission. Once F0 injected were ~3-4 months old, the fish were incrossed or crossed to wt fish and their progeny were analysed for the presence of a deletion allele. The F0 injected were not transmitted the mutations to their offspring.

3.14 CRISPR/Cas9 system was used to generate a *lncRNA1* promoter/*lncRNA2* mutant allele

3.14.1 Introduction

Deletion of the sequences between at the beginning of exon 1 of the *lncRNA2* gene and at the beginning of exon 1 of *lncRNA1* was expected to remove the lncRNA2 gene but leave the *lncRNA1* gene almost completely intact. The expression of the latter would, however, come under the control of the *lncRNA2* promoter. This could lead to the misexpression of *lncRNA1* in various cellular contexts.

To target TS67, a TS67-specific gRNA expression vector was generated and *in vitro* transcribed to generate TS67 gRNA. The gRNAs for TS67 and TS65 (Figure 3.20A-C) were injected together with Cas9 protein and mini ruby into one-cell stage zebrafish embryos. gDNAs from one uninjected and 8 single injected 24 hpf embryos were extracted for genotyping. External oligos (DB778-DB1396) and internal oligos (DB874-DB878) were used for the nested PCR amplification of the genomic region containing both TSs. The wt fragment was expected to be approximately 1.5 kb. Successful targeting should generate a ~260 bp fragment in the genomic PCR. From the uninjected embryo, a 1.5 kb wt fragment was amplified. PCR on the gDNAs isolated from the injected embryos resulted in different-sized fragments, including a 260 bp fragment (Figure 3.20D). These variations could be due to the NHEJ repair process that joins the two ends of the DNA strand and variably adds and deletes individual nucleotides. The 260 bp PCR fragment amplified on gDNA isolated from embryo 7 was isolated from the agarose gel, purified and sequenced. The sequence of the amplified PCR product showed a break in the PAM region of TS67, followed by a 38 bp insertion and a segment of TS65, including the PAM region.

Because this result indicated the successful targeting of the *lncRNA1* promoter/*lncRNA2* locus in these embryos, siblings of embryo 7 were raised.

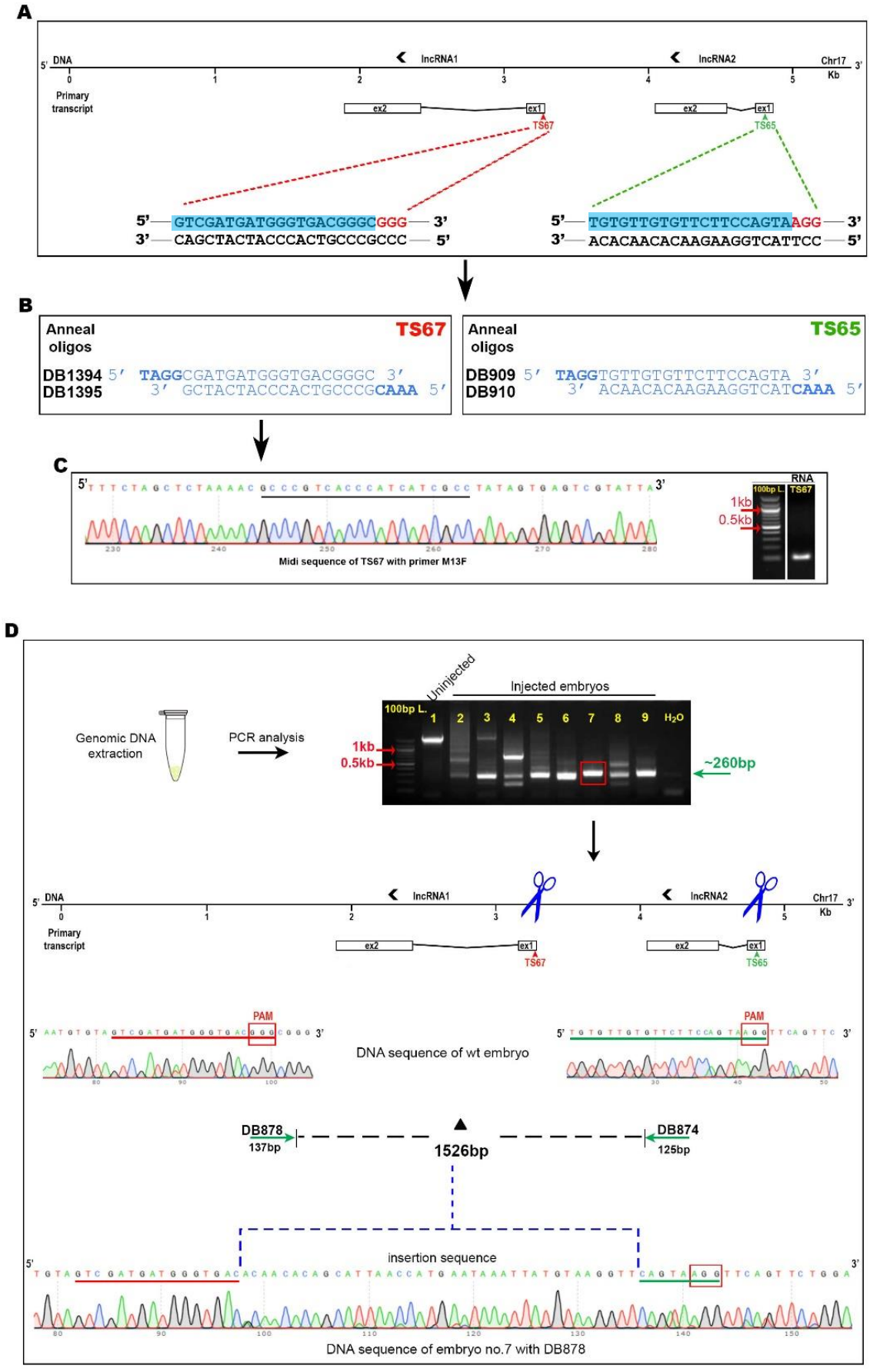

A
DNA
lncRNA1
lncRNA2
Chr17
Kb
Primary transcript
ex2
ex1
TS67
TS65
5' GTCGATGATGGGTGACGGGCGGG 3'
3' CAGCTACTACCCACTGCCCGCCC 5'
5' TGTGTTGTGTTCTTCCAGTAAGG 3'
3' ACACAACACAAGAAGGTCATTCC 5'
B
Anneal oligos
TS67
DB1394 5' TAGGCGATGATGGGTGACGGGC 3'
DB1395 3' GCTACTACCCACTGCCCGCAAA 5'
Anneal oligos
TS65
DB909 5' TAGGTGTTGTGTTCTTCCAGTA 3'
DB910 3' ACAACACAAGAAGGTCATCAAA 5'
C
Midi sequence of TS67 with primer M13F
RNA
100bp L.
TS67
1kb
0.5kb
D
Genomic DNA extraction
PCR analysis
Uninjected
Injected embryos
100bp L.
1 2 3 4 5 6 7 8 9 H2O
1kb
0.5kb
~260bp
DNA
lncRNA1
lncRNA2
Chr17
Kb
Primary transcript
ex2
ex1
TS67
TS65
PAM
DNA sequence of wt embryo
PAM
DB878
137bp
1526bp
DB874
125bp
insertion sequence
DNA sequence of embryo no.7 with DB878

Figure 3.20 Knockout *lncRNA2* gene of zebrafish using the CRISPR/Cas9 system

(A) Schematics showing the location of targeted genomic sequences; the PAM sequence is highlighted in red and the TS-specific section of the gRNA is highlighted in blue. **(B)** Complementary DNA oligos designed representing the target site sequences harbouring a T7 polymerase binding site and the GG dinucleotide which represents the start of the transcribed sequence (Bold) and annealed to make double-stranded linker fragment. **(C)** The presence of the linker (TS67) sequenced using primer M13F. Exemplary chromatograms are shown that underlined the presence of the target site sequence. The digested fragment been used as a template and was subjected to IVT and analysed by gel electrophoresis. **(D)** At 24 hpf, genomic DNA was extracted from a subset of embryos. The region of interest was PCR amplified using primers that flank the deleted region, the primers used were DB878 and DB874. The PCR product was then run on an agarose gel to determine whether the edit was successful, the predictable deleted band should be ~260 bp. The successful mutant was confirmed by DNA sequencing (embryo 7, highlighted in red) and was related to the wt sequence. As this indicated successful targeting of the *lncRNA2* locus in these embryos, their siblings were raised.

The ~3 months injected fish were screened for the germ-line transmission of the deletion alleles. For this purpose, potential founders were crossed with wt fish. Randomly selected individual embryos were collected and genotyped as before. PCR amplification with the primers DB878 and DB874 revealed the presence of the 250 bp fragment in 7 out of 11 embryos derived from one of the potential founders (Figure 3.21Ai). The PCR fragment was isolated and sequenced. The sequencing data showed that 1538 bp of sequence between the 2 TSs had been deleted and that no extra nucleotides had been added. This mutant allele of the lncRNA1/2 locus was called *qmc222*. The *qmc222* allele is fusion at the beginning of exon 1 of the *lncRNA2* gene with *lncRNA1* gene with the expression of *lncRNA2* promoter (Figure 3.21B).

F1 progeny from the *qmc222* founder was raised to adulthood. When they reached sexual maturity, fin clips were taken to identify the *qmc222* carriers (Figure 3.21Aii). Adult fish that were found to carry the *qmc222* allele were incrossed to generate a homozygous mutant line. To reduce the number of fish we needed to raise the F1 progeny were fin-clipped at 3 days post fertilisation (Figure 3.22). Then, gDNA was used as a template to amplify the wt and mutant sequences to identify wt, *qmc222* heterozygous and *qmc222* homozygous fish. First, PCR amplification was performed to detect the ~250 bp mutant fragment, using the same oligos as were used for the injected embryos. The second PCR amplification used the oligos DB750 and DB665 to amplify the wt fragment, which is a ~370 bp in size (Figure 3.22). Embryos number 3, 5, 7 and 10 were identified as being qmc222 homozygous. The qmc222 homozygous embryos were raised.

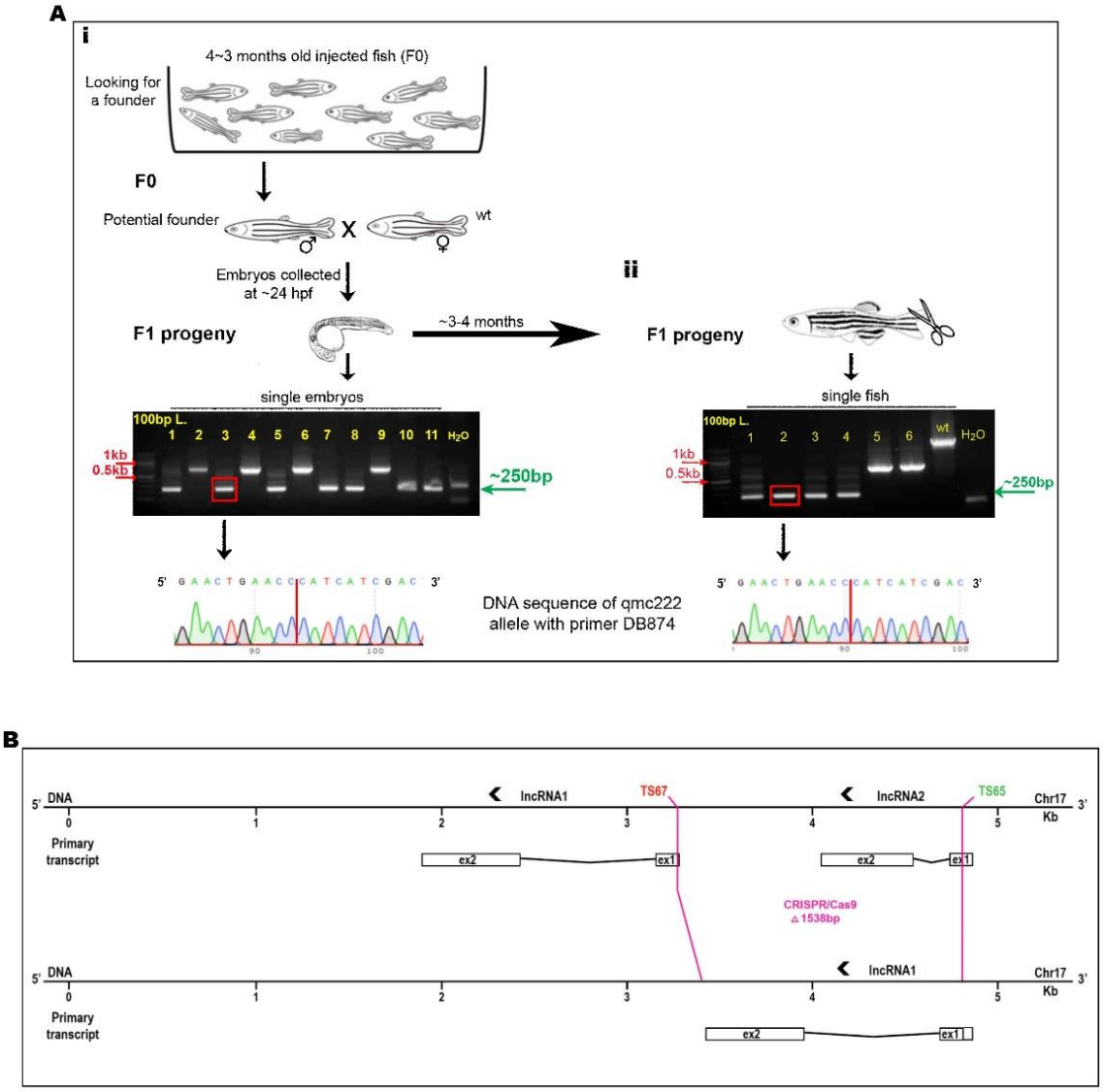


Figure 3.21 Generation and identification of qmc222 carrier

(A) Check offspring for germ transmission. (i) Once F0 injected were ~3-4 months old, the fish were crossed to wt fish and their progeny were analysed for the presence of a deletion allele. The PCR was performed as described in Figure 3.20. The PCR product was then run on an agarose gel, the predictable deleted band is ~251bp. The detected mutant was confirmed by DNA sequencing on embryo No.3 (highlighted with red). This way the qmc222 allele was identified that carried a 1538 bp deletion. (ii) F1 progeny of the qmc222 founder was raised to adulthood. Adult F1 fish were fin-clipped and genotyped. For this purpose, genomic DNA was isolated from the fin clips. PCR was performed as described in Figure 3.20 to identify the presence of the qmc222 allele of the lncRNA1/2 locus, embryo No.2 (highlighted in red) was confirmed the mutant by sequencing. **(B)** A diagram illustrates an overview of the qmc222 gene mutant structure.

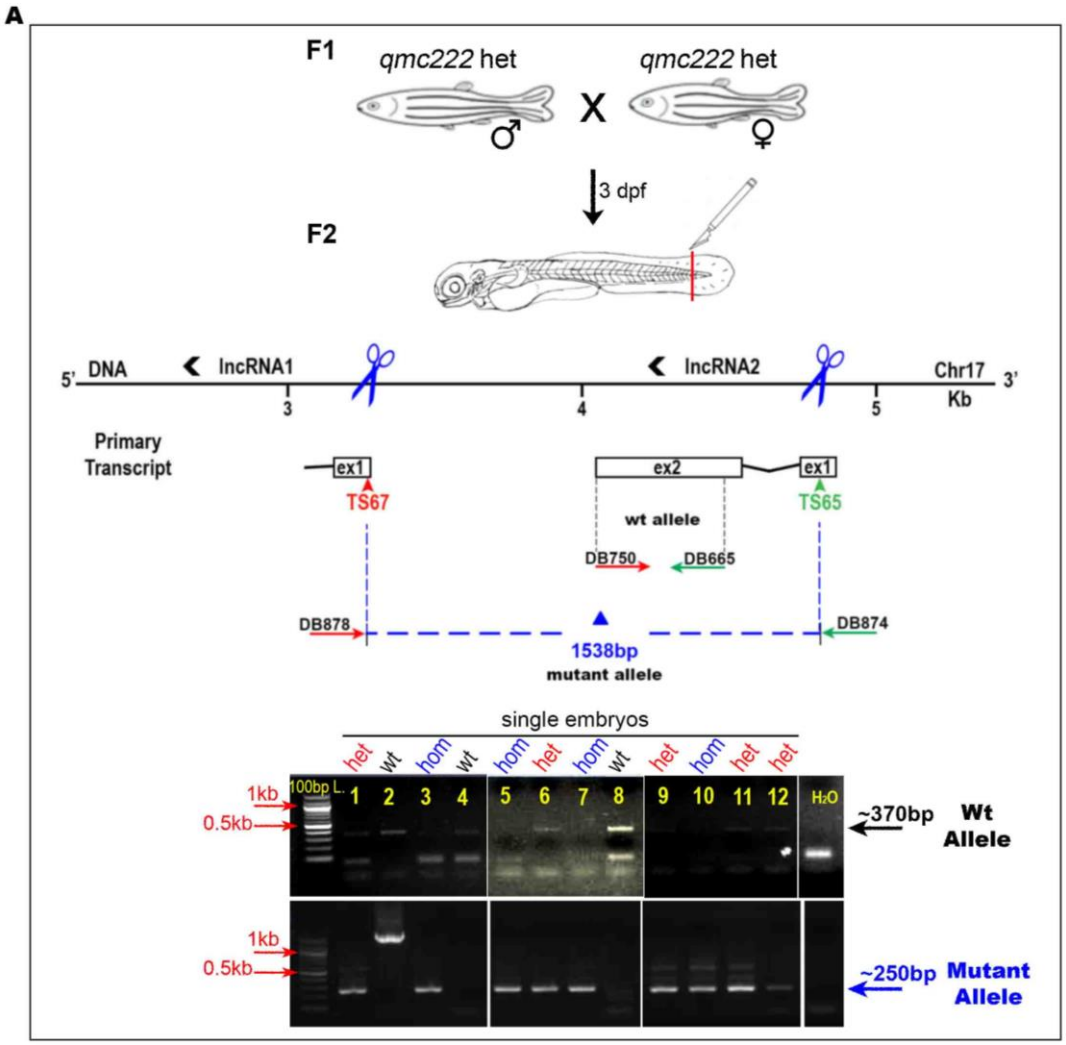


Figure 3.22 Generation and identification of qmc222 homozygous larvae

Identified homozygous carriers were incrossed to generate F2 progeny. These were genotyped at 3 dpf by fin clipping. Genomic DNA extracted from the fin clips was used in two PCR reactions. In the first, primers DB750 and DB665 were used to amplify the wt allele of the lncRNA1/2 locus. In the second PCR, oligos DB878 and DB874 were applied to detect the presence of the qmc222 deletion allele, the qmc222 hom fin clip embryos were raised.

Next, expression of *lncRNA1* and *lncRNA2* was studied in embryos homozygous for the *qmc222* allele. For this purpose, whole-mount *in situ* hybridisation experiments were performed on embryos obtained from a *qmc222* heterozygous incross. First, it was confirmed that the loss of the *lncRNA2* gene causes the expected loss of *lncRNA2* expression. At 36 hpf, embryos showed no staining (Figure 3.23Bi-iii). Genotyping these embryos confirmed that all five were homozygous *qmc222* carriers. No homozygous embryos were found in the cohort of the stained embryo. Interestingly, genotyping the latter embryos also revealed that the loss of one copy of *lncRNA2* correlated with reduced staining in the position of the ventral wall of the dorsal aorta (Figure 3.23D). Then, the expression of *lncRNA1* was examined to see whether its expression was altered under the control of the *lncRNA2* promoter. At 36 hpf, *lncRNA1* expression was detected in all embryos, but clear differences in the expression level were visible that correlated with the embryos' genotype. *lncRNA1* expression was observed only in the CHT in the single wt embryos, whereas expression was observed in the vDA and in the CHT in both *qmc222* heterozygous and homozygous embryos (Figure 3.23Ai-iii), suggesting that the *lncRNA2* promoter supported lncRNA1 expression from the *qmc222* allele at levels that were similar to the wt levels of *lncRNA2* and substantially higher than that of *lncRNA1*. Next, lncRNA*1* in situ hybridisation was performed at 3 dpf to see whether its characteristic T cell was lost when under the control of the *lncRNA2* promoter. Surprisingly, however, thymic *lncRNA1* expression was retained in all embryos (Figure 3.24). This suggests that it is not the *lncRNA1* promoter but an enhancer that directs *lncRNA1* expression to T cells. This enhancer is able to interact with lncRNA2 promoter, however, only does so in the absence of the lncRNA1 promoter (Appendix B).

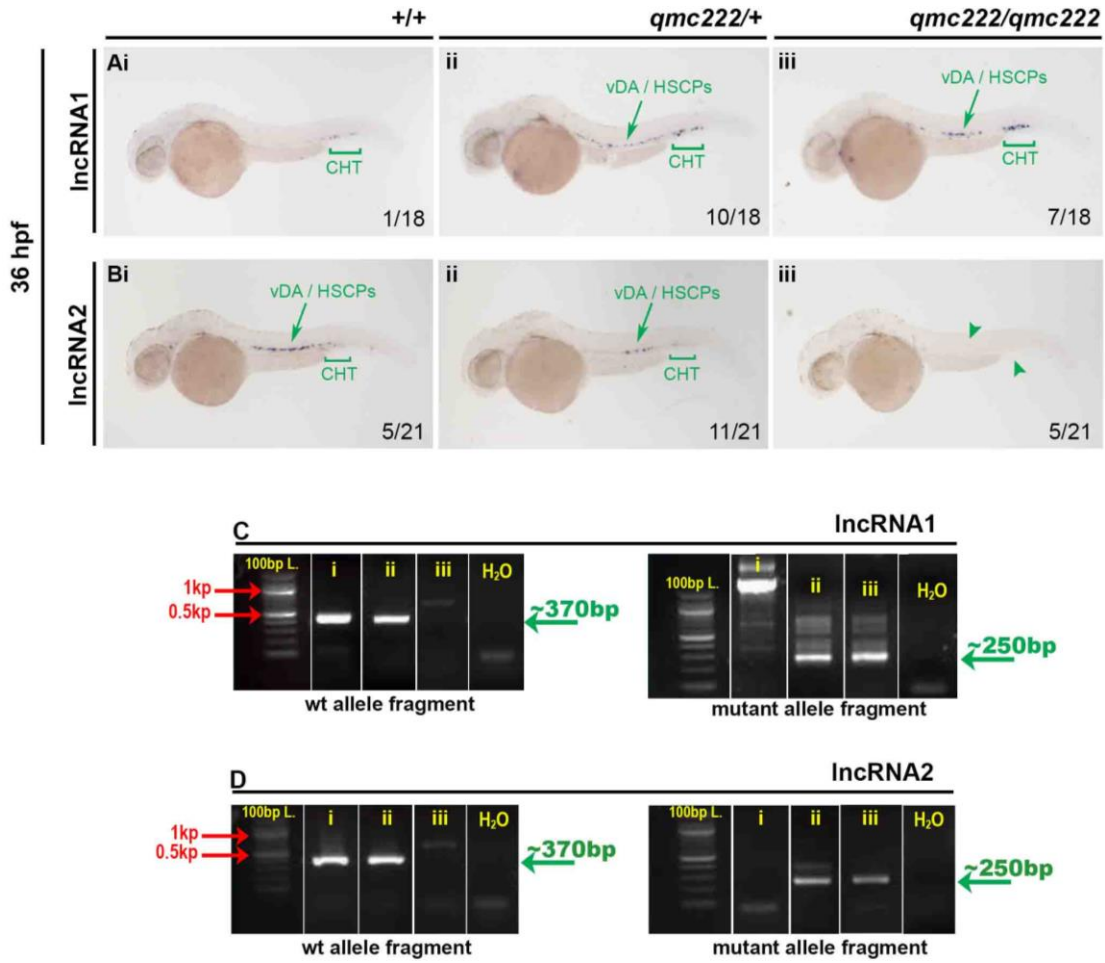


Figure 3.23 Loss of *lncRNA2* expression in *qmc222* mutant confirmed that the CRISPR/Cas9 system deleted *lncRNA2* gene of zebrafish

(A-B) Expression of *lncRNA1* and *lncRNA2* in ~36 hpf (i) wt, (ii) *qmc222* heterozygous, and (iii) *qmc222* homozygous embryos detected by WISH. All embryos are presented in a lateral view with anterior facing the left. Green arrow: endothelial cells of the vDA and HSPCs, green bracket: the CHT, green arrowhead: missing in expression in endothelial cells of the vDA and HSPCs, and missing in expression in HSPCs of the CHT. The numbers of genotyped embryos over the total numbers of embryos analysed are provided on the panels. **(B)** PCR amplification of WISH embryos to distinguish *wt* and *qmc222* alleles.

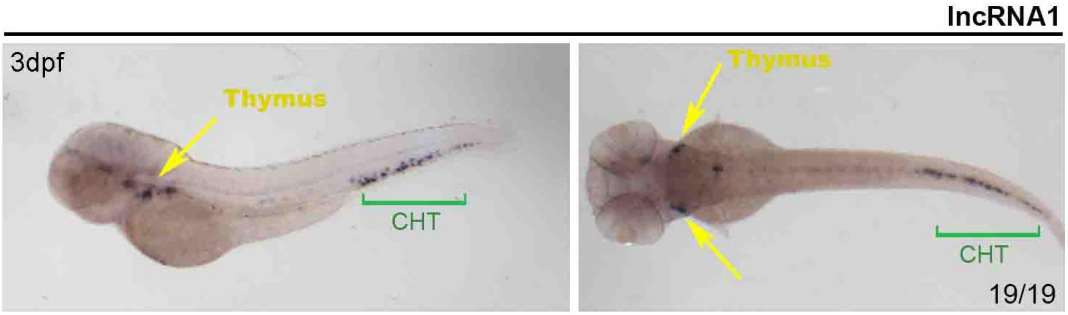


Figure 3.24 *lncRNA1* expression is retained in thymic in *qmc222*

Expression of *lncRNA1* in 3 dpf of qmc222 embryos detected by WISH. Embryo is presented in a lateral and dorsal view with anterior facing the left. Yellow arrow: thymus, green bracket: the CHT.

3.15 Mutant *lncRNA2* alleles were generated using the CRISPR/Cas9 system

In a second attempt to delete the lncRNA2 gene in the genome an approach was chosen that should maintain the integrity of the *lncRNA1* gene and its promoter. Again, two sequences were targeted with the aim to delete the sequence between the target sites. One of the sites, TS65 (the site in exon 1 of the lncRNA2 gene) was identical to the one used in the previous experiment. The second target site was 78 bp downstream of the *lncRNA2* gene and 665 bp upstream of the *lncRNA1* coding sequence. Deletion of the sequence between the two TSs was expected to delete the lncRNA2 gene only (Figure 3.25A).

The two guide RNAs (TS65 and TS81) were injected with Cas9 and mini red ruby into 1-cell stage wt embryos. The following day, ten injected embryos and one uninjected embryo were randomly collected for PCR-based genotyping using nested pairs of oligos (DB778-DB878 and DB874-DB876, respectively). Multiple PCR fragments of different sizes were amplified including a ~500 bp fragment that could represent the mutant allele (Figure 3.25D). The 500 bp fragment of embryo 8 was isolated and sequenced. The sequence analysis showed that the corresponded to the *lncRNA1/2* locus. It represented a mutant sequence that must have been derived from Cas9 cleavage inside the PAM sequence of TS81 and 2 base pairs upstream of the TS65. There was no insertion of additional base pairs. Because the genotyping suggested the successful targeting of the *lncRNA2* locus in these embryos, their siblings were raised to adulthood.

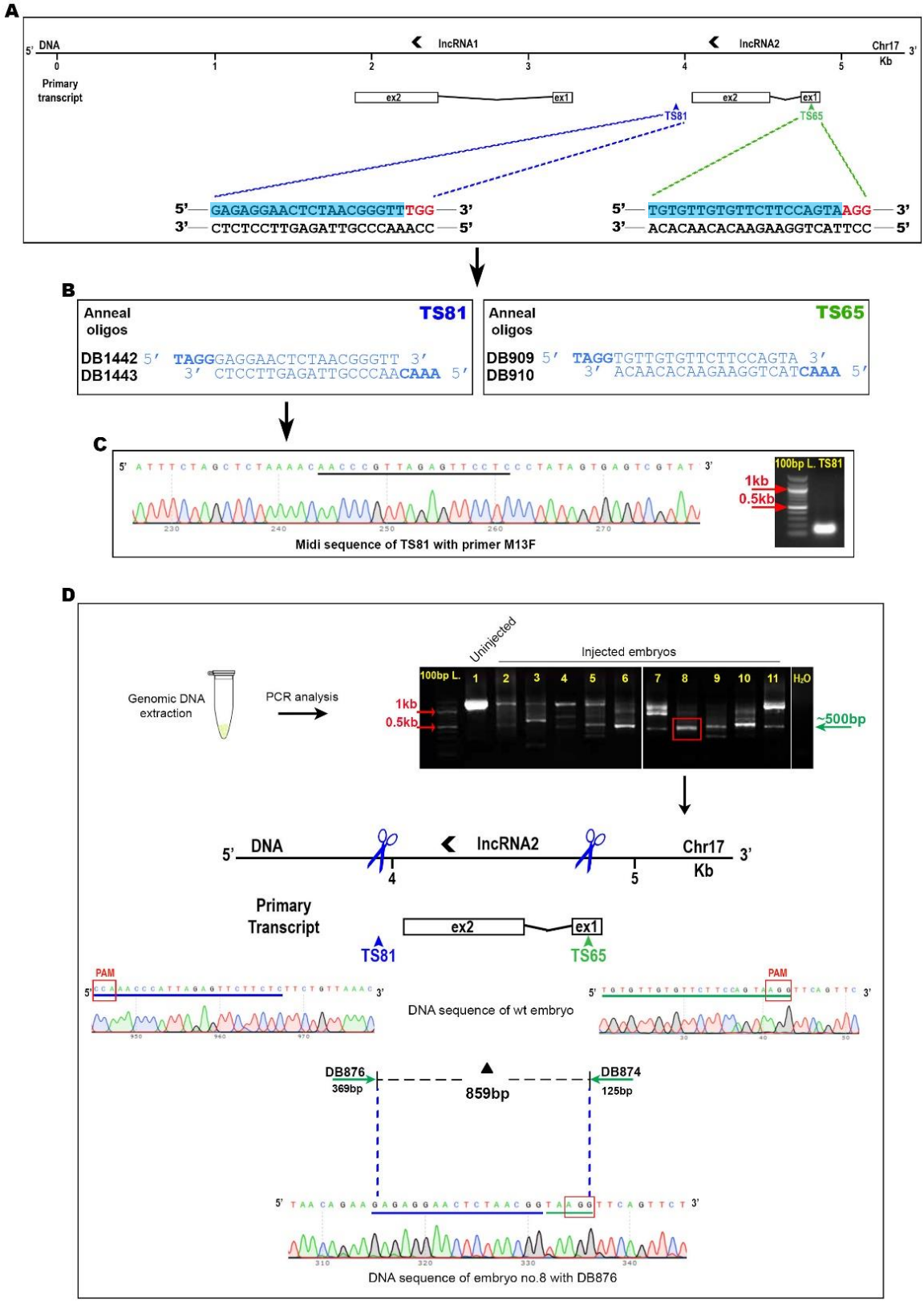


Figure 3.25 Knockout *lncRNA2* gene of zebrafish using the CRISPR/Cas9 system

(A) Schematics showing the location of targeted genomic sequences; the PAM sequence is highlighted in red and the TS-specific section of the gRNA

is highlighted in blue. **(B)** Complementary DNA oligos designed representing the target site sequences harbouring a T7 polymerase binding site and the GG dinucleotide which represents the start of the transcribed sequence (Bold) and annealed to make double-stranded linker fragment. **(C)** The presence of the linker (TS81) sequenced using primer M13F. Exemplary chromatograms are shown that underlined the presence of the target site sequence. The digested fragment been used as a template and was subjected to IVT and analysed by gel electrophoresis. **(D)** At 24 hpf, genomic DNA was extracted from a subset of embryos. The region of interest was PCR amplified using primers that flank the deleted region, the primers used were DB876 and DB874. The PCR product was then run on an agarose gel to determine whether the edit was successful, the predictable deleted band should be ~500 bp. The successful mutant was confirmed by DNA sequencing (embryo 8, highlighted in red) and was related to the wt sequence. As this indicated successful targeting of the lncRNA2 locus in these embryos, their siblings were raised.

To identify a founder that carried this particular mutation, the F0 injected fish were crossed with wt fish when they were 3-4 months old. From each successful cross, embryos were genotyped individually to detect the presence of the deletion allele. The expected 500 bp PCR product was amplified on the gDNA isolated from 3 out of 10 embryos that were the progeny of a female fish (Figure 3.26Ai). The sequence of the PCR product representing embryo 1 showed that 861 bp were deleted between the two TSs and that 2 bp had been inserted during non-homologous end joining (Figure 3.26Ai). Siblings of the founder were fin-clipped at 3 dpf and carriers of the mutant *lncRNA2* allele were raised to adulthood (Figure 3.26Aii). This new *lncRNA2* mutant allele was named *qmc223*. The *qmc223* allele is fusion at the beginning of exon 1 of the *lncRNA2* gene with *lncRNA1* promoter (Figure 3.26B).

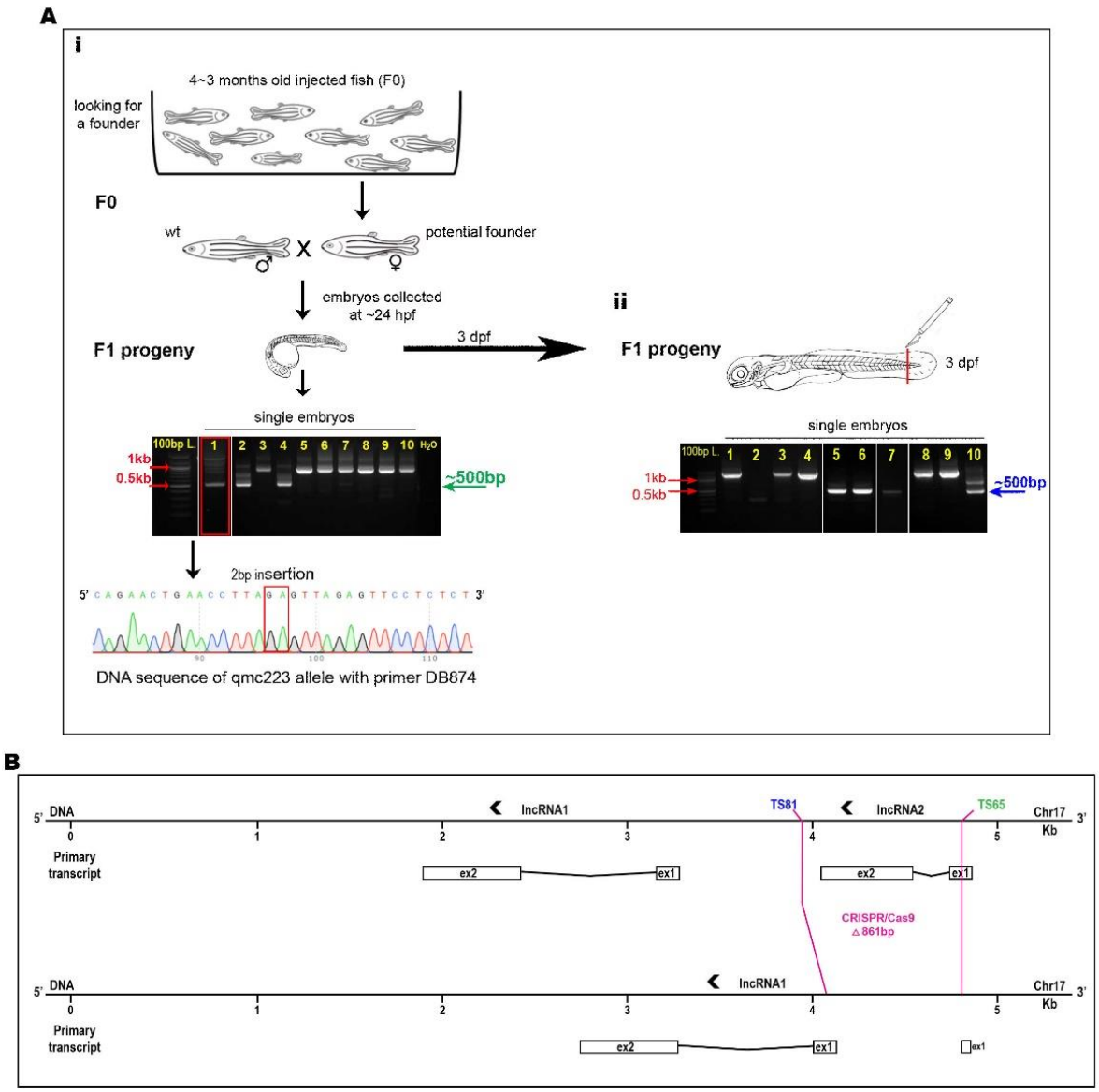


Figure 3.26 Generation and identification of *qmc223* heterozygous larvae

(A) Check offspring for germ transmission. (i) Once F0 injected were ~3-4 months old, the fish were crossed to wt fish and their progeny were analysed for the presence of a deletion allele. The PCR was performed as described in (Figure 3.25). The PCR product was then run on an agarose gel, the predictable deleted band is ~500 bp. The detected mutant was confirmed by DNA sequencing on embryo No.1 (highlighted with red). This way the *qmc223* allele was identified that carried a ~800 bp deletion. (ii) F1 day 3 progeny were fin-clipped and genotyped. For this purpose, genomic DNA was isolated from the fin clips. PCR was performed as described (Figure 3.25) to identify the presence of the *qmc223* allele of the *lncRNA2* locus, the *qmc223* het fin clip embryos were raised. **(B)** A diagram illustrates an overview of the *qmc223* gene mutant structure.

WISH was performed to confirm that the *lncRNA2* was no longer expressed in *qmc223* homozygous mutant embryos. For this experiment, *qmc223* heterozygous fish were incrossed to obtain 36 hpf *wt*, *qmc223* heterozygous and *qmc223* homozygous embryos. Three of the 22 embryos that were stained in the WISH experiment displayed no *lncRNA2* expression in the vDA and in the CHT. Genotyping revealed that these 3 embryos were homozygous carriers of the *qmc223* allele. PCR experiments on their gDNAs showed the successful amplification of the qmc223-specific 500 bp fragment. By contrast, amplification of a wt-specific 370 bp PCR fragment (between oligos DB750 and DB665) failed. The 8 embryos with the highest level of vDA and CHT staining turned out to be *wt* embryos. Heterozygous embryos displayed weaker staining in the vDA and the CHT. This is likely due to a gene dosage effect. In summary, the WISH data confirmed that there is no *lncRNA2* expression from the *qmc223* allele of the *lncRNA1/2* locus (Figure 3.27Ai-iii) (Appendix C).

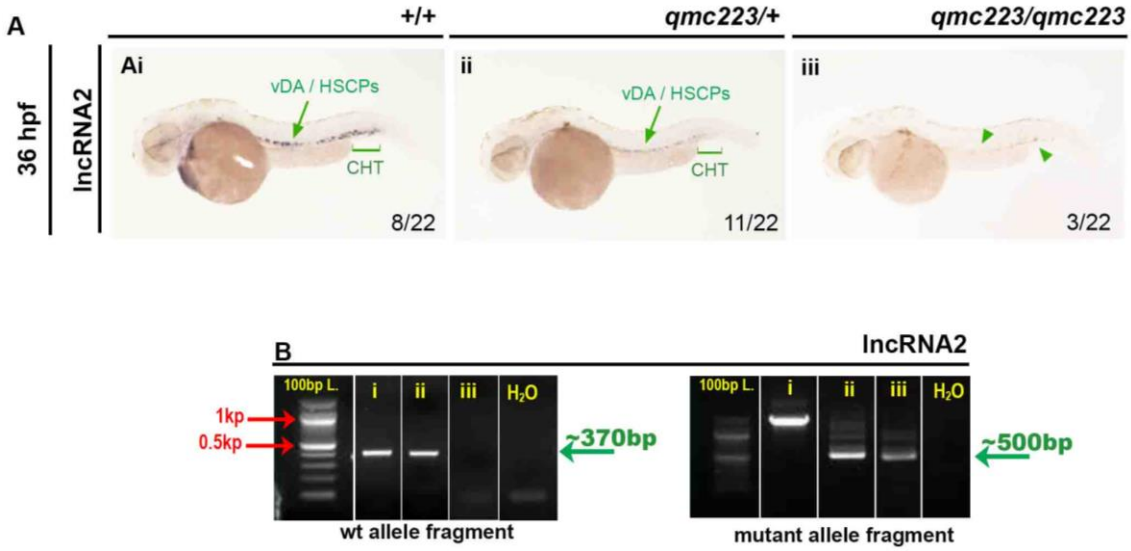


Figure 3.27 Loss of lncRNA2 expression in *qmc223* mutant confirmed that the CRISPR/Cas9 system deleted *lncRNA2* gene of zebrafish

(A) Expression of *lncRNA2* in ~36 hpf **(i)** wt, **(ii)** *qmc223* heterozygous, and **(iii)** *qmc223* homozygous embryos detected by WISH. All embryos are presented in a lateral view with anterior facing the left. Green arrow: endothelial cells of the vDA and HSPCs, green bracket: the CHT, green arrowhead: missing in expression in endothelial cells of the vDA and HSPCs and missing in expression in HSPCs of the CHT. The numbers of genotyped embryos over the total numbers of embryos analysed are provided on the panels. **(B)** PCR amplification of WISH embryos to distinguish *wt* and *qmc223* alleles.

Summary

- The CRISPR/Cas9 technology was used to delete *lncRNA* genes individually. To investigate their functions in blood development separately.
- Wt zebrafish embryos were injected with *lncRNA1* gRNAs and Cas9 and were raised to adulthood. Unfortunately, the mutant alleles detected in injected embryos failed to be transmitted to the next generation.
- We successfully generated a deletion of the region that encompassed the *lncRNA1* promoter and *lncRNA2* gene, and the line was named *qmc222*. In the qmc222 allele, the beginning of exon 1 of the *lncRNA2* gene is fused to the *lncRNA1* gene.
- *qmc222* homozygous embryos, larvae and adults were viable.
- Confirming the loss of lncRNA2 expression in *qmc222* homozygous embryos and by contrast, *lncRNA1* expression was retained in *qmc222* homozygous embryos by WISH experiments
- In order to investigate the role of lncRNA2 only, a *qmc223* line was generated with mutation of lncRNA2 only. Heterozygous embryos survive to adulthood and are fertile. WISH analysis confirmed the loss of *lncRNA2* expression.

4. Discussion

Long non-coding RNAs (lncRNAs) have been found to perform various functions, associated with a wide variety of important biological processes, and demonstrate tissue- and cell-type-specific expression patterns. They can function as regulators of cell fate, via a number of mechanisms, ranging from gene expression regulation to effects on mRNA and protein stability (Delás and Hannon, 2017). lncRNAs have been linked to the development of haematopoiesis, through effects on cell lineage specification, proliferation, and differentiation (Delas et al., 2019). Novel zebrafish lncRNAs have been identified in our lab through transcriptome analyses, performed on GFP-positive and GFP-negative cells isolated from the zebrafish line *qmc551*. LncRNAs appear to be specifically expressed in GFP-positive cells, which represent primitive and definitive blood cells. In this project, RNA *in situ* hybridisation was performed in wt and mutant embryos, to closely inspect gene expression patterns and identify cell types. In addition, to understand the roles that lncRNAs may play during haematopoiesis, reverse genetics strategies, such as CRISPR/Cas9 technology, were used to delete genes, both individually and in combination, from the zebrafish genome.

4.1 lncRNA1 and lncRNA2 expression is associated with haematopoiesis during zebrafish development

The spatiotemporal expression patterns of lncRNAs in wt zebrafish embryos during development are consistent with the expression patterns of primitive haematopoietic genes, such as *scl* and *cloche,* and definitive haematopoietic genes, such *runx1* and *cmyb* (Gering et al., 1998, Green et al., 1992, Drake et al., 1997, Kallianpur et al., 1994). These results corroborate the findings of previous studies in mice (Cabezas-Wallscheid et al., 2014). A transcriptome analysis was performed on progenitor populations, which corresponded to HSCs and were isolated from mouse bone marrow by fluorescence-activated cell sorting. In this study, they identified lncRNAs that were strongly expressed in HSCs, relative to other cell populations (Cabezas-Wallscheid et al., 2014). Functional studies have determined that lncRNAs regulate various aspects of HSCs. For instance, an lncRNA required for HSC differentiation, named Spehd, was identified in mouse bone marrow. Reduced levels of Spehd decreased the capacity for multilineage differentiation, which was identified due to deficient oxidative phosphorylation in the common myeloid progenitor (CMP) population when the transcriptomic analysis was performed (Delas et al., 2019).

4.2 *lncRNA1* and *2* expression is blood-specific but variable in different haematopoietic lineages

RNA *in situ* hybridisation experiments, performed in *cloche* mutant embryos, confirmed that both *lncRNA1* and *2* are expressed in haematopoietic cells and ECs. Stainier et al. demonstrated that *npas4l* (*cloche*) acts at the top of the haemato-vascular specification cascade. defects in *npas4l* result in the failure to develop blood, ECs

and endocardial cells (Stainier et al., 1995). Therefore, the lack of expression for both lncRNAs in *cloche* mutants indicated that these genes are downstream of *npas4l*. Further *in situ* hybridisations experiments were conducted to uncover the cell types that express lncRNAs, which demonstrated that both genes are expressed in primitive and definitive haematopoietic cells. However, the lack of *lncRNA*1 and *lncRNA2* expression in a subset of cells in *mib* mutants revealed that these cells are definitive progenitors. *Notch* is normally expressed in the DA, but not in the postcardinal vein (PCV), during embryonic vascular development. Embryos lacking Notch activity, such as *mib* mutant embryos, fail to induce arterial-specific ephrinB2 (*efnb2a*) expression and exhibit the ectopic expression of venous markers within the DA (Lawson et al., 2001). This defect results in the inability to generate definitive blood cells, but normal numbers of prRBCs (Gering and Patient, 2005, Burns et al., 2005) and presumptive pre-definitive erythromyeloid progenitors (EMPs) were generated (Bertrand et al., 2007). *Notch1* knockout mice form primitive blood normally; however, they die early, due to EC defects, preventing the analysis of definitive haematopoiesis. However, their retention in the posterior trunk and in the PBI is likely due to lncRNA expression in prRBCs and/or presumptive EMPs, which struggle to enter the circulation in *mib* embryos (Lawson et al., 2001). At the onset of definitive haematopoiesis, the vDA expresses *runx1* and *cmyb* (Kalev-Zylinska et al., 2002, Burns et al., 2002). *Runx1* and *cmyb* mutants present specific defects in definitive haematopoiesis. *Runx1* morphant and mutant embryos specify HECs (Thambyrajah et al., 2016) but fail to undergo endothelial to haematopoietic transition (Kissa and Herbomel, 2010; Bonkhofer et al., 2019). In *cmyb* mutants, normal numbers of prRBCs are formed, but severe defects in definitive haematopoiesis occur, including a loss of thymocytes (Soza-Ried et al., 2010). *Runx1* knockout mice fail to develop definitive erythroid cells, myeloid cells, and lymphoid cells

(Wang et al., 1996, Okuda et al., 1996). Moreover, cmyb expression decreases with *runx1* knockdown in the DA and, later, lymphopoiesis is also affected (Kalev-Zylinska et al., 2002, Gering and Patient, 2005, Burns et al., 2005). In both of these mutants, *lncRNA1* expression is depleted in the definitive haematopoietic system. A clear loss of *lncRNA1* expression was observed specifically for the vDA portion of the trunk and in T-cells of the thymus, in *runx1* and *cmyb* mutants, respectively (Figure 3.7 and Figure 3.8), suggesting that lncRNA1 is expressed in definitive cells, where it downregulates *runx1* and *cmyb*. LncRNA appears to tightly regulate the survival and function of mouse CD8 T-cells by regulating the expression of other factors (Kotzin et al., 2019). In addition, lncRNA controls the survival of other haematopoietic cells, such as neutrophils, eosinophils, and monocytes, in response to the regulation of the transcription of its neighbouring pro-apoptotic genes (Kotzin et al., 2016).

Interestingly, the expression of *lncRNA2* was missing at 30 hpf in *qmc551/sh320* double homozygous embryos that lack the expression of *gfi1aa* and *gfi1ab* (Figure 3.6). Whether this lack of expression is a transient phenomenon caused by a delay in the endothelial to haematopoietic transition or whether it highlights the failure of emerging cells to express *lncRNA2* is currently under investigation. Preliminary data produced by Mariam Jaddah, another PhD student in the lab, suggested that HECs in the DA that lack Gfi1aa and Gfi1ab can still differentiate into definitive blood cells that seed the CHT, the thymus, and the kidney, suggesting that *lncRNA2* expression most likely recovers in the *gfi1aa*/*gfi1ab* double mutant embryos. The *lncRNA2* probe may, thus, prove to be a good indicator for progression through the endothelial to haematopoietic transition.

4.3 *lncRNA2* expression in prRBC is repressed by Gfi1aa

WISH experiments revealed that 30 hpf qmc551 homozygous gfi1aa mutant embryos display a significant increase in *lncRNA2* expression in the ICM and in the PBI. It had been noted previously that *qmc551*-carrying embryos display perivascular GFP+ prRBC progenitors that co-express the gata1:dsRed transgene (Thambyrajah et al., 2016). It has since been noted that *qmc551* homozygous embryos carry more of these cells than heterozygous embryos. This increase is likely the delay in prRBCs differentiation that our lab has documented in the absence of Gfi1aa (Moore et al., 2018). These delayed cells may struggle to find their way into circulation at the time when the DA and the posterior cardinal vein form. They, thus, end up in the mesenchyme outside the vessels. Consistent with this interpretation, transcriptome analysis that was performed on prRBCs isolated from *qmc551* homozygous and heterozygous embryos and gfi1b morphants (Moore et al., 2018) established that *lncRNA2* expression was elevated in *qmc551* homozygous prRBCs relative to the *qmc551* heterozygous control cells (Figure 4.1). The expression of *lncRNA1* was also elevated in the absence of Gfi1aa but remained at much lower levels overall, levels that were difficult to assess by RNA in situ hybridisation. These findings suggest that in the absence of Gfi1aa prRBCs were delayed in activating late erythroid differentiation and in the absence of both Gfi1aa and Gfi1b progressively delayed prRBCs even further (Moore et al., 2018). This delay due to struggled to suppress early erythroid genes which *lncRNA2* could be one of them.

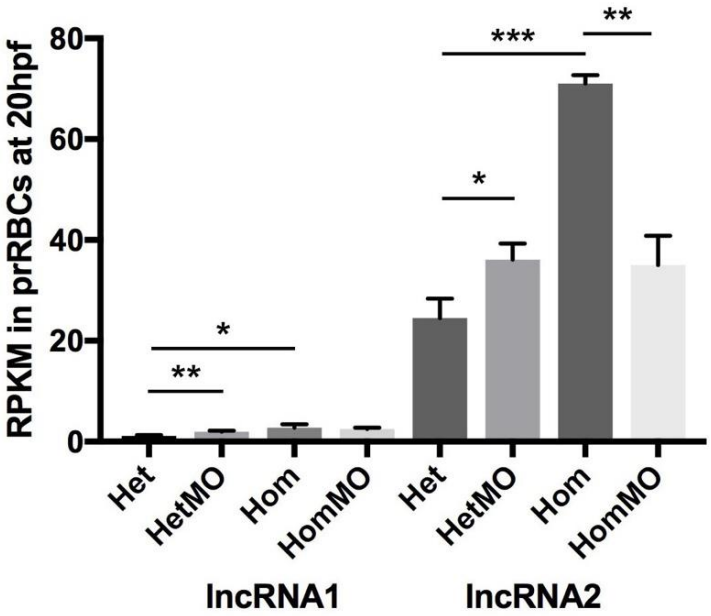


Figure 4.1 The levels of *lncRNA1* and *lncRNA2* expression in prRBCs at 20hpf in Gfi1aa and/or Gfi1b-depleted embryos

The bar chart shows the *lncRNA1* and *2* gene expression data from an RNA-Seq experiment that was performed on prRBCs isolated from *qmc551* heterozygous, *gfi1b* morphant *qmc551* heterozygous, *qmc551* homozygous and *gfi1b* morphant *qmc551* homozygous 20 hpf embryos (Moore et al., 2018). The expression levels are given in RPKMs. Standard deviations derived from triplicate measurements are provided. An unpaired t-test with Welch's correction was performed on the data to demine whether differences between the means were statistically significant: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.4 *qmc222* homozygous fish survive to adulthood and are fertile

We successfully generated a deletion of the region that encompassed the *lncRNA1* promoter and *lncRNA2* gene, and the line was named qmc222 (Figure 3.20 and Figure 3.21). In the *qmc222* allele, the beginning of exon 1 of the *lncRNA2* gene is fused to the lncRNA1 gene. The *lncRNA1* gene should, therefore, be expressed under the control of the *lncRNA2* promoter.

qmc222 homozygous embryos were identified and raised to adulthood. It turned out that the qmc222 homozygous embryos, larvae and adults were viable. Then WISH experiments were performed and confirmed the loss of *lncRNA2* expression in qmc222 homozygous embryos (Figure 3.23). By contrast, lncRNA1 expression was retained in qmc222 homozygous embryos. There was slightly increased lncRNA1 expression relative to the wt embryos in the trunk and in the CHT region (Figure 3.23). Besides, at 30 hpf, a close up to the heart region clearly showed that *lncRNA1* expression is in prRBCs cells that are pooled around the heart that was not expressed in wt embryos (Appendix D). Suggesting that the lncRNA2 promoter supported lncRNA1 expression from the qmc222 allele at levels that were similar to the wt levels of lncRNA2 and substantially higher than that of lncRNA1. In addition, the WISH analysis performed the expression of *lncRNA1* on 3 dpf qmc222 heterozygous in-cross embryos showed that lncRNA1 expression was retained in the thymus. Given that only *lncRNA1* was normally expressed in T cells, this result suggests that it is not the *lncRNA1* promoter but an enhancer that directs *lncRNA1* expression to T cells. This enhancer is obviously able to interact with lncRNA2 promoter, however, only does so in the qmc22 allele of the lncRNA1/2 locus. Insulators may block their interaction in the wt allele. A natural

progression of this work is to analyse qPCR on these cells of mutant line.

4.5 qmc221 allele is not transmitted to the next generation

An attempt to uncover the functions of the lncRNAs genes during zebrafish haematopoiesis, here, the CRISPR/Cas9 system was used to delete the lncRNA genes from the wt and *qmc551het/qmc554het* zebrafish genome. We were able to generate lncRNA1/2 deletion allele with background mutant of *qmc551het/qmc554het,* which called *qmc221*. However, this mutant allele was not transmitted to the next generation. It was suggested that one of the targets sites have a low score (TS14) that could affect the efficiency of Cas9 generating a double strand break (Moreno-Mateos et al., 2015). Moreover, the mutagenesis rates could be partially due to embryo cells that amass Cas9-sgRNA complex become insufficient NHEJ at the target locus specified by the inefficient sgRNA. This results in a complex genetic mosaic of mutant and wild-type alleles (Jao et al., 2013, Moreno-Mateos et al., 2015). Consequently, during blastula period (1k-cell), there are only approximately four germ cells (Raz, 2003). Therefore, if Cas9-induced mutagenesis occurs around this stage, there will be more chance of targeting in the somatic cells than in the germ cells. Hence, Cas9-induced mutagenesis during this embryo stage will result in low of germ-line mutations but highly diverse somatic mutations. Such incomplete mutagenesis is obstructed for creating stable germline mutants. These suggestions are consistent with that an attempt to delete *lncRNA1*, as where the same TS was used. Consequently, anther TS was designed in an attempt to delete both genes from zebrafish genome and this time with high score value.

4.6 investigation the role of lncRNA2 in prRBC by generating triple mutant of qmc222/qmc551/qmc554 homozygous

In order to investigate the role of lncRNA2 in prRBC, we generated a new triple heterozygous line of qmc222/qmc551/qmc554 from the outcross of qmc222 homozygous with qmc551/qmc554 homozygous fish. The heterozygous fish survive to adulthood. As mentioned previously that the Gfi1aa and Gfi1b in zebrafish suppress endothelial gene expression and promote late erythroid gene expression in immature primitive erythroblasts. Therefore, mutations in both genes result in zebrafish with delayed differentiation of primitive erythroblast from mesenchymal haemangioblasts, although differentiation is not completely blocked (Moore et al., 2018). Beside *lncRNA2* is upregulated in the vDA and PBI of qmc551 homozygous mutants by WISH analysis. Therefore, we would like to investigate its role by generated triple mutant.

4.7 Future perspectives

This project presents data that suggested that *lncRNA1* and *lncRNA2* are expressed in primitive and definitive haematopoietic cells. These findings should be confirmed by performing a cross-section of wt embryos, to identify and confirm the location of *lncRNA1*- and *lncRNA2*-expressing cells. Moreover, further *in situ* hybridisation experiments in a mutant line, such as *Runx1* mutants, should be performed to examine the expression of *lncRNa2* during early developmental stages. This project has also generated several mutant lines, in which both *lncRNA1* and *lncRNA2* have been deleted or in which *lncRNA2* alone was deleted, with or without the promoter region for *lncRNA1*. The following remain to be examined:

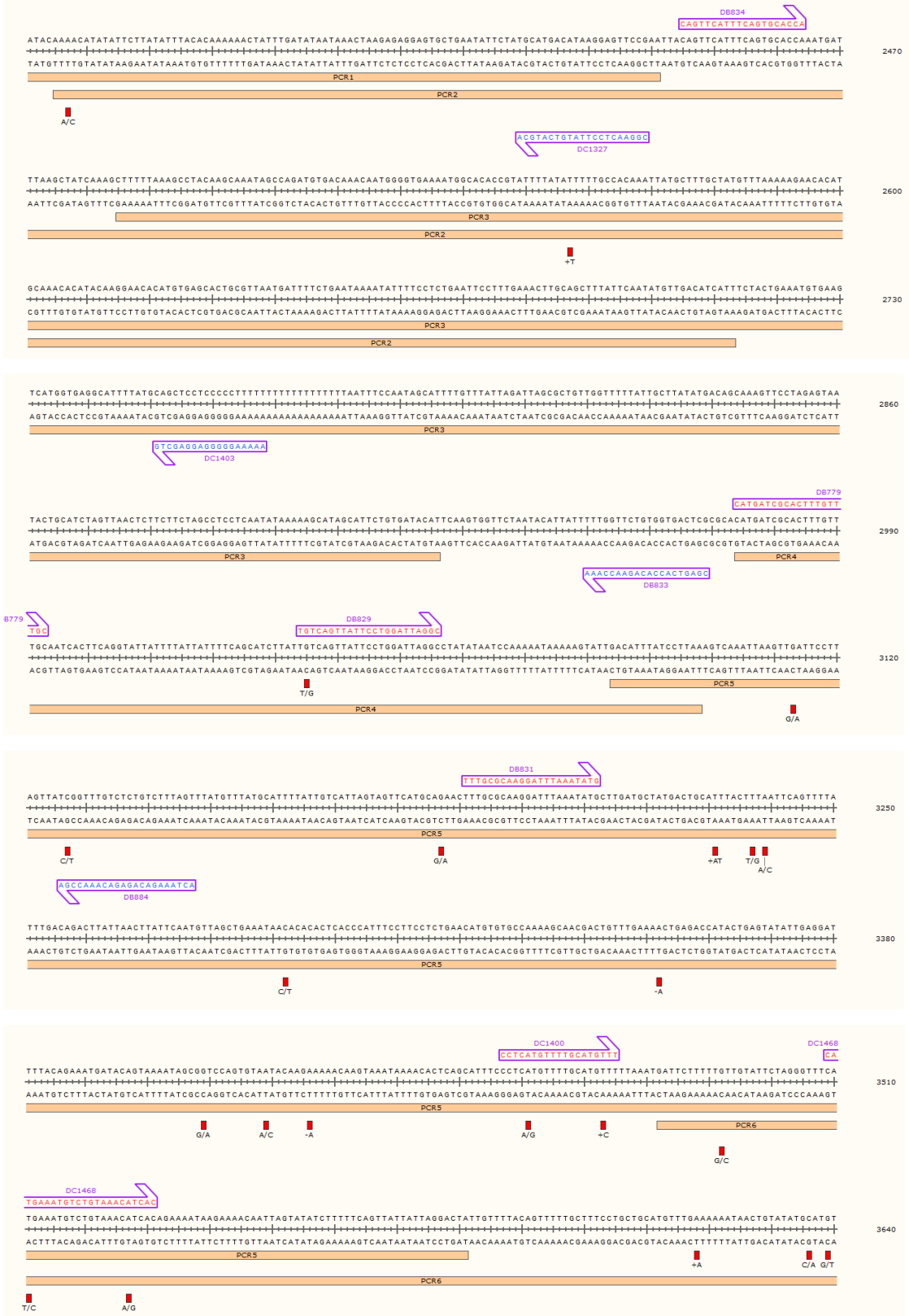
1. qPCR-based expression levels of each mutant (*qmc222*, *qmc223*, and *qmc224*).
2. Isolated RBCs from these mutants should be examined microscopically to determine whether any morphological changes occur.
3. Future studies should examine the expression of *lncRNA1* at early stages in *qmc222* homozygous embryos to determine whether the *lncRNA2* promoter supports *lncRNA1* expression.

5. Appendix

5.1 Appendix A

Sequence of the zebrafish *lncRNA1* and *2* loci. Red box presents the polymorphism. Orange bars represent the PCR products.



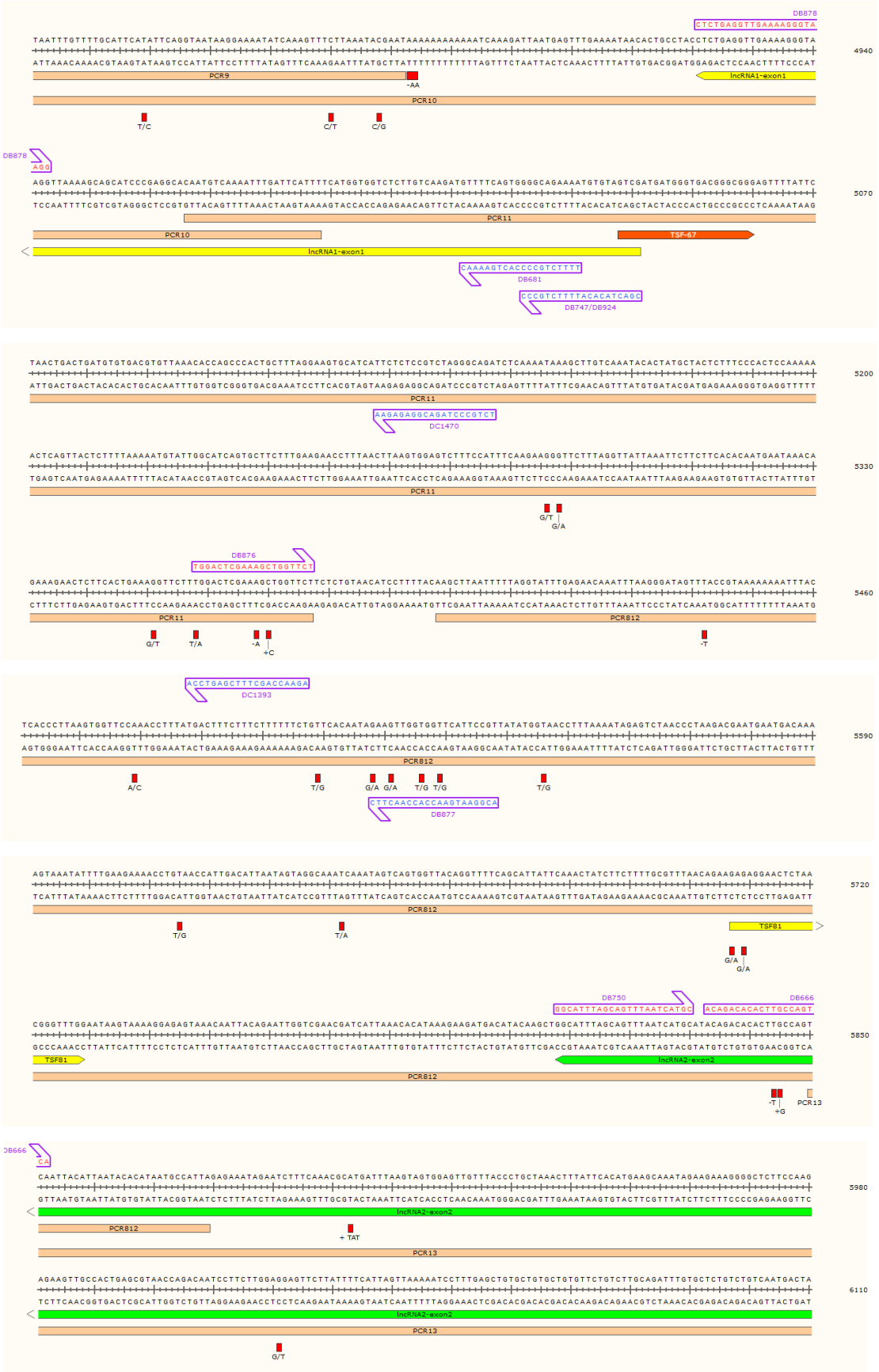

DB834
CAGTTCATTTCAGTGCACCA
ATACAAAACATATATTCTTATATTTACACAAAAAACTATTTGATATAATAAACTAAGAGAGGAGTGCTGAATATTCTATGCATGACATAAGGAGTTCCGAATTACAGTTCATTTCAGTGCACCAAATGAT
TATGTTTTGTATATAAGAATATAAATGTGTTTTTTGATAAACTATATTATTTGATTCTCTCCTCACGACTTATAAGATACGTACTGTATTCCTCAAGGCTTAATGTCAAGTAAAGTCACGTGGTTTACTA
2470
PCR1
PCR2
A/C
ACGTACTGTATTCCTCAAGGC
DC1327
TTAAGCTATCAAAGCTTTTTAAAGCCTACAAGCAAATAGCCAGATGTGACAAACAATGGGGTGAAAATGGCACACCGTATTTTATATTTTTGCCACAAATTATGCTTTGCTATGTTTAAAAAGAACACAT
AATTCGATAGTTTCGAAAAATTTCGGATGTTCGTTTATCGGTCTACACTGTTTGTTACCCCACTTTTACCGTGTGGCATAAAATATAAAAACGGTGTTTAATACGAAACGATACAAATTTTTCTTGTGTA
2600
PCR3
PCR2
+T
GCAAACACATACAAGGAACACATGTGAGCACTGCGTTAATGATTTTCTGAATAAAATATTTTCCTCTGAATTCCTTTGAAACTTGCAGCTTTATTCAATATGTTGACATCATTTCTACTGAAATGTGAAG
CGTTTGTGTATGTTCCTTGTGTACACTCGTGACGCAATTACTAAAAGACTTATTTTATAAAAGGAGACTTAAGGAAACTTTGAACGTCGAAATAAGTTATACAACTGTAGTAAAGATGACTTTACACTTC
2730
PCR3
PCR2

TCATGGTGAGGCATTTTATGCAGCTCCTCCCCCTTTTTTTTTTTTTTTTTAATTTCCAATAGCATTTTGTTTATTAGATTAGCGCTGTTGGTTTTTATTGCTTATATGACAGCAAAGTTCCTAGAGTAA
AGTACCACTCCGTAAAATACGTCGAGGAGGGGGAAAAAAAAAAAAAAAAAAAATTAAAGGTTATCGTAAAACAAATAATCTAATCGCGACAACCAAAAATAACGAATATACTGTCGTTTCAAGGATCTCATT
2860
PCR3
GTCGAGGAGGGGGAAAAA
DC1403
DB779
CATGATCGCACTTTGTT
TACTGCATCTAGTTAACTCTTCTTCTAGCCTCCTCAATATAAAAAGCATAGCATTCTGTGATACATTCAAGTGGTTCTAATACATTATTTTTGGTTCTGTGGTGACTCGCGCACATGATCGCACTTTGTT
ATGACGTAGATCAATTGAGAAGAAGATCGGAGGAGTTATATTTTTCGTATCGTAAGACACTATGTAAGTTCACCAAGATTATGTAATAAAAACCAAGACACCACTGAGCGCGTGTACTAGCGTGAAACAA
2990
PCR3
PCR4
AAACCAAGACACCACTGAGC
DB833
B779
TGC
DB829
TGTCAGTTATTCCTGGATTAGGC
TGCAATCACTTCAGGTATTATTTTATTATTTTCAGCATCTTATTGTCAGTTATTCCTGGATTAGGCCTATATAATCCAAAAATAAAAAGTATTGACATTTATCCTTAAAGTCAAATTAAGTTGATTCCTT
ACGTTAGTGAAGTCCATAATAAAATAATAAAAGTCGTAGAATAACAGTCAATAAGGACCTAATCCGGATATATTAGGTTTTTATTTTTCATAACTGTAAATAGGAATTTCAGTTTAATTCAACTAAGGAA
3120
T/G
PCR5
PCR4
G/A

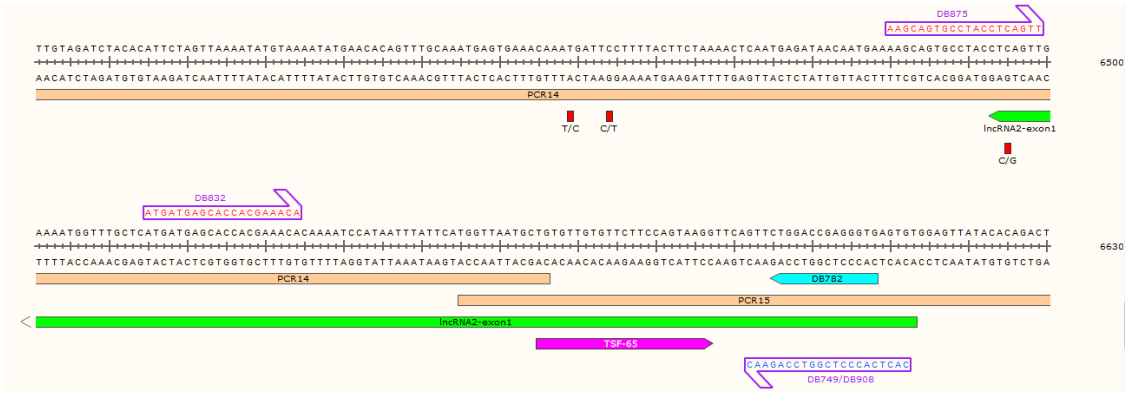
DB831
TTTGCGCAAGGATTTAAATATG
AGTTATCGGTTTGTCTCTGTCTTTAGTTTATGTTTATGCATTTTATTGTCATTAGTAGTTCATGCAGAACTTTGCGCAAGGATTTAAATATGCTTGATGCTATGACTGCATTTACTTTAATTCAGTTTTA
TCAATAGCCAAACAGAGACAGAAATCAAATACAAATACGTAAAATAACAGTAATCATCAAGTACGTCTTGAAACGCGTTCCTAAATTTATACGAACTACGATACTGACGTAAATGAAATTAAGTCAAAAT
3250
PCR5
C/T
G/A
+AT
T/G
A/C
AGCCAAACAGAGACAGAAATCA
DB884
TTTGACAGACTTATTAACTTATTCAATGTTAGCTGAAATAACACACACTCACCCATTTCCTTCCTCTGAACATGTGTGCCAAAAGCAACGACTGTTTGAAAACTGAGACCATACTGAGTATATTGAGGAT
AAACTGTCTGAATAATTGAATAAGTTACAATCGACTTTATTGTGTGTGAGTGGGTAAAGGAAGGAGACTTGTACACACGGTTTTCGTTGCTGACAAACTTTTGACTCTGGTATGACTCATATAACTCCTA
3380
PCR5
C/T
-A

DC1400
CCTCATGTTTTGCATGTTT
DC1468
CA
TTTACAGAAATGATACAGTAAAATAGCGGTCCAGTGTAATACAAGAAAAACAAGTAAATAAAACACTCAGCATTTCCCTCATGTTTTGCATGTTTTTAAATGATTCTTTTTGTTGTATTCTAGGGTTTCA
AAATGTCTTTACTATGTCATTTTATCGCCAGGTCACATTATGTTCTTTTTGTTCATTTATTTTGTGAGTCGTAAAGGGAGTACAAAACGTACAAAAATTTACTAAGAAAAACAACATAAGATCCCAAAGT
3510
PCR5
PCR6
G/A
A/C
-A
A/G
+C
G/C
DC1468
TGAAATGTCTGTAAACATCAC
TGAAATGTCTGTAAACATCACAGAAAATAAGAAAACAATTAGTATATCTTTTTCAGTTATTATTAGGACTATTGTTTTACAGTTTTTGCTTTCCTGCTGCATGTTTGAAAAAATAACTGTATATGCATGT
ACTTTACAGACATTTGTAGTGTCTTTTATTCTTTTGTTAATCATATAGAAAAAGTCAATAATAATCCTGATAACAAAATGTCAAAAACGAAAGGACGACGTACAAACTTTTTTATTGACATATACGTACA
3640
PCR5
PCR6
+A
C/A
G/T
T/C
A/G


AAGGACGACGTACAAACT
DB830
DB682
TCATCCACTTTAATTGTGCATATAGA
DB748
ATCATCCACTTTAATTGTGCATATA
GAGTAAAATAGGACACTCAAACATATACAATCATCCACTTTAATTGTGCATATAGACATATTCTTAATCATTATAATTACAATCAAGCCAAAAATATTGTATAACTGAATTATCATGTTTAATTTTGAAC
CTCATTTTATCCTGTGAGTTTGTATATGTTAGTAGGTGAAATTAACACGTATATCTGTATAAGAATTAGTAATATTAATGTTAGTTCGGTTTTTATAACATATTGACTTAATAGTACAAATTAAAACTTG
3770
G/A
PCR7
lncRNA1-exon2
PCR6
TSF-14
AAGGAAATCACAGCTGATGGGTATTTTCAAAAATGGGGCAAGAGTCCATCTGTGTGGCTTTGTTGGGTGTTGTCCTTTTAAAGCCGTTTTTTTTTTTTCCAGAAATATTTATATTAACAAATAAACAAAA
TTCCTTTAGTGTCGACTACCCATAAAAGTTTTTACCCCGTTCTCAGGTAGACACACCGAAACAACCCACAACAGGAAAATTTCGGCAAAAAAAAAAAAGGTCTTTATAAATATAATTGTTTATTTGTTTT
3900
PCR7
lncRNA1-exon2
PCR6
+T
C/T
TAGACACACCGAAACAACCC
DB883
CGAAACAACCCACAACAGGA
DB835
TCTTTTTCTAATTTGAATTGCTGTGGCATCTTGCTTTATTTATGTGTCTTCCAGAACTGTGGAAAATCTCTCTCTTCATGTGACTGGCTGTCCGTCTTTCTGGGTTACATCAAAACACTCTCCAGCCAGT
AGAAAAAGATTAAACTTAACGACACCGTAGAACGAAATAAATACACAGAAGGTCTTGACACCTTTTAGAGAGAGAAGTACACTGACCGACAGGCAGAAAGACCCAATGTAGTTTTGTGAGAGGTCGGTCA
4030
PCR7
lncRNA1-exon2
A/C
CACTGACCGACAGGCAGAAA
DC1397
DB882
CAACACATCCAGTCCTGCAG
TCAGTGCATGAAGCTGAGGGTGGCAGCTGGAATCAGTAAACACAGCATGGTGGGCAACACATCCAGTCCTGCAGCATCGGTGGCCCCCGCAGTGTTGGTATTGGTGTTGTTTGTCGGTTGCGAGGCTGGT
AGTCACGTACTTCGACTCCCACCGTCGACCTTAGTCATTTGTGTCGTACCACCCGTTGTGTAGGTCAGGACGTCGTAGCCACCGGGGGCGTCACAACCATAACCACAACAAACAGCCAACGCTCCGACCA
4160
PCR7
lncRNA1-exon2
PCR8
TTTTGCTCACCATCTTTTGCTGAGCCTTTGTCCTACACAAAATCAGAATAAACAGGCAGATCATTATAATGTCCACTATTGTGAAAAATGTGCAGTTGTATACCCTAACGGACATCTGGCTTTTATAAAG
AAAACGAGTGGTAGAAAACGACTCGGAAACAGGATGTGTTTTAGTCTTATTTGTCCGTCTAGTAATATTACAGGTGATAACACTTTTTACACGTCAACATATGGGATTGCCTGTAGACCGAAAATATTTC
4290
PCR7
C/A
lncRNA1-exon2
A/C
PCR8
ACGACTCGGAAACAGGATGT
DB879
TAATTAAAAAATTTTTTTCATTTGTTCTTTTGTTTTGTTTTAGGTAACCCGCCTTTTTCATATGCAATGTGAAATCAGCTTAAAGTTATTTAAAGTTGAATTAAAGTTATTTTGATAGAAGGTGCCTTAA
ATTAATTTTTTAAAAAAAGTAAACAAGAAAACAAAACAAAATCCATTGGGCGGAAAAAGTATACGTTACACTTTAGTCGAATTTCAATAAATTTCAACTTAATTTCAATAAAACTATCTTCCACGGAATT
4420
PCR8
+T
DC1469
TCACCCAAATATGTCCTCACC
AATGCACTTTATTGTTATTACATTTGTAATTCTTTATGATGCTTCATCAGTCATTCACCCAAATATGTCCTCACCATTTACTCACACTAAAACCATTATGATTTTTTATCTGTTGAGCCCAAAACTAAAT
TTACGTGAAATAACAATAATGTAAACATTAAGAAATACTACGAAGTAGTCAGTAAGTGGGTTTATACAGGAGTGGTAAATGAGTGTGATTTTGGTAATACTAAAAAATAGACAACTCGGGTTTTGATTTA
4550
PCR8
PCR9
C/T
C/T
+A
+T
DC1396
GGAAACCTGCAGCCATTG
GTTCTGAAGAATGTTGGAAACCTGCAGCCATTGATATCCATAGTCGGAAAAATATTATTAAAAGGAAGATAATGGCTGTTTTTCTCAACATTCTTCAGTATATCTTCTTTTGTGTTCAACTGAAAAAAAG
CAAGACTTCTTACAACCTTTGGACGTCGGTAACTATAGGTATCAGCCTTTTTATAATAATTTTCCTTCTATTACCGACAAAAAGAGTTGTAAGAAGTCATATAGAAGAAAACACAAGTTGACTTTTTTTC
4680
PCR9
A/G
-A
AAATTGATTTGGAACAAGTGGAGGATGAGTGATTAATGGCCGAATTTTCAGTTTTGGTGAACTAGCCAAGAGTCCATTTTAACAAATGACCCAAACTTCTAATTACTTGGTTTAAAAGAACTGGAGGCAA
TTTAACTAAACCTTGTTCACCTCCTACTCACTAATTACCGGCTTAAAAGTCAAAACCACTTGATCGGTTCTCAGGTAAAATTGTTTACTGGGTTTGAAGATTAATGAACCAAATTTTCTTGACCTCCGTT
4810
PCR9
PCR10
A/T
A/T
T/G


DB878
CTCTGAGGTTGAAAAGGGTA
TAATTTGTTTTGCATTCATATTCAGGTAATAAGGAAAATATCAAAGTTTCTTAAATACGAATAAAAAAAAAAAATCAAAGATTAATGAGTTTGAAAATAACACTGCCTACCTCTGAGGTTGAAAAGGGTA
4940
ATTAAACAAAACGTAAGTATAAGTCCATTATTCCTTTTATAGTTTCAAAGAATTTATGCTTATTTTTTTTTTTTAGTTTCTAATTACTCAAACTTTTATTGTGACGGATGGAGACTCCAACTTTTCCCAT
PCR9
-AA
lncRNA1-exon1
PCR10
T/C
C/T
C/G
DB878
AGG
AGGTTAAAAGCAGCATCCCGAGGCACAATGTCAAAATTTGATTCATTTTCATGGTGGTCTCTTGTCAAGATGTTTTCAGTGGGGCAGAAAATGTGTAGTCGATGATGGGTGACGGGCGGGAGTTTTATTC
5070
TCCAATTTTCGTCGTAGGGCTCCGTGTTACAGTTTTAAACTAAGTAAAAGTACCACCAGAGAACAGTTCTACAAAAGTCACCCCGTCTTTTACACATCAGCTACTACCCACTGCCCGCCCTCAAAATAAG
PCR11
PCR10
TSF-67
lncRNA1-exon1
CAAAAGTCACCCCGTCTTTT
DB681
CCCGTCTTTTACACATCAGC
DB747/DB924
TAACTGACTGATGTGTGACGTGTTAAACACCAGCCCACTGCTTTAGGAAGTGCATCATTCTCTCCGTCTAGGGCAGATCTCAAAATAAAGCTTGTCAAATACACTATGCTACTCTTTCCCACTCCAAAAA
5200
ATTGACTGACTACACACTGCACAATTTGTGGTCGGGTGACGAAATCCTTCACGTAGTAAGAAGAGGCAGATCCCGTCTAGAGTTTTATTTCGAACAGTTTATGTGATACGATGAGAAAGGGTGAGGTTTTT
PCR11
AAGAGAGGCAGATCCCGTCT
DC1470
ACTCAGTTACTCTTTTAAAAATGTATTGGCATCAGTGCTTCTTTGAAGAACCTTTAACTTAAGTGGAGTCTTTCCATTTCAAGAAGGGTTCTTTAGGTTATTAAATTCTTCTTCACACAATGAATAAACA
5330
TGAGTCAATGAGAAAATTTTTACATAACCGTAGTCACGAAGAAACTTCTTGGAAATTGAATTCACCTCAGAAAGGTAAAGTTCTTCCCAAGAAATCCAATAATTTAAGAAGAAGTGTGTTACTTATTTGT
PCR11
G/T
G/A
DB876
TGGACTCGAAAGCTGGTTCT
GAAAGAACTCTTCACTGAAAGGTTCTTTGGACTCGAAAGCTGGTTCTTCTCTGTAACATCCTTTTACAAGCTTAATTTTTAGGTATTTGAGAACAAATTTAAGGGATAGTTTACCGTAAAAAAAATTTAC
5460
CTTTCTTGAGAAGTGACTTTCCAAGAAACCTGAGCTTTCGACCAAGAAGAGACATTGTAGGAAAATGTTCGAATTAAAAATCCATAAACTCTTGTTTAAATTCCCTATCAAATGGCATTTTTTTTAAATG
PCR11
PCR812
G/T
T/A
-A
+C
-T
ACCTGAGCTTTCGACCAAGA
DC1393
TCACCCTTAAGTGGTTCCAAACCTTTATGACTTTCTTTCTTTTTTCTGTTCACAATAGAAGTTGGTGGTTCATTCCGTTATATGGTAACCTTTAAAATAGAGTCTAACCCTAAGACGAATGAATGACAAA
5590
AGTGGGAATTCACCAAGGTTTGGAAATACTGAAAGAAAGAAAAAAGACAAGTGTTATCTTCAACCACCAAGTAAGGCAATATACCATTGGAAATTTTATCTCAGATTGGGATTCTGCTTACTTACTGTTT
PCR812
A/C
T/G
G/A
G/A
T/G
T/G
T/G
CTTCAACCACCAAGTAAGGCA
DB877
AGTAAATATTTTGAAGAAAACCTGTAACCATTGACATTAATAGTAGGCAAATCAAATAGTCAGTGGTTACAGGTTTTCAGCATTATTCAAACTATCTTCTTTTGCGTTTAACAGAAGAGAGGAACTCTAA
5720
TCATTTATAAAACTTCTTTTGGACATTGGTAACTGTAATTATCATCCGTTTAGTTTATCAGTCACCAATGTCCAAAAGTCGTAATAAGTTTGATAGAAGAAAACGCAAATTGTCTTCTCTCCTTGAGATT
PCR812
T/G
T/A
TSF81
G/A
G/A
DB750
GGCATTTAGCAGTTTAATCATGC
DB666
ACAGACACACTTGCCAGT
CGGGTTTGGAATAAGTAAAAGGAGAGTAAACAATTACAGAATTGGTCGAACGATCATTAAACACATAAAGAAGATGACATACAAGCTGGCATTTAGCAGTTTAATCATGCATACAGACACACTTGCCAGT
5850
GCCCAAACCTTATTCATTTTCCTCTCATTTGTTAATGTCTTAACCAGCTTGCTAGTAATTTGTGTATTTCTTCTACTGTATGTTCGACCGTAAATCGTCAAATTAGTACGTATGTCTGTGTGAACGGTCA
TSF81
lncRNA2-exon2
PCR812
-T
+G
PCR13
DB666
CA
CAATTACATTAATACACATAATGCCATTAGAGAAATAGAATCTTTCAAACGCATGATTTAAGTAGTGGAGTTGTTTACCCTGCTAAACTTTATTCACATGAAGCAAATAGAAGAAAGGGGCTCTTCCAAG
5980
GTTAATGTAATTATGTGTATTACGGTAATCTCTTTATCTTAGAAAGTTTGCGTACTAAATTCATCACCTCAACAAATGGGACGATTTGAAATAAGTGTACTTCGTTTATCTTCTTTCCCCGAGAAGGTTC
lncRNA2-exon2
PCR812
+ TAT
PCR13
AGAAGTTGCCACTGAGCGTAACCAGACAATCCTTCTTGGAGGAGTTCTTATTTTCATTAGTTAAAAATCCTTTGAGCTGTGCTGTGCTGTGTTCTGTCTTGCAGATTTGTGCTCTGTCTGTCAATGACTA
6110
TCTTCAACGGTGACTCGCATTGGTCTGTTAGGAAGAACCTCCTCAAGAATAAAAGTAATCAATTTTTAGGAAACTCGACACGACACGACACAAGACAGAACGTCTAAACACGAGACAGACAGTTACTGAT
lncRNA2-exon2
PCR13
G/T

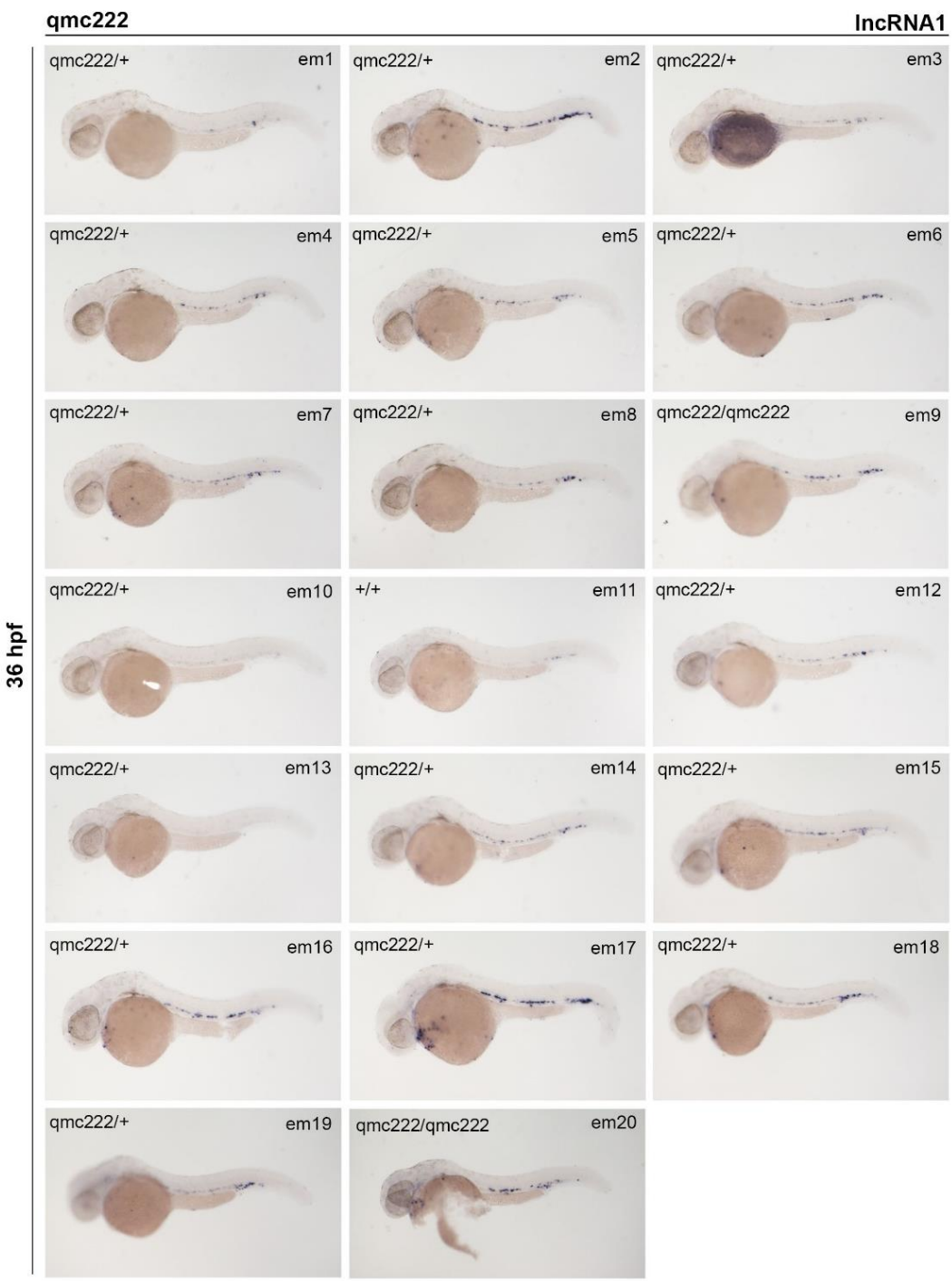

DB907
TCTGCCACTCCCACATTGAT
TCTATTCTGCCACTCCCACATTGATGCTCGCCTTTGGTGAATGTTCAGGATCTGCTGTGGATGAAGCCCAGGGTGGCAGCCGGGAACAGAAGGTACAGTGTTTTGGGCACAGGACTCAGTCCTACAGCAC
AGATAAGACGGTGAGGGTGTAACTACGAGCGGAAACCACTTACAAGTCCTAGACGACACCTACTTCGGGTCCCACCGTCGGCCCTTGTCTTCCATGTCACAAAACCCGTGTCCTGAGTCAGGATGTCGTG
lncRNA2-exon2
PCR14
PCR13
A/T
A/T
TAGACGACACCTACTTCGGG
DB665
TGGAGTTAGTAGATGCGCTGGGCGATGGGACAGTGTTTTCACTTGACTTCTGACTCTCCTACATCAGACATGAAGCAGTCACTTTACAGCATCACAGATAAAATAAAAAATAAAAACATTCTGGAGCTGT
ACCTCAATCATCTACGCGACCCGCTACCCTGTCACAAAAGTGAACTGAAGACTGAGAGGATGTAGTCTGTACTTCGTCAGTGAAATGTCGTAGTGTCTATTTTATTTTTTATTTTTGTAAGACCTCGACA
lncRNA2-exon2
PCR14
GTGAACTGAAGACTGAGAGGAT
DB925
6240
6370

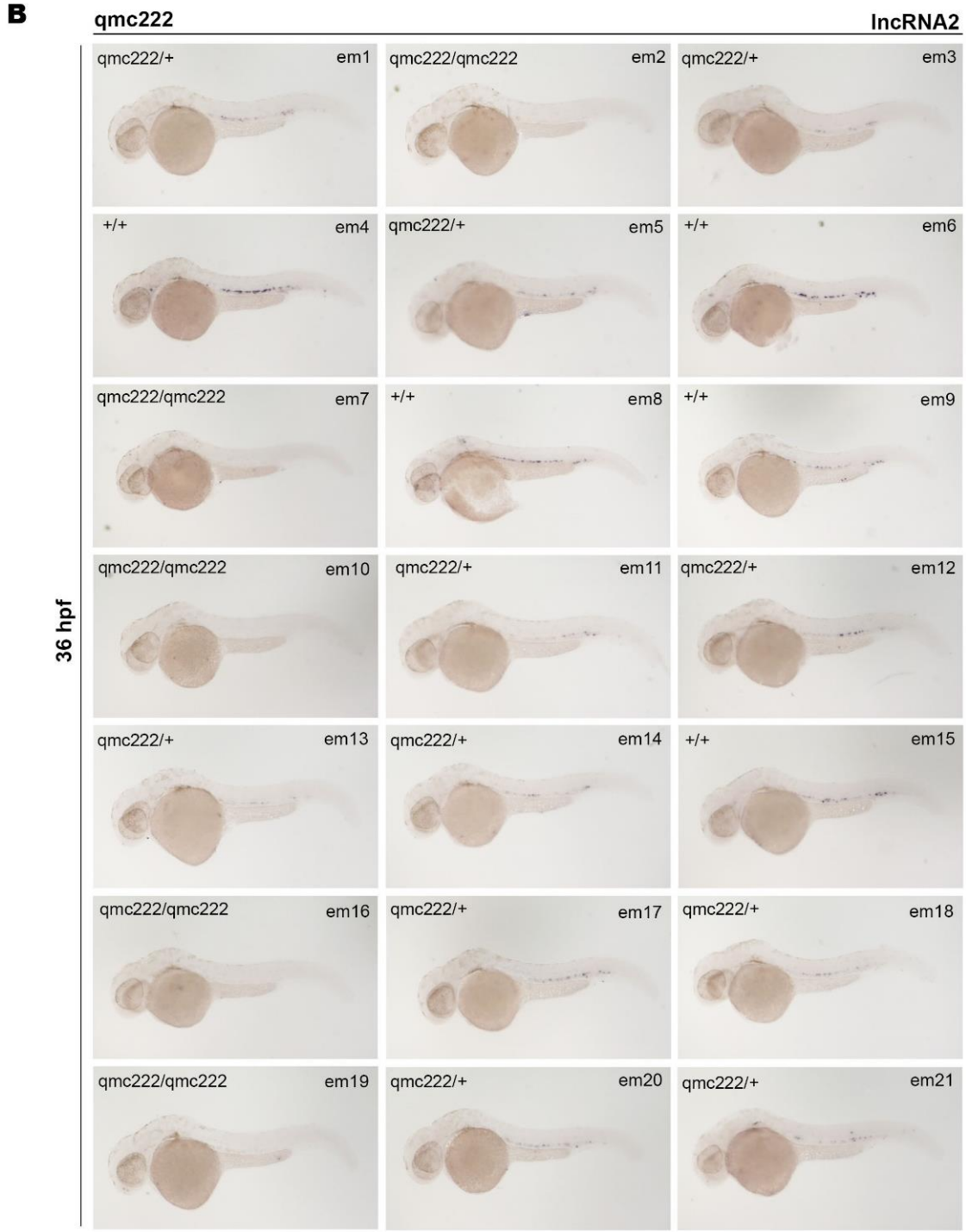

DB875
AAGCAGTGCCTACCTCAGTT
TTGTAGATCTACACATTCTAGTTAAAATATGTAAAATATGAACACAGTTTGCAAATGAGTGAAACAAATGATTCCTTTTACTTCTAAAACTCAATGAGATAACAATGAAAAGCAGTGCCTACCTCAGTTG
AACATCTAGATGTGTAAGATCAATTTTATACATTTTATACTTGTGTCAAACGTTTACTCACTTTGTTTACTAAGGAAAATGAAGATTTTGAGTTACTCTATTGTTACTTTTCGTCACGGATGGAGTCAAC
PCR14
T/C
C/T
lncRNA2-exon1
C/G
DB832
ATGATGAGCACCACGAAACA
AAAATGGTTTGCTCATGATGAGCACCACGAAACACAAAATCCATAATTTATTCATGGTTAATGCTGTGTTGTGTTCTTCCAGTAAGGTTCAGTTCTGGACCGAGGGTGAGTGTGGAGTTATACACAGACT
TTTTACCAAACGAGTACTACTCGTGGTGCTTTGTGTTTTAGGTATTAAATAAGTACCAATTACGACACAACACAAGAAGGTCATTCCAAGTCAAGACCTGGCTCCCACTCACACCTCAATATGTGTCTGA
PCR14
DB782
PCR15
lncRNA2-exon1
TSF-65
CAAGACCTGGCTCCCACTCAC
DB749/DB908
6500
6630


DB704
TAGGAAAAATGGGCGGGGGAAA
GGACATTTCAGTTATAAACACCCTCCTATCTGTGAGTTTGAGAATTGCAGATAAGAGATGACCAATGGAGGAGGGGATTTAGGAAAAATGGGCGGGGGAAAGGGTGTGTATTTTTTAGGTTCTGACTATT
CCTGTAAAGTCAATATTTGTGGGAGGATAGACACTCAAACTCTTAACGTCTATTCTCTACTGGTTACCTCCTCCCCTAAATCCTTTTTACCCGCCCCCTTTCCCACACATAAAAAATCCAAGACTGATAA
PCR15
CTGGTTACCTCCTCCCCTAA
DB874
TTTGCAGCTTCTGGCCCATGAACATCTCCTAGCACAAAGTAATGTGACTTGTTTTCACATTTTTTACTGAACTGTACTACACATATGCCCCATGGTGTTACAGCAAGAGGTTATAATGATGTACAAATAG
AAACGTCGAAGACCGGGTACTTGTAGAGGATCGTGTTTCATTACACTGAACAAAAGTGTAAAAAATGACTTGACATGATGTGTATACGGGGTACCACAATGTCGTTCTCCAATATTACTACATGTTTATC
GGTACTTGTAGAGGATCGTGTTT
DB881
ATGTGTATACGGGGTACCAC
DB778
CATGAGAGTGATTAATCTTCACTGCTGGATGCATATGGTTATACAATGCTAGTTATTTTTTATGAAAATATGAATAACTTTATTTTTTATTTCATCACAAGCCAGACTTGTTGGATATCTTCAGTAAATA
GTACTCTCACTAATTAGAAGTGACGACCTACGTATACCAATATGTTACGATCAATAAAAAATACTTTTATACTTATTGAAATAAAAAATAAAGTAGTGTTCGGTCTGAACAACCTATAGAAGTCATTTAT
CGGTCTGAACAACCTATAGAAGT
DB880
CAAGACAAACTTTGCTGTCATTTTTTCAAAACTACAGAAAATCATATAATTTTAAGATGAGTGCTGATTATATATCCATTACTGCAATACTCAATTTTTCTTTTAAGATGTTGTGAGATAAAAACAAATA
GTTCTGTTTGAAACGACAGTAAAAAAGTTTTGATGTCTTTTAGTATATTAAAATTCTACTCACGACTAATATATAGGTAATGACGTTATGAGTTAAAAAGAAAATTCTACAACACTCTATTTTTGTTTAT
6760
6890
7020
7150

5.2 Appendix B

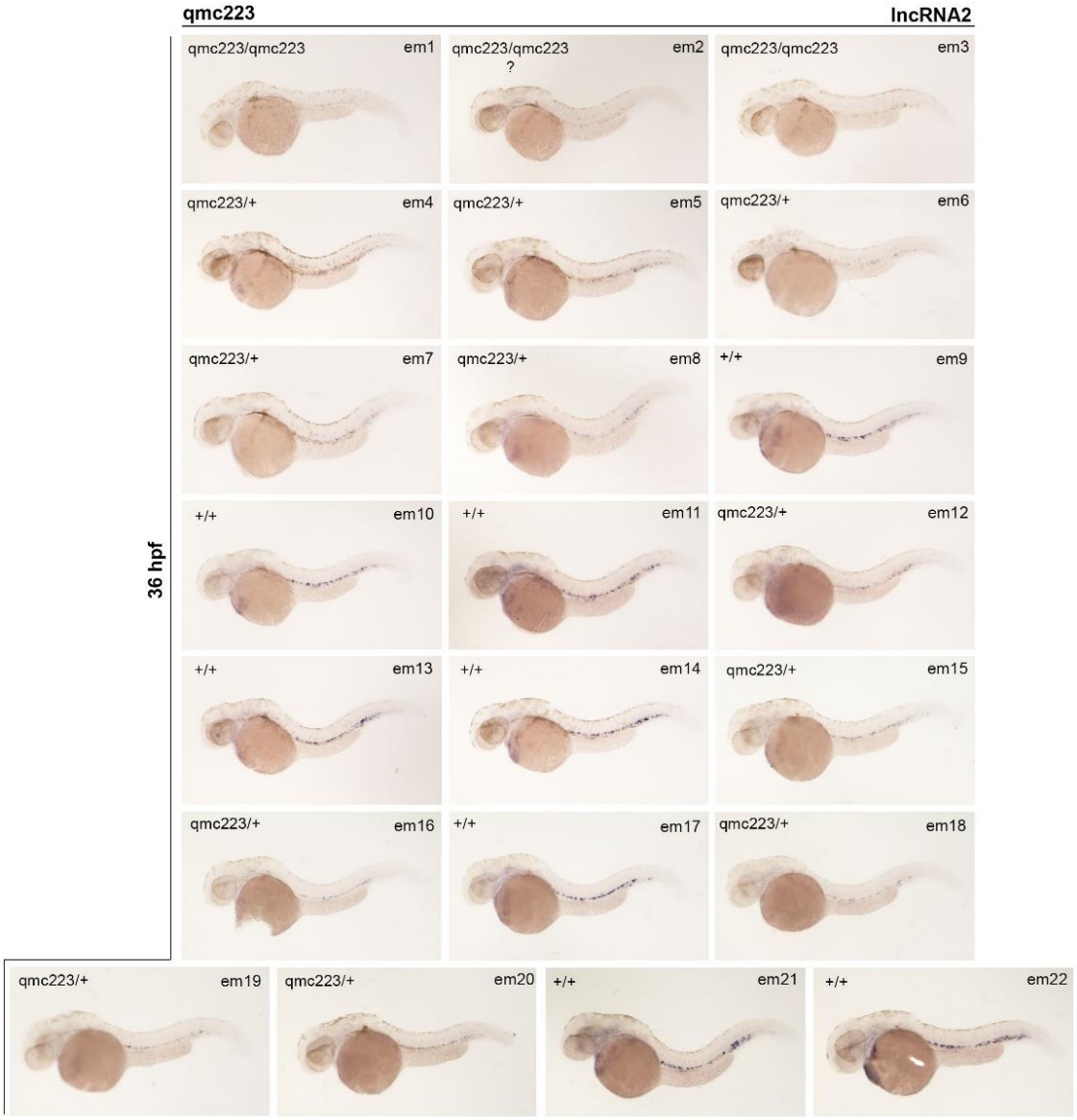
WISH and genotyping data for all *qmc222* embryos of (*lncRNA1* (A) *lncRNA2* (B)).




B
qmc222
lncRNA2
36 hpf
qmc222/+ em1
qmc222/qmc222 em2
qmc222/+ em3
+/+ em4
qmc222/+ em5
+/+ em6
qmc222/qmc222 em7
+/+ em8
+/+ em9
qmc222/qmc222 em10
qmc222/+ em11
qmc222/+ em12
qmc222/+ em13
qmc222/+ em14
+/+ em15
qmc222/qmc222 em16
qmc222/+ em17
qmc222/+ em18
qmc222/qmc222 em19
qmc222/+ em20
qmc222/+ em21

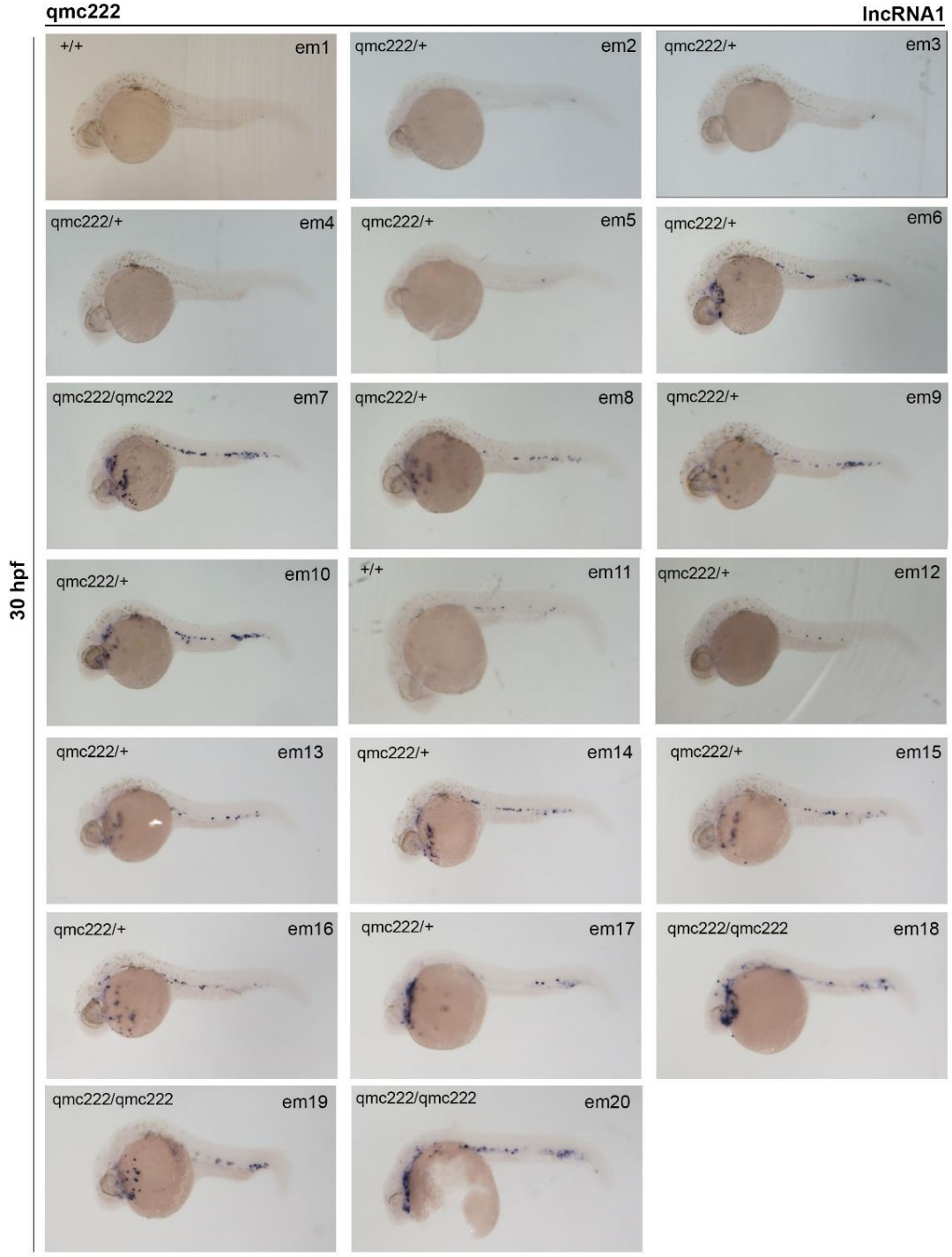
5.3 Appendix C

WISH and genotyping data for all *qmc223* embryos of *lncRNA2*.



5.4 Appendix D

WISH and genotyping data for all *qmc222* embryos of *lncRNA1*.



5.5 Appendix E

lncRNA1 and lncRNA2 probes generation

lncRNA1 and *lncRNA2* cDNAs were amplified by PCR using sequence specific primers (Figure 5.1A). The PCR fragment was ligated to the pGEM-Teasy plasmid fragment using T4 DNA ligase. The cDNA was, thereby, inserted between the SP6 and T7 bacteriophage promoters (Figure 5.1Bi-ii). The ligation mix was then used to transform *E.coli,* and Ampicillin-resistant colonies were selected (Figure 5.1C). Plasmid DNA was isolated first in a mini scale followed by EcoRI restriction digests to confirm the insertion of a DNA fragment of the correct size (Figure 5.2Ai-Bi). Recombinant plasmids were then prepared on a medium scale and sequenced using primer M13F to confirm that the correct fragment was inserted into the plasmid (Figure 5.2Aii-Bii). The DNA sequence also established the orientation of the inserts in the respective plasmids, defining the restriction enzyme and the bacteriophage RNA polymerase needed to linearise the plasmids and in vitro transcribe the DNAs, respectively. Both plasmids were first linearised using the restriction enzyme *ApaI* that cuts upstream of the genes of interest. Next, in vitro transcription was performed in the presence of digoxigenin (DIG)-labelled uridine-triphosphate (UTP) using RNA polymerase SP6 to create DIG-labelled antisense RNA probes (Figure 5.2Aiii-Biii).

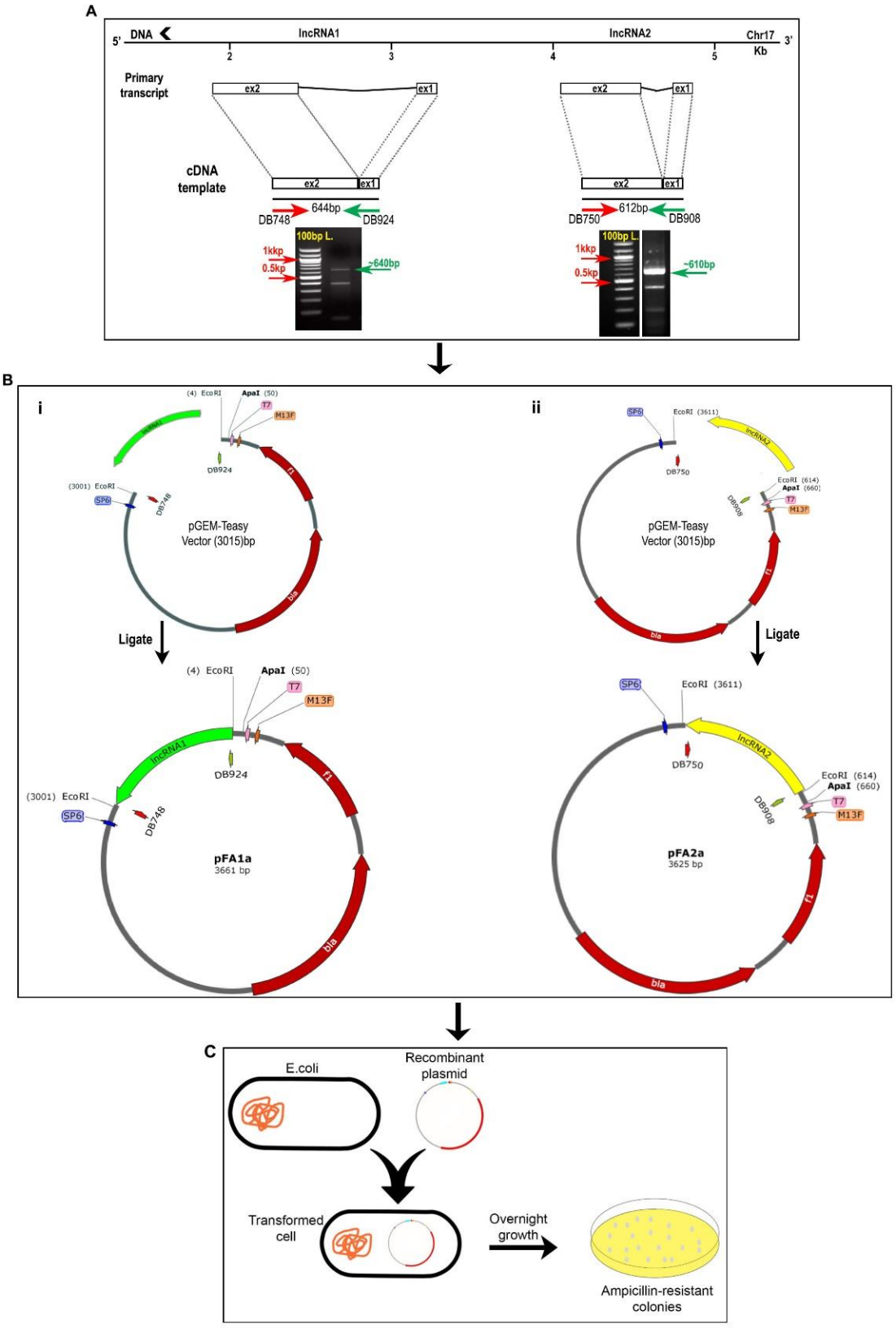


Figure 5.1 *lncRNA1* and *lncRNA2* probes synthesis for whole-mount in situ hybridisation

(A) The sequences of interest were amplified by PCR using sequence-specific primers. **(B)** (i) a 644 bp *lncRNA1* fragment and (ii) a 612 bp

lncRNA2 fragment were ligated into a pGEM-Teasy vector using T4 DNA ligase. **(C)** The ligation mix was used to transform *E.coli*. Ampicillin-resistant colonies were identified on ampicillin-containing agar plates.

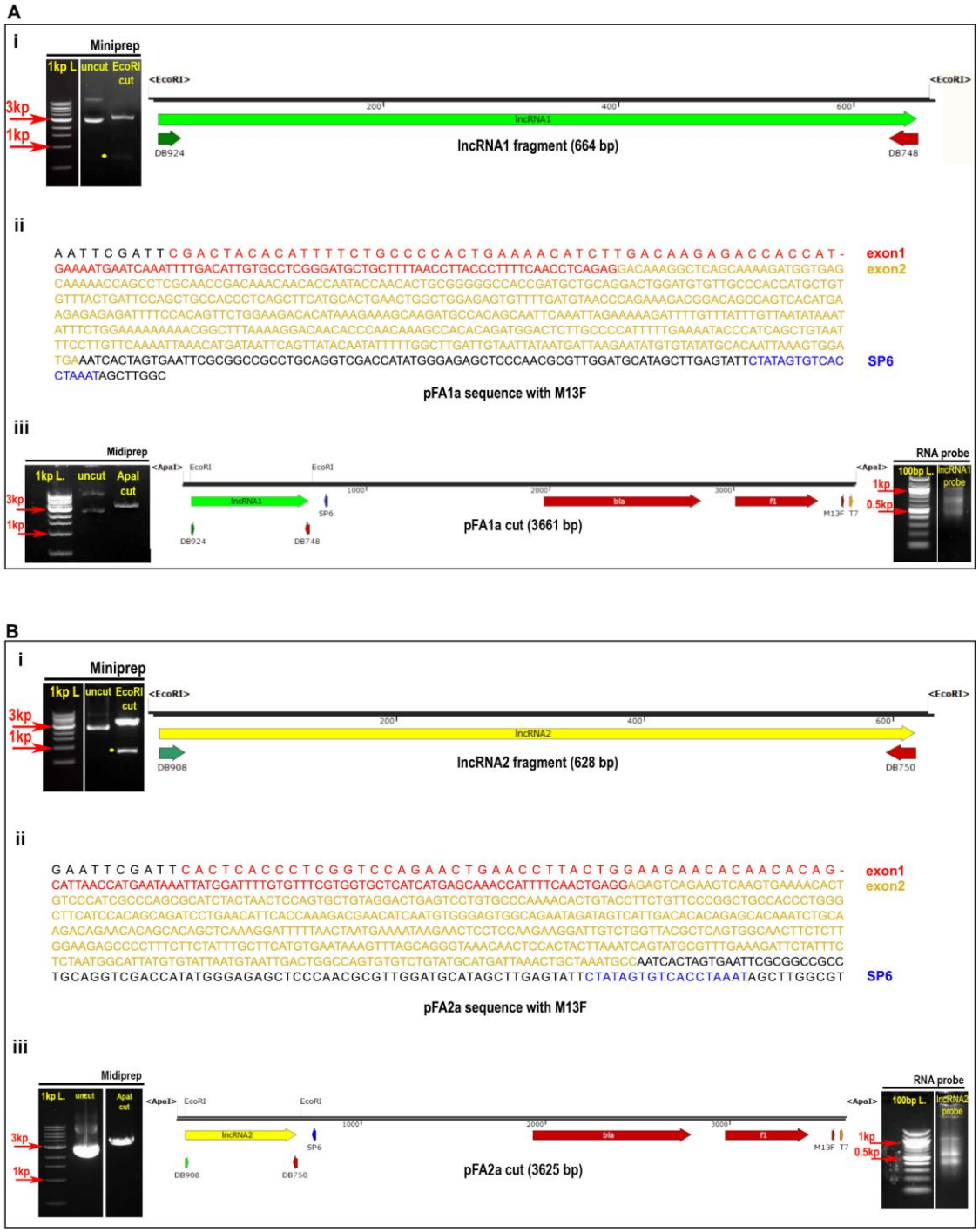


Figure 5.2 Continue *lncRNA1* and *lncRNA2* probes synthesis for whole-mount in situ hybridisation. (A) *lncRNA1* and (B) *lncRNA2*

(Ai) Plasmid DNAs isolated in mini scale preparations were subjected to EcoRI restriction digests. **(Aii)** Recombinant plasmids were prepared at medium scale and then sequenced using primer M13F. The plasmid sequence colour coded, as follows: black: pGEM-Teasy, red: lncRNA1 or lncRNA2 exon1, yellow: *lncRNA1* or *lncRNA2* exon2, and blue: polymerase SP6 site. **(Aiii)** Plasmid DNA was linearised with the restriction enzyme ApaI. The digests were then examined by agarose gel electrophoresis. The

linearised plasmid was then used as a template for *in vitro* translation. RNA products were analysed by gel electrophoresis.

References

ALADHAMI, M. A. & KUNZ, Y. W. 1977. ONTOGENESIS OF HEMATOPOIETIC SITES IN BRACHYDANIO-RERIO (HAMILTON-BUCHANAN) (TELEOSTEI). *Development Growth & Differentiation,* 19**,** 171-179.

ALVAREZ-DOMINGUEZ, J. R., HU, W., YUAN, B., SHI, J., PARK, S. S., GROMATZKY, A. A., OUDENAARDEN, A. V. & LODISH, H. F. 2014. Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood,* 123**,** 570-581.

BARRANGOU, R. 2015. The roles of CRISPR–Cas systems in adaptive immunity and beyond. *Current Opinion in Immunology,* 32**,** 36-41.

BARTOLOMEI, M. S., ZEMEL, S. & TILGHMAN, S. M. 1991. Parental imprinting of the mouse H19 gene. *Nature,* 351**,** 153-5.

BATISTA, P. J. & CHANG, H. Y. 2013. Long noncoding RNAs: cellular address codes in development and disease. *Cell,* 152**,** 1298-307.

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. & LOOK, A. T. 2001. Myelopoiesis in the zebrafish, Danio rerio. *Blood,* 98**,** 643-651.

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.

BLAGOEVA, P. & STOICHEV, I. 1979. In vivo inhibition of the tyrosinase of hamster melanoma with sodium diethyldithiocarbamate. *Experientia,* 35**,** 1389-90.

BOIERS, C., CARRELHA, J., LUTTEROPP, M., LUC, S., GREEN, J. C., AZZONI, E., WOLL, P. S., MEAD, A. J., HULTQUIST, A., SWIERS, G., PERDIGUERO, E. G., MACAULAY, I. C., MELCHIORI, L., LUIS, T. C., KHARAZI, S., BOURIEZ-JONES, T., DENG, Q., PONTEN, A., ATKINSON, D., JENSEN, C. T., SITNICKA, E., GEISSMANN, F., GODIN, I., SANDBERG, R., DE BRUIJN, M. F. & JACOBSEN, S. E. 2013. Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell Stem Cell,* 13**,** 535-48.

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.

BRANNAN, C. I., DEES, E. C., INGRAM, R. S. & TILGHMAN, S. M. 1990. The product of the H19 gene may function as an RNA. *Mol Cell Biol,* 10**,** 28-36.

BRUNSTEIN, C. G. & WAGNER, J. E. 2006. Umbilical cord blood transplantation and banking. *Annu Rev Med,* 57**,** 403-17.

BURGER, A., LINDSAY, H., FELKER, A., HESS, C., ANDERS, C., CHIAVACCI, E., ZAUGG, J., WEBER, L. M., CATENA, R., JINEK, M., ROBINSON, M. D. & MOSIMANN, C. 2016. Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development,* 143**,** 2025-37.

BURNS, C. E., DEBLASIO, T., ZHOU, Y., ZHANG, J., ZON, L. & NIMER, S. D. 2002. Isolation and characterization of runxa and runxb, zebrafish members of the runt family of transcriptional regulators. *Exp Hematol,* 30**,** 1381-9.

BURNS, C. E., TRAVER, D., MAYHALL, E., SHEPARD, J. L. & ZON, L. I. 2005. Hematopoietic stem cell fate is established by the Notch-Runx pathway. *Genes Dev,* 19**,** 2331-42.

CABEZAS-WALLSCHEID, N., KLIMMECK, D., HANSSON, J., LIPKA, D. B., REYES, A., WANG, Q., WEICHENHAN, D., LIER, A., VON PALESKE, L., RENDERS, S., WUNSCHE, P., ZEISBERGER, P., BROCKS, D., GU, L., HERRMANN, C.,

HAAS, S., ESSERS, M. A. G., BRORS, B., EILS, R., HUBER, W., MILSOM, M. D., PLASS, C., KRIJGSVELD, J. & TRUMPP, A. 2014. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell,* 15**,** 507-522.

CANTOR, A. B. & ORKIN, S. H. 2002. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene,* 21**,** 3368-76.

CECH, T. R. & STEITZ, J. A. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell,* 157**,** 77-94.

CHEN, A. T. & ZON, L. I. 2009. Zebrafish blood stem cells. *J Cell Biochem,* 108**,** 35-42.

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

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

CIAU-UITZ, A., LIU, F. & PATIENT, R. 2010. Genetic control of hematopoietic development in Xenopus and zebrafish. *Int J Dev Biol,* 54**,** 1139-49.

CIAU-UITZ, A., MONTEIRO, R., KIRMIZITAS, A. & PATIENT, R. 2014. Developmental hematopoiesis: ontogeny, genetic programming and conservation. *Exp Hematol,* 42**,** 669-83.

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. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science,* 306**,** 636-40.

CONSORTIUM, E. P., BIRNEY, E., STAMATOYANNOPOULOS, J. A., DUTTA, A., GUIGÓ, R., GINGERAS, T. R., MARGULIES, E. H., WENG, Z., SNYDER, M., DERMITZAKIS, E. T., THURMAN, R. E., KUEHN, M. S., TAYLOR, C. M., NEPH, S., KOCH, C. M., ASTHANA, S., MALHOTRA, A., ADZHUBEI, I., GREENBAUM, J. A., ANDREWS, R. M., FLICEK, P., BOYLE, P. J., CAO, H., CARTER, N. P., CLELLAND, G. K., DAVIS, S., DAY, N., DHAMI, P., DILLON, S. C., DORSCHNER, M. O., FIEGLER, H., GIRESI, P. G., GOLDY, J., HAWRYLYCZ, M., HAYDOCK, A., HUMBERT, R., JAMES, K. D., JOHNSON, B. E., JOHNSON, E. M., FRUM, T. T., ROSENZWEIG, E. R., KARNANI, N., LEE, K., LEFEBVRE, G. C., NAVAS, P. A., NERI, F., PARKER, S. C. J., SABO, P. J., SANDSTROM, R., SHAFER, A., VETRIE, D., WEAVER, M., WILCOX, S., YU, M., COLLINS, F. S., DEKKER, J., LIEB, J. D., TULLIUS, T. D., CRAWFORD, G. E., SUNYAEV, S., NOBLE, W. S., DUNHAM, I., DENOEUD, F., REYMOND, A., KAPRANOV, P., ROZOWSKY, J., ZHENG, D., CASTELO, R., FRANKISH, A., HARROW, J., GHOSH, S., SANDELIN, A., HOFACKER, I. L., BAERTSCH, R., KEEFE, D., DIKE, S., CHENG, J., HIRSCH, H. A., SEKINGER, E. A., LAGARDE, J., ABRIL, J. F., SHAHAB, A., FLAMM, C., FRIED, C., HACKERMÜLLER, J., HERTEL, J., LINDEMEYER, M., MISSAL, K., TANZER, A., WASHIETL, S., KORBEL, J., EMANUELSSON, O., PEDERSEN, J. S., HOLROYD, N., TAYLOR, R., SWARBRECK, D., MATTHEWS, N., DICKSON, M. C., THOMAS, D. J., WEIRAUCH, M. T., et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447**,** 799-816.

COONEY, J. D., HILDICK-SMITH, G. J., SHAFIZADEH, E., MCBRIDE, P. F., CARROLL, K. J., ANDERSON, H., SHAW, G. C., TAMPLIN, O. J., BRANCO, D. S., DALTON, A. J., SHAH, D. I., WONG, C., GALLAGHER, P. G., ZON, L. I., NORTH, T. E. & PAW, B. H. 2013. Teleost growth factor independence

(gfi) genes differentially regulate successive waves of hematopoiesis. *Dev Biol,* 373**,** 431-41.
COSTA, G., KOUSKOFF, V. & LACAUD, G. 2012. Origin of blood cells and HSC production in the embryo. *Trends Immunol,* 33**,** 215-23.
CUMANO, A. & GODIN, I. 2007. Ontogeny of the hematopoietic system. *Annu Rev Immunol,* 25**,** 745-85.
DELÁS, M. J. & HANNON, G. J. 2017. lncRNAs in development and disease: from functions to mechanisms. *Open Biol,* 7.
DELAS, M. J., JACKSON, B. T., KOVACEVIC, T., VANGELISTI, S., MUNERA MARAVILLA, E., WILD, S. A., STORK, E. M., ERARD, N., KNOTT, S. R. V. & HANNON, G. J. 2019. lncRNA Spehd Regulates Hematopoietic Stem and Progenitor Cells and Is Required for Multilineage Differentiation. *Cell Rep,* 27**,** 719-729.e6.
DETRICH, H. W., 3RD, KIERAN, M. W., CHAN, F. Y., BARONE, L. M., YEE, K., RUNDSTADLER, J. A., PRATT, S., RANSOM, D. & ZON, L. I. 1995. Intraembryonic hematopoietic cell migration during vertebrate development. *Proc Natl Acad Sci U S A,* 92**,** 10713-7.
DRAKE, C. J., BRANDT, S. J., TRUSK, T. C. & LITTLE, C. D. 1997. TAL1/SCL is expressed in endothelial progenitor cells/angioblasts and defines a dorsal-to-ventral gradient of vasculogenesis. *Dev Biol,* 192**,** 17-30.
DRISSEN, R., BUZA-VIDAS, N., WOLL, P., THONGJUEA, S., GAMBARDELLA, A., GIUSTACCHINI, A., MANCINI, E., ZRIWIL, A., LUTTEROPP, M., GROVER, A., MEAD, A., SITNICKA, E., JACOBSEN, S. E. W. & NERLOV, C. 2016. Distinct myeloid progenitor–differentiation pathways identified through single-cell RNA sequencing. *Nature Immunology,* 17**,** 666.
DUFOURCQ, P., RASTEGAR, S., STRÄHLE, U. & BLADER, P. 2004. Parapineal specific expression of gfi1 in the zebrafish epithalamus. *Gene Expression Patterns,* 4**,** 53-57.
EBRALIDZE, A. K., GUIBAL, F. C., STEIDL, U., ZHANG, P., LEE, S., BARTHOLDY, B., JORDA, M. A., PETKOVA, V., ROSENBAUER, F., HUANG, G., DAYARAM, T., KLUPP, J., O'BRIEN, K. B., WILL, B., HOOGENKAMP, M., BORDEN, K. L., BONIFER, C. & TENEN, D. G. 2008. PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes Dev,* 22**,** 2085-92.
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**,** 4393-403.
FRAME, J. M., FEGAN, K. H., CONWAY, S. J., MCGRATH, K. E. & PALIS, J. 2016. Definitive Hematopoiesis in the Yolk Sac Emerges from Wnt-Responsive Hemogenic Endothelium Independently of Circulation and Arterial Identity. *Stem Cells,* 34**,** 431-44.
GANSNER, J. M., MADSEN, E. C., MECHAM, R. P. & GITLIN, J. D. 2008. Essential role for fibrillin-2 in zebrafish notochord and vascular morphogenesis. *Dev Dyn,* 237**,** 2844-61.
GERING, M. & PATIENT, R. 2005. Hedgehog Signaling Is Required for Adult Blood Stem Cell Formation in Zebrafish Embryos. *Developmental Cell,* 8**,** 389-400.
GERING, M., RODAWAY, A. R. F., GÖTTGENS, B., PATIENT, R. K. & GREEN, A. R. 1998. The SCL gene specifies haemangioblast development from early mesoderm. *The EMBO Journal,* 17**,** 4029-4045.
GORE, A. V., PILLAY, L. M., VENERO GALANTERNIK, M. & WEINSTEIN, B. M. 2018. The zebrafish: A fintastic model for hematopoietic development and disease. *Wiley Interdiscip Rev Dev Biol,* 7**,** e312.
GOYAMA, S., YAMAGUCHI, Y., IMAI, Y., KAWAZU, M., NAKAGAWA, M., ASAI, T., KUMANO, K., MITANI, K., OGAWA, S., CHIBA, S., KUROKAWA, M. & HIRAI, H. 2004. The transcriptionally active form of AML1 is required for

hematopoietic rescue of the AML1-deficient embryonic para-aortic splanchnopleural (P-Sp) region. *Blood,* 104**,** 3558-3564.
GREEN, A. R., LINTS, T., VISVADER, J., HARVEY, R. & BEGLEY, C. G. 1992. SCL is coexpressed with GATA-1 in hemopoietic cells but is also expressed in developing brain. *Oncogene,* 7**,** 653-60.
GRIMES, H. L., CHAN, T. O., ZWEIDLER-MCKAY, P. A., TONG, B. & TSICHLIS, P. N. 1996. The Gfi-1 proto-oncoprotein contains a novel transcriptional repressor domain, SNAG, and inhibits G1 arrest induced by interleukin-2 withdrawal. *Mol Cell Biol,* 16**,** 6263-72.
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.
HOCK, H., HAMBLEN, M. J., ROOKE, H. M., TRAVER, D., BRONSON, R. T., CAMERON, S. & ORKIN, S. H. 2003. Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation. *Immunity,* 18**,** 109-120.
HOFFBRAND, A. V., CATOVSKY, D. & TUDDENHAM, E. G. D. 2010. *Postgraduate haematology,* Malden, Mass., Blackwell Pub.
HU, W., YUAN, B., FLYGARE, J. & LODISH, H. F. 2011. Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. *Genes Dev,* 25**,** 2573-8.
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**,** 625-30.
HWANG, W. Y., FU, Y., REYON, D., MAEDER, M. L., KAINI, P., SANDER, J. D., JOUNG, J. K., PETERSON, R. T. & YEH, J. R. 2013a. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One,* 8**,** e68708.
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.
ISHINO, Y., KRUPOVIC, M. & FORTERRE, P. 2018. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. *J Bacteriol,* 200.
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., WEISSMAN, A. M., LEWIS, J., CHANDRASEKHARAPPA, S. C. & CHITNIS, A. B. 2003. Mind bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by Delta. *Dev Cell,* 4**,** 67-82.
JAGANNATHAN-BOGDAN, M. & ZON, L. I. 2013. Hematopoiesis. *Development (Cambridge, England),* 140**,** 2463-2467.
JAO, L.-E., WENTE, S. R. & CHEN, W. 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences,* 110**,** 13904-13909.
JI, R. P., PHOON, C. K., ARISTIZABAL, O., MCGRATH, K. E., PALIS, J. & TURNBULL, D. H. 2003. Onset of cardiac function during early mouse embryogenesis coincides with entry of primitive erythroblasts into the embryo proper. *Circ Res,* 92**,** 133-5.
JIN, H., SOOD, R., XU, J., ZHEN, F., ENGLISH, M. A., LIU, P. P. & WEN, Z. 2009. Definitive hematopoietic stem/progenitor cells manifest distinct differentiation output in the zebrafish VDA and PBI. *Development,* 136**,** 647-654.
KALEV-ZYLINSKA, M. L., HORSFIELD, J. A., FLORES, M. V., 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**,** 2015-30.
KALLIANPUR, A. R., JORDAN, J. E. & BRANDT, S. J. 1994. The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis. *Blood,* 83**,** 1200-8.
KARSUNKY, H., ZENG, H., SCHMIDT, T., ZEVNIK, 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.
KOTZIN, J. J., ISEKA, F., WRIGHT, J., BASAVAPPA, M. G., CLARK, M. L., ALI, M. A., ABDEL-HAKEEM, M. S., ROBERTSON, T. F., MOWEL, W. K., JOANNAS, L., NEAL, V. D., SPENCER, S. P., SYRETT, C. M., ANGUERA, M. C., WILLIAMS, A., WHERRY, E. J. & HENAO-MEJIA, J. 2019. The long noncoding RNA Morrbid regulates CD8 T cells in response to viral infection. *Proc Natl Acad Sci U S A,* 116**,** 11916-11925.
KOTZIN, J. J., SPENCER, S. P., MCCRIGHT, S. J., KUMAR, D. B. U., COLLET, M. A., MOWEL, W. K., ELLIOTT, E. N., UYAR, A., MAKIYA, M. A., DUNAGIN, M. C., HARMAN, C. C. D., VIRTUE, A. T., ZHU, S., BAILIS, W., STEIN, J., HUGHES, C., RAJ, A., WHERRY, E. J., GOFF, L. A., KLION, A. D., RINN, J. L., WILLIAMS, A., FLAVELL, R. A. & HENAO-MEJIA, J. 2016. The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. *Nature,* 537**,** 239-243.
KUMANO, K., CHIBA, S., KUNISATO, A., SATA, M., SAITO, T., NAKAGAMI-YAMAGUCHI, E., YAMAGUCHI, T., MASUDA, S., SHIMIZU, K., TAKAHASHI, T., OGAWA, S., HAMADA, Y. & HIRAI, H. 2003. Notch1 but Not Notch2 Is Essential for Generating Hematopoietic Stem Cells from Endothelial Cells. *Immunity,* 18**,** 699-711.
LAWSON, N. D., SCHEER, N., PHAM, V. N., KIM, C.-H., CHITNIS, A. B., CAMPOS-ORTEGA, J. A. & WEINSTEIN, B. M. 2001. Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development,* 128**,** 3675-3683.
LIAO, W., BISGROVE, B. W., SAWYER, H., HUG, B., BELL, B., PETERS, K., GRUNWALD, D. J. & STAINIER, D. Y. 1997. The zebrafish gene cloche acts upstream of a flk-1 homologue to regulate endothelial cell differentiation. *Development,* 124**,** 381-389.
LUO, M., JEONG, M., SUN, D., PARK, HYUN J., RODRIGUEZ, BENJAMIN A. T., XIA, Z., YANG, L., ZHANG, X., SHENG, K., DARLINGTON, GRETCHEN J., LI, W. & GOODELL, MARGARET A. 2015. Long Non-Coding RNAs Control Hematopoietic Stem Cell Function. *Cell Stem Cell,* 16**,** 426-438.
MA, H., HAO, Y., DONG, X., GONG, Q., CHEN, J., ZHANG, J. & TIAN, W. 2012. Molecular Mechanisms and Function Prediction of Long Noncoding RNA. *The Scientific World Journal,* 2012**,** 541786.
MA, L., BAJIC, V. B. & ZHANG, Z. 2013. On the classification of long non-coding RNAs. *RNA Biol,* 10**,** 925-33.
MARSHALL, C. J. & THRASHER, A. J. 2001. The embryonic origins of human haematopoiesis. *Br J Haematol,* 112**,** 838-50.
MCGRATH, K. E., FRAME, J. M., FEGAN, K. H., BOWEN, J. R., CONWAY, S. J., CATHERMAN, S. C., KINGSLEY, P. D., KONISKI, A. D. & PALIS, J. 2015. Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Rep,* 11**,** 1892-904.

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 β-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**,** 1017-31.

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 Adv,* 2**,** 2589-2606.

MOORE, M. A. S. & METCALF, D. 1970. Ontogeny of the Haemopoietic System: Yolk Sac Origin of In Vivo and In Vitro Colony Forming Cells in the Developing Mouse Embryo*. *British Journal of Haematology,* 18**,** 279-296.

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. *Nat Methods,* 12**,** 982-8.

MÖRÖY, T., VASSEN, L., WILKES, B. & KHANDANPOUR, C. 2015. From cytopenia to leukemia: the role of Gfi1 and Gfi1b in blood formation. *Blood,* 126**,** 2561-2569.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5**,** 621-8.

MUCENSKI, M. L., MCLAIN, K., KIER, A. B., SWERDLOW, S. H., SCHREINER, C. M., MILLER, T. A., PIETRYGA, D. W., SCOTT, W. J., JR. & POTTER, S. S. 1991. A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell,* 65**,** 677-89.

NORTH, T., GU, T. L., STACY, T., WANG, Q., HOWARD, L., BINDER, M., MARIN-PADILLA, M. & SPECK, N. A. 1999. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development,* 126**,** 2563-75.

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.

ORKIN, S. H. & ZON, L. I. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell,* 132**,** 631-44.

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.

PALIS, J. & YODER, M. C. 2001. Yolk-sac hematopoiesis: the first blood cells of mouse and man. *Exp Hematol,* 29**,** 927-36.

PALLISTER, C. J. & WATSON, M. S. 2011. *Haematology,* Bloxham, Scion.

PARALKAR, V. R., MISHRA, T., LUAN, J., YAO, Y., KOSSENKOV, A. V., ANDERSON, S. M., DUNAGIN, M., PIMKIN, M., GORE, M., SUN, D., KONUTHULA, N., RAJ, A., AN, X., MOHANDAS, N., BODINE, D. M., HARDISON, R. C. & WEISS, M. J. 2014. Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood,* 123**,** 1927-37.

PAULI, A., VALEN, E., LIN, M. F., GARBER, M., VASTENHOUW, N. L., LEVIN, J. Z., FAN, L., SANDELIN, A., RINN, J. L., REGEV, A. & SCHIER, A. F. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res,* 22**,** 577-91.

PETERSDORF, E. W. 2013. The major histocompatibility complex: a model for understanding graft-versus-host disease. *Blood,* 122**,** 1863-72.
PORCHER, C., SWAT, W., ROCKWELL, K., FUJIWARA, Y., ALT, F. W. & ORKIN, S. H. 1996. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell,* 86**,** 47-57.
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., FIDDES, I., KUO, T., CHUNG, W. S., SALEK, S., LERRIGO, R., ALSIO, J., LUO, S., TWORUS, D., AUGUSTINE, S. M., MUCENIEKS, S., NYSTEDT, B., GIRALDEZ, A. J., SCHROTH, G. P., ANDERSSON, O. & STAINIER, D. Y. 2016. Cloche is a bHLH-PAS transcription factor that drives haemato-vascular specification. *Nature,* 535**,** 294-8.
RIDDELL, J., GAZIT, R., GARRISON, B. S., GUO, G., SAADATPOUR, A., MANDAL, P. K., EBINA, W., VOLCHKOV, P., YUAN, G. C., ORKIN, S. H. & ROSSI, D. J. 2014. Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell,* 157**,** 549-64.
ROBB, L., LYONS, I., LI, R., HARTLEY, L., KONTGEN, F., HARVEY, R. P., METCALF, D. & BEGLEY, C. G. 1995. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc Natl Acad Sci U S A,* 92**,** 7075-9.
SALEQUE, S., CAMERON, S. & ORKIN, S. H. 2002. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes and Development,* 16**,** 301-306.
SANDER, J. D. & JOUNG, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol,* 32**,** 347-55.
SEITA, J. & WEISSMAN, I. L. 2010. Hematopoietic Stem Cell: Self-renewal versus Differentiation. *Wiley interdisciplinary reviews. Systems biology and medicine,* 2**,** 640-653.
SHALABY, F., ROSSANT, J., YAMAGUCHI, T. P., GERTSENSTEIN, M., WU, X. F., BREITMAN, M. L. & SCHUH, A. C. 1995. Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature,* 376**,** 62-6.
SILVER, L. & PALIS, J. 1997. Initiation of Murine Embryonic Erythropoiesis: A Spatial Analysis. *Blood,* 89**,** 1154-1164.
SOOD, R., ENGLISH, M. A., JONES, M., MULLIKIN, J., WANG, D.-M., ANDERSON, M., WU, D., CHANDRASEKHARAPPA, S. C., YU, J., ZHANG, J. & PAUL LIU, P. 2006. Methods for reverse genetic screening in zebrafish by resequencing and TILLING. *Methods,* 39**,** 220-227.
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.
SRIVASTAVA, M., HSIEH, S., GRINBERG, A., WILLIAMS-SIMONS, L., HUANG, S. P. & PFEIFER, K. 2000. H19 and Igf2 monoallelic expression is regulated in two distinct ways by a shared cis acting regulatory region upstream of H19. *Genes Dev,* 14**,** 1186-95.
STAINIER, D. Y., WEINSTEIN, B. M., DETRICH, H. W., 3RD, ZON, L. I. & FISHMAN, M. C. 1995. Cloche, an early acting zebrafish gene, is required by both the endothelial and hematopoietic lineages. *Development,* 121**,** 3141-50.
TAKAHASHI, K., YAMAMURA, F. & NAITO, M. 1989. Differentiation, maturation, and proliferation of macrophages in the mouse yolk sac: a light-microscopic, enzyme-cytochemical, immunohistochemical, and ultrastructural study. *J Leukoc Biol,* 45**,** 87-96.
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. *Dev Biol,* 417**,** 25-39.
TOBER, J., KONISKI, A., MCGRATH, K. E., VEMISHETTI, R., EMERSON, R., DE MESY-BENTLEY, K. K., WAUGH, R. & PALIS, J. 2007. The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood,* 109**,** 1433-41.
TONG, B., GRIMES, H. L., YANG, T. Y., BEAR, S. E., QIN, Z., DU, K., EL-DEIRY, W. S. & TSICHLIS, P. N. 1998. The Gfi-1B proto-oncoprotein represses p21WAF1 and inhibits myeloid cell differentiation. *Mol Cell Biol,* 18**,** 2462-73.
VOGELI, K. M., JIN, S. W., MARTIN, G. R. & STAINIER, D. Y. 2006. A common progenitor for haematopoietic and endothelial lineages in the zebrafish gastrula. *Nature,* 443**,** 337-9.
WAHLESTEDT, C. 2013. Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nature Reviews Drug Discovery,* 12**,** 433.
WALLIS, D., HAMBLEN, M., ZHOU, Y., VENKEN, K. J. T., SCHUMACHER, A., LEIGHTON GRIMES, H., ZOGHBI, H. Y., ORKIN, S. H. & BELLEN, H. J. 2003. The zinc finger transcription factor Gfi1, implicated in lymphomagenesis, is require for inner ear hair cell differentiation and survival. *Development,* 130**,** 221-232.
WANG, H., LI, W., GUO, R., SUN, J., CUI, J., WANG, G., HOFFMAN, A. R. & HU, J. F. 2014. An intragenic long noncoding RNA interacts epigenetically with the RUNX1 promoter and enhancer chromatin DNA in hematopoietic malignancies. *Int J Cancer,* 135**,** 2783-94.
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. *Proc Natl Acad Sci U S A,* 93**,** 3444-9.
WARD, M., MCEWAN, C., MILLS, J. D. & JANITZ, M. 2015. Conservation and tissue-specific transcription patterns of long noncoding RNAs. *J Hum Transcr,* 1**,** 2-9.
WEI, W., WEN, L., HUANG, P., ZHANG, Z., CHEN, Y., XIAO, A., HUANG, H., ZHU, Z., ZHANG, B. & LIN, S. 2008. Gfi1.1 regulates hematopoietic lineage differentiation during zebrafish embryogenesis. *Cell Research,* 18**,** 677-685.
WESTERFIELD, M. 1993. *The zebrafish book; a guide for the laboratory use of zebrafish (brachydanio rerio),* Eugene, University of Oregon Press.
WESTERFIELD, M. 1995. *The zebrafish book. A guide for the laboratory use of zebrafish (danio rerio),* Eugene, University of Oregon Press.
WIEDENHEFT, B., STERNBERG, S. H. & DOUDNA, J. A. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature,* 482**,** 331-8.
WILUSZ, J. E., SUNWOO, H. & SPECTOR, D. L. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes & development,* 23**,** 1494-1504.
YAMANE, T., HOSEN, N., YAMAZAKI, H. & WEISSMAN, I. L. 2009. Expression of AA4.1 marks lymphohematopoietic progenitors in early mouse development. *Proc Natl Acad Sci U S A,* 106**,** 8953-8.
YAMANE, T., WASHINO, A. & YAMAZAKI, H. 2013. Common developmental pathway for primitive erythrocytes and multipotent hematopoietic progenitors in early mouse development. *Stem Cell Reports,* 1**,** 590-603.
YOKOMIZO, T., HASEGAWA, K., ISHITOBI, H., OSATO, M., EMA, M., ITO, Y., YAMAMOTO, M. & TAKAHASHI, S. 2008. Runx1 is involved in primitive erythropoiesis in the mouse. *Blood,* 111**,** 4075-4080.
YOKOMIZO, T., OGAWA, M., OSATO, M., KANNO, T., YOSHIDA, H., FUJIMOTO, T., FRASER, S., NISHIKAWA, S., OKADA, H., SATAKE, M., NODA, T., NISHIKAWA, S. & ITO, Y. 2001. Requirement of Runx1/AML1/PEBP2alphaB

for the generation of haematopoietic cells from endothelial cells. *Genes Cells,* 6**,** 13-23.
ZOVEIN, A. C., TURLO, K. A., PONEC, R. M., LYNCH, M. R., CHEN, K. C., HOFMANN, J. J., COX, T. C., GASSON, J. C. & IRUELA-ARISPE, M. L. 2010. Vascular remodeling of the vitelline artery initiates extravascular emergence of hematopoietic clusters. *Blood,* 116**,** 3435-3444.
ZWEIDLER-MCKAY, P. A., GRIMES, H. L., FLUBACHER, M. M. & TSICHLIS, P. N. 1996. Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor. *Mol Cell Biol,* 16**,** 4024-34.