

UNIVERSITY OF OXFORD

Biotagging, a genetically encoded toolkit in the zebrafish, reveals novel non-coding RNA players during neural crest and myocardium development

Author:
Vanessa Mae Ren CHONG

Supervisor:
Professor Tatjana SAUKA-SPENGLER

A thesis submitted for the degree of
D.Phil. in Medical Sciences

at

Lincoln College

Trinity 2017

University of Oxford

Abstract

Lincoln College

Trinity 2017

D.Phil. in Medical Sciences

Biotagging, a genetically encoded toolkit in the zebrafish, reveals novel non-coding RNA players during neural crest and myocardium development

by Vanessa Mae Ren CHONG

Complex multicellular organisms are composed of at least 200 cell types, which contain the same DNA "black box" of genetic information. It is the precise regime according to which they express their genes, exquisitely controlled by gene regulatory circuits, that defines their cellular identity, morphology and function. We have developed an *in vivo* biotinylation method that uses genetically encoded components in zebrafish, termed *biotagging*, for genome-wide regulatory analysis of defined embryonic cell populations. By labelling selected proteins in specific cell types, *biotagging* eliminates background inherent to analyses of complex embryonic environments via highly stringent biochemical procedures and targeting of specific interactions without the need for cell sorting. We utilised *biotagging* to characterise the *in vivo* translational landscape on polysomes as well as the transcriptional regulatory landscape in nuclei of migratory neural crest cells, which intermix with environing tissues during their migration. Our migratory neural crest translatome presented both known and novel players of the neural crest gene regulatory network. An in depth look into the active nuclear transcriptome uncovered a complex world of non-coding regulatory RNAs that potentially specify migratory neural crest identity and present evidence of active bidirectional transcription on regions of open chromatin that include putative *cis*-regulatory elements. Analysis of our transcribed *cis*-regulatory modules functionally links these elements to known genes that are key to migratory neural crest function and its derivatives. We also identified a novel cohort of circular RNAs enriched at regions of tandem duplicated genes. Last but not least, we recovered developmentally regulated long non-coding RNAs and transcribed transposable elements. To functionally dissect the biological roles of these factors, we have built two *Ac/Ds*-mediated *in vivo* toolkits for efficient screening of putative enhancers and for CRISPR/Cas9-based transcriptional modulation. Overall, our methods and findings present a comprehensive view of the active coding and non-coding landscapes of migratory neural crest on a genome-wide scale that refine the current regulatory architecture underlying neural crest identity.

Acknowledgements

First and foremost, my heartfelt gratitude goes to Professor Tatjana Sauka-Spengler for her unwavering support throughout this *D.Phil*. as both a mentor and supervisor. The body of work represented in this thesis would not have been possible without her deep belief in my potential and constant encouragement to push my scientific limits.

I would also like to thank the rest of the Sauka-Spengler lab, both current and former members, for being a superb group of comrades in the ever delicate task of navigating a work/life balance while trying to do great science.

To my Thesis Advisory Committee members - Professor Tudor Fulga, Dr. Andrew Bassett and Dr. Sarah de Val - thank you for the enthusiastic support and meaningful discussions that have shaped this thesis in both small and big ways.

To all past and current collaborators - thank you for your faith and for presenting me with multiple opportunities to think creatively beyond the scope of my own research.

A big thank you goes to the funding bodies that made it financially-possible for me to be here at Oxford - the Clarendon Fund, Keith Murray Trust (Lincoln College) and the Radcliffe Department of Medicine.

Last but not least, I would like to thank my parents back in Malaysia, for the constant trust in my judgement and decision-making as I navigate through life (including deciding to pursue this *D.Phil*.). Special thanks goes to my supportive partner-in-crime, Alasdair, and my constant writing companion without opposable thumbs, Crispr.

Contents

Abstract **iii**

Acknowledgements **v**

List of Figures **xiii**

List of Tables **xv**

List of Abbreviations **xvii**

1 General introduction **1**

1.1 The neural crest drives vertebrate development 1

1.1.1 An overview of early neural crest ontogeny 3

Induction and specification 3

EMT and migration 4

1.1.2 Gene regulatory networks underlying neural crest ontogeny . 5

1.1.3 Enhancers are non-coding players in the neural crest GRN . . 8

1.2 The non-coding genome in development and disease 10

1.2.1 Enhancer-derived RNAs (eRNAs) 12

ERNAs respond to biological stimuli and may affect chromatin looping 12

ERNAs as regulators of chromatin remodelling 13

ERNAs as molecular tethers that regulate transcription of genes 14

Global eRNA transcription has specific spatiotemporal patterns 14

1.2.2 RNA Polymerase II pausing at promoters 15

Differences between eRNAs and promoter RNAs 15
Potential function of RNAPII pausing at promoters 16
1.2.3 Long non-coding RNAs (lncRNAs) 16
LncRNAs regulate biological processes 17
Genome-wide identification and characterisation of lncRNAs 19
Debates surrounding lncRNA form and function 21
lncRNAs and metazoan evolution 23
1.2.4 eRNAs and lncRNAs - different, same, or both? 24
1.2.5 Circular RNAs 25
Biogenesis of circRNAs 26
1.2.6 Transposable elements 27
1.3 Current methods and limitations in studying non-coding RNA 28
1.4 The non-coding genome in embryonic development 31

2 The *biotagging* toolkit for zebrafish **33**
2.1 An *in vivo* biotinylation toolkit for zebrafish 33
2.1.1 Principles and history of *in vivo* biotinylation 33
2.1.2 The binary *in vivo* biotinylation system for zebrafish 34
2.2 Establishment of *Avi effector* transgenic lines 35
2.2.1 *Avi effectors* for subcellular compartmentalisation 36
2.2.2 *Avi effector* chromatin regulators for epigenomic profiling . . . 39
2.3 Establishment of *BirA driver* transgenic lines 42
2.3.1 Ubiquitous and tissue-specific *BirA drivers* 42
2.4 Establishment and refinement of the *biotagging* method 44
2.4.1 Specificity and efficiency of *biotagging* 45
2.4.2 Experimental conditions for successful nuclear or polysomal isolation 48
2.5 Genome-wide profiling of the cardiomyocyte nuclear transcriptome . 49
2.5.1 Enrichment of myocardial molecular signature by *biotagging* . 49

2.6 Discussion . . . 52
2.6.1 The finer details of *biotagging* . . . 52
2.6.2 Biotin-ChIP of chromatin regulators in defined cell types . . 54

3 The neural crest translatome and nuclear transcriptome 57
3.1 Current approaches to studying translatomes . . . 57
3.1.1 Ribosome profiling . . . 57
3.1.2 Polysome profiling . . . 58
3.2 *biotagging*-TRAP-seq of the migrating neural crest . . . 58
3.2.1 Exploratory data analysis of the neural crest translatome . . . 61
3.2.2 Differential expression analysis of the neural crest translatome 65
Genes with higher abundance in the *sox10* translatome . . . 66
Genes with lower abundance in the *sox10* translatome . . . 67
3.3 *biotagging*-RanGAP-seq of the migratory neural crest . . . 70
3.3.1 Bidirectional transcription across open promoters is enriched in the neural crest . . . 70
3.3.2 Neural crest molecular signature is encoded in the *cis*-regulatory transcriptome . . . 74
3.4 *In vivo* tools for studying gene regulatory networks . . . 79
3.4.1 *Ac/Ds*-mediated transposition of genetic elements . . . 80
The *Ac/Ds* toolkit for assaying enhancer activity in the embryo 80
3.4.2 Transcribed *cis*-regulatory elements have biological activity . 81
3.5 Discussion . . . 83
3.5.1 Current insights from the *sox10* translatome at 16-18 ss . . . 83
3.5.2 Current insights from the *sox10* nuclear transriptome at 16-18 ss . . . 86
3.5.3 Compartmentalising the identity of migratory neural crest . . 87

4 Identification and characterisation of zebrafish non-coding RNAs 89
4.1 Non-coding RNAs are new game players of development . . . 89

4.2 Circular RNAs present a new paradigm in the transcription of genes 90
4.2.1 Putative zebrafish circular RNAs form genomic clusters and are compartment-specific 90
4.2.2 Predicted circularisation events are weakly correlated to underlying mRNA expression 93
4.2.3 Circularisation "hotspots" reveal enrichment at tandem duplicated genes 94
4.2.4 A new method for multiplexed screening and sequencing of putative circRNAs 97
Two novel tandem-circRNAs: *circ-cyt1* and *circ-smyhc* 98
4.3 Identification of developmentally regulated non-coding RNAs 103
4.3.1 Long non-coding RNAs (lncRNAs) 103
4.3.2 Transposable elements 106
Testing enhancer activity of transposable elements in zebrafish 106
4.4 In vivo tools for studying gene regulatory networks II 109
4.4.1 CRISPR/Cas9-based *interference* or *activation* of gene expression 110
A genetic CRISPR/Cas9-based system for epigenome engineering 111
4.5 Discussion 113

5 Concluding remarks and future work 117

6 Materials and methods 121

6.1 Zebrafish husbandry 121
6.2 Plasmids 121
6.3 Microinjection of zebrafish embryos 121
6.3.1 Generation of *tol2* plasmid transgenic lines 121
6.3.2 Generation of BAC transgenic lines 122
6.3.3 *Ac/Ds*-mediated enhancer screening assay 122

6.4 *Biotagging*: Isolation of subcellular compartments using Streptavidin beads . . . 123
6.4.1 Preparation of Streptavidin beads for RNA procedures . . . 123
6.4.2 Nuclei isolation . . . 123
6.4.3 Polysomal isolation . . . 125
6.4.4 RNA extraction of isolated nuclei and polysomes . . . 126
6.4.5 ATAC-seq from FAC-sorted cells . . . 126
6.4.6 Protein extraction and Western Blot . . . 127
6.5 RNA-seq . . . 128
6.5.1 Preparation of strand-specific libraries . . . 128
6.5.2 Bioinformatics processing . . . 128
Validation of transgene alleles using *de novo* transcript assembly 129
6.6 Circular RNA validation . . . 131
6.6.1 Simultaneous screening and amplification of circRNA candidates . . . 131
6.6.2 Bioinformatics processing . . . 133
6.7 Whole-mount Immunohistochemistry . . . 134

A Gene Lists **135**
A.1 Chapter 3 . . . 137
A.1.1 Genes more abundant in the *sox10* translatome . . . 138
A.1.2 Genes less abundant in the *sox10* translatome . . . 143
A.2 Chapter 4 . . . 153
A.2.1 Differentially expressed lncRNAs in *sox10 nuclear* versus *bactin2 nuclear* at 16-18 ss . . . 154
A.2.2 Differentially expressed lncRNAs in *myl7 nuclear* versus *bactin2 nuclear* at 24-26 hpf . . . 161

B circRNA sequences **165**
B.1 Chapter 4 . . . 167

B.1.1 Transcript c119_g1_i1 containing *circ-cyt1* 167
B.1.2 Transcript c408_g1_i1 containing *circ-smyhc* 168

List of Figures

1.1 Neural crest development 2
1.2 Neural crest gene regulatory network (GRN) 7
1.3 Enhancers regulate gene expression 9
1.4 Long non-coding RNAs (lncRNAs) regulate gene expression 19
1.5 Formation of circular RNAs (circRNAs) 27

2.1 The *biotagging* toolkit 35
2.2 *Avi effectors* for subcellular compartmentalisation 38
2.3 *Avi effector* chromatin regulators for epigenomic profiling 41
2.4 *BirA drivers* for biotinylation of *Avi effectors* 44
2.5 Specificity and efficiency of the *biotagging* method 47
2.6 Enrichment of myocardial molecular signature by *biotagging* 51

3.1 *Sox10* is a neural crest marker 60
3.2 Exploratory data analysis of the neural crest translatome 63
3.3 Differential expression analysis of the neural crest translatome 69
3.4 Bidirectional transcription across open promoters is enriched in the neural crest 73
3.5 Neural crest molecular signature is encoded in the *cis*-regulatory transcriptome 76
3.6 The *Ac/Ds* toolkit for assaying enhancer activity in the embryo 82

4.1 Putative zebrafish circular RNAs form genomic clusters and are compartment-specific 92

4.2 Correlation between expression of putative circRNAs and underlying mRNA expression 93
4.3 Validating circularisation "hotspots" enriched at tandem duplicated genes 96
4.4 Characterisation of *circ-cyt1* 100
4.5 Characterisation of *circ-smyhc* 102
4.6 Identification of developmentally regulated non-coding RNAs 105
4.7 Enhancer activity of *cis*-regulatory transposable elements in zebrafish 108
4.8 Tissue-specific Cas9 transgenic lines for epigenome engineering . . . 112

List of Tables

2.1 *Biotagging* transgenic lines 36

3.1 Chromosome positions and primer sequences of new enhancers identified from *biotagging* 81

4.1 Identification of putative circRNAs in *biotagging* RNA-seq datasets . 90

4.2 Chromosome positions and primer sequences of *tfap2α* enhancers containing transposable elements 107

A.1 Genes more abundant in the *sox10* translatome 142

A.2 Genes less abundant in the *sox10* translatome 152

A.3 Differentially expressed lncRNAs in *sox10 nuclear* 160

A.4 Differentially expressed lncRNAs in *myl7 nuclear* 164

List of Abbreviations

aa	amino acid
AFC	Additive Fold Change
ASO	antisense oligonucleotide
ATAC	Assay for Transposase Accessible Chromatin
BCCP	biotin carboxyl carrier protein
BirA	biotin holoenzyme synthetase, biotin ligase
BLAT	BLAST-like alignment tool
bp	base pair
CBP	CREB (cAMP response element-binding protein)-binding protein
cDNA	complementary DNA
CHART	Capture Hybridisation Analysis of RNA Targets
ChIP	Chromatin Immunoprecipitation
ChIRP	Chromatin Isolation by RNA Purification
circRNA	circular RNA
CNC	cranial neural crest
CNS	central nervous system
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRM	*cis*-regulatory module
DGD	Duplicated Genes Database
DNA	deoxyribonucleic acid
EMT	epithelial-to-mesenchymal transition
ERV	endogenous retrovirus
eRNA	enhancer RNA

(m)ESCs	(mouse) embryonic stem cells
EZH2	Enhancer of zeste homolog 2
FACS	Fluorescence Activated Cell Sorting
FC	fold change
FDR	false discovery rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GO	Gene ontology
GREAT	Genomic Regions Enrichment of Annotations Tool
GSEA	Gene Set Enrichment Analysis
GWAS	genome-wide association study
HDR	homology-directed repair
HOTAIR	HOX transcript antisense RNA
HOTTIP	HOXA transcript at the distal tip
hpf	hours post fertilisation
IHC	immunohistochemistry
INTACT	Isolation of Nuclei TAgged in specific Cell Types
kb	kilobase
lincRNA	long intergenic/intervening non-coding RNA
lncRNA	long non-coding RNA
miRNA	micro RNA
MLL	mixed-lineage leukaemia
mRNA	messenger RNA
ncRNA	non-coding RNA
NC	neural crest
NELF	negative elongation factor
NGS	next generation sequencing
NLS	Nuclear Localisation Signal
NPC	nuclear pore complex

NSD3	nuclear receptor binding SET domain protein 3
ORF	open reading frame
PANTHER	Protein ANalysis THrough Evolutionary Relationships
PBAF	polybromo-associated BAF (BRG1-associated factors)
PDGF-A	platelet-derived growth factor subunit A
piRNA	Piwi-interacting RNA
PRC	Polysome Repressive Complex
qPCR	quantitative polymerase chain reaction
RanGAP1	Ran GTPase-activating protein 1
RAP	RNA Antisense Purification
RNA	ribonucleic acid
RNAPII	RNA Polymerase II
rRNA	ribosomal RNA
RT-PCR	reverse transription (followed by) polymerase chain reaction
SINE	Short Interspersed Nuclear Element
ss	somite stage
SUMO-1	small ubiquitin-related modifier 1
TALEN	transcription activator-like effector nucleases
TE	transposable element
TF	transcription factor
TLR4	Toll-like receptor 4
TPM	Transcript Per Million
TRAP	Translating Ribosome Affinity Purification
TSS	transcription start site
UTR	untranslated region
WDR5	WD repeat-containing protein 5
WISH	whole-mount *in situ* hybridisation
WPP	tryptophan-proline-proline (domain found in RanGAP)

ZFIN The Zebrafish Information Network

Chapter 1

General introduction

1.1 The neural crest drives vertebrate development

The neural crest (NC) is a transient and multipotent vertebrate cell population that arises following gastrulation in the developing embryo. At the end of gastrulation, prospective neural crest occupy a strip of cells termed neural plate border, located between the future neural and non-neural ectoderm. As neurulation proceeds, NC cells come to reside within the dorsal aspect of the forming neural tube, where they undergo epithelial-to mesenchymal transition (EMT) in order to segregate from the neural tube, delaminate and migrate across the embryo to sometimes very distant final destinations (Sauka-Spengler and Bronner-Fraser, 2008).

Pioneering research in the neural crest using cell labelling techniques and quail-chick chimaera transplantation experiments formed the cornerstones of NC ontogeny during embryogenesis, spearheaded by descriptions of their migratory pathways and derivatives (Le Douarin and Kalcheim, 1999). The NC has captured the interests of both developmental and evolutionary biologists alike, with its highly conserved ability to migrate extensively throughout the embryo along well-defined canonical pathways to specific regions within the embryo where they settle and differentiate into a multitude of cell types. Non-ectomesenchymal derivatives of trunk and cranial neural crest (CNC) include sensory and autonomic neurons, glia, melanocytes, iridophores, xanthophores, erythrophores and cells of the adrenal medulla.

In the head, the CNC also differentiates into ectomesenchymal derivatives such as connective tissue, odontoblasts, osteocytes, chrondrocytes and smooth muscle

cells (Graham, 2003; Donoghue et al., 2008). The CNC is also known to give rise to the corneal endothelium and stroma in the eye and pigment cells in the inner ear (Dupin and Sommer, 2012), as well as pericytes and the meninges in the forebrain (Etchevers et al., 2001). Due to their heavy contribution to the patterning of vertebrate craniofacial structures (reviewed in (Santagati and Rijli, 2003)), the CNC is of particular interest as a key contributor to the evolution of jawed vertebrates (Cerny et al., 2010). This is supported by evidence in the lamprey (a basal vertebrate) highlighting functional interactions between main components of the gene regulatory network underlying NC ontogeny (see below) (Sauka-Spengler et al., 2007; Nikitina et al., 2008).

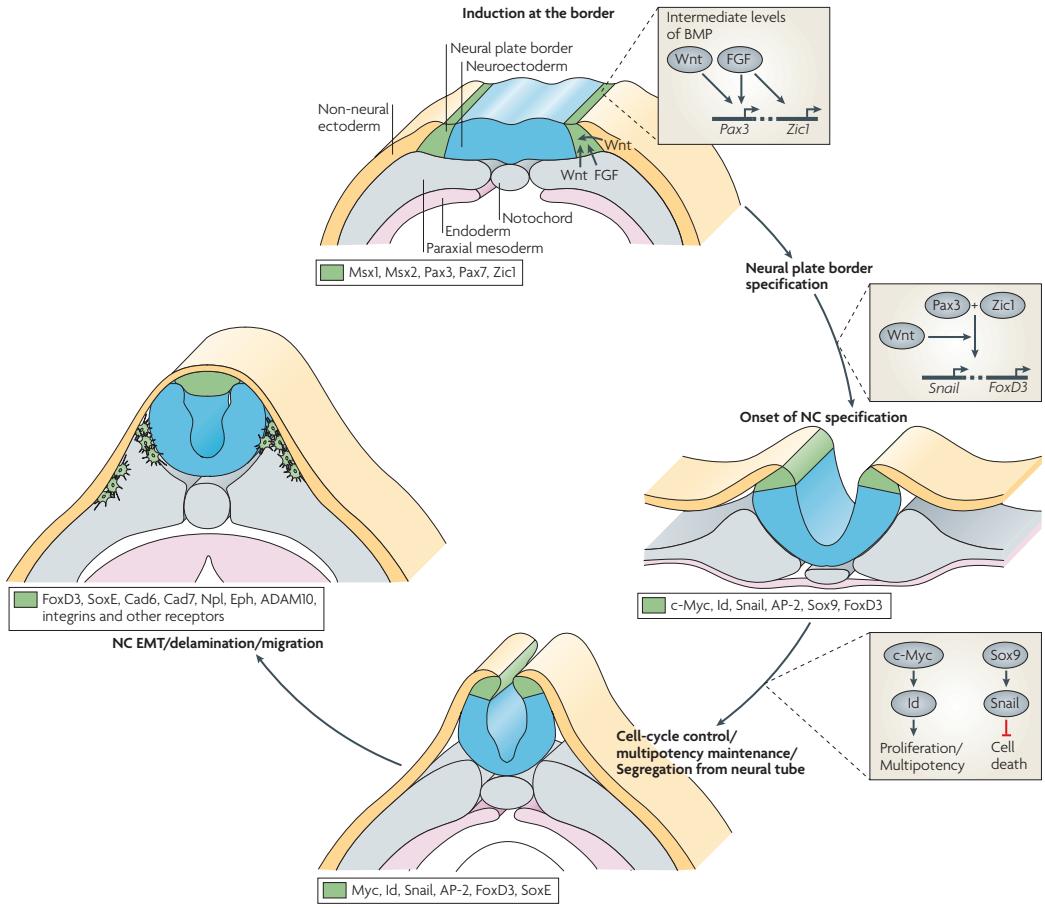


FIGURE 1.1: Neural crest development in vertebrate embryos (Sauka-Spengler and Bronner-Fraser, 2008).

Schematic representation of neural crest induction, specification and delamination from the dorsal neural tube. Key genes involved in the different stages are also shown.

1.1.1 An overview of early neural crest ontogeny

Induction and specification

Neural crest induction at the neural plate border, mainly controlled by BMPs and WNTs, precedes its specification and subsequent migration (reviewed in (Knecht and Bronner-Fraser, 2002; Prasad et al., 2012; Simões-Costa and Bronner, 2015)). In the zebrafish, *foxd3* has been demonstrated to be a key player in specification of premigratory NC. *Foxd3* mutants demonstrated reduced expression of early NC markers such as *snai1b*, *sox10* and *crestin*, and exhibited developmental phenotypes linked to impaired neural crest migration and differentation (Stewart et al., 2006). Using novel *foxd3* gene trap zebrafish lines that express endogenous *foxd3-Citrine* or *foxd3-mCherry* fusion proteins, *in vivo* movements of *foxd3*-expressing cells were live-tracked at developmental stages key to NC ontogeny (Hochgreb-Hägele and Bronner, 2013). This presented a leap in the field, as a previous *foxd3* reporter line utilising its proximal promoter led to delayed onset of its reporter expression and thus a missed opportunity to observe NC specification *in vivo* (Gilmour et al., 2002). In the gene trap lines presented by Hochgreb-Hägele *et al.*, *foxd3-Citrine* expression was reported to start at 1-2 ss as bilateral patches at the neural plate border. However, closer examination of this line by us revealed that *foxd3* expression is also detected earlier at 70% epiboly stage on the dorsal side (Gavriouchkina et al., 2017), recapitulating a previous observation using *in situ hybridisation* (Wang et al., 2011c). These observations suggests uncharacterised roles of *foxd3* during NC induction, upstream of its characterised role in NC specification. *Tfap2*α is also a key specifier of the NC. At 70% epiboly stage, *tfap2*α expression is complementary to *foxd3*, covering most of the ventral regions (Knight et al., 2003; Wang et al., 2011c). NC specification and migration were negatively affected in *Tfap2*α mutants, although *foxd3* expression was only affected in hindbrain NC and *snai2* was largely unaffected. Most importantly, double mutants of *tfap2*α*/foxd3* completely lack all NC derivatives (Wang et al., 2011c). These results suggested distinct modes of action

for *tfap2*α and *foxd3* in specifying different subpopulations during the birth of the NC. Other than transcription factors (TFs), chromatin regulators also play a key role in NC specification, highlighted by a study demonstrating the role of a lysine methyltransferase, NSD3, in regulating gene expression of NC transcription factors (Jacques-Fricke and Gammill, 2014). However, it is important to note that this "neural plate induction of NC" model has come into question recently, with evidence presented by Buitrago-Delgado, Nordin *et al.* that *Xenopus* NC and pluripotent blastula cells expressed the same cohort of genes including *Id3*, *Tfap2*, *Ets1*, *FoxD3* and *Snail1* (Buitrago-Delgado et al., 2015). These factors were demonstrated to be required for maintenance of pluripotency in blastula cells, suggesting that the NC represents a subset of these stem cells that remained pluripotent as neurulation proceeds until they become specified as the latter.

EMT and migration

Following specification, NC cells delaminate from the dorsal neural tube before migrating into the embryo. EMT is widely-accepted to orchestrate this process by triggering genetic pathways that include TFs such as *Twist*, *SoxE*, *Fox*, *Snail/Slug* and *Ets* family of proteins - all known to be implicated in both NC development and the onset of cancer metastasis. Furthermore, an extensive body of the evidence in the literature has led to a detailed description of the migratory pathways following delamination (both reviewed in (Theveneau and Mayor, 2012)) . During migration, NC cells integrate positive and negative signals from their surroundings via cell-cell/extracellular matrix interactions to migrate in distinctive ventromedial or dorsolateral pathways in the trunk, as well as dorsoventral streams to form cranial ganglia or to invade branchial arches and form the craniofacial skeleton in the head. Post-otic NC i.e. cardiac and enteric NC migrate to their aptly-named final destinations, where they form non-ectomesencyhmal derived structures such as connective tissue in blood vessel walls, neurons and ganglia.

Recent work has provided new insights into the genetic programme underlying NC EMT and migration. In chick, Cadherin6B is processed by metalloproteinase proteins to generate C-terminal fragments (CTF2s) which upregulate expression of a pro-EMT gene, *Snail2* (Schiffmacher et al., 2016). Delamination of NC cells requires inhibition of Wnt/β-catenin signalling upstream of EMT, a process recently shown in frog and chick embryos to be mediated by Dact1/2 scaffold proteins expressed in premigratory NC cells but were found dispensable once onset of migration was complete (Rabadán et al., 2016). Recently, another metalloproteinase, MMP16, expressed in chick CNC was demonstrated to coordinate NC migration by enhancing degradation of N-cadherin and laminin in the dorsal neural tube (Roth et al., 2017). Downregulation of N-cadherin in the dorsal neural tube is coupled with upregulation of N-cadherin in the NC during EMT. Moreover, the latter was demonstrated to be controlled by a platelet-derived growth factor (PDGF-A) and its receptor (PDGFRα), both expressed in migrating CNCs (Bahm et al., 2017). With the continuing advancement of molecular and genetic techniques, we predict the continuing emergence of fresh evidence highlighting the multi-faceted details that underlie this complex behaviour *in vivo*.

1.1.2 Gene regulatory networks underlying neural crest ontogeny

Gene regulatory networks (GRNs) co-ordinate and control the expression of genes encoding TFs and cell signalling pathway components, set within defined programmes of genetic cascades mediated by *cis*-regulatory elements during development (reviewed in (Levine and Davidson, 2005)). GRNs present a different standpoint in the understanding of developmental pathways from a genomic instead of inter-/intracellular process perspective. The lengthy process of NC development, that starts at the end of gastrulation and proceeds in late organogenesis has been proposed to be orchestrated by a multi-module gene regulatory network (Sauka-Spengler and Bronner-Fraser, 2008; Simões-Costa and Bronner, 2015). WNT, FGF and BMP

signals at the neural plate border activate expression of genes from the *Msx*, *Pax* and *Zic* families during NC induction. *Pax3* and *Zic1* activates expression of *bona fide* NC factors, such as *Snai1* and *FoxD3*, thus driving the onset of NC specification defined by the expression of *Tfap2*, *Id*, *Myc*, *Myb*, *SoxE* and *Ets* gene families. The continued expression of these factors, as well as the activation of cadherins, integrins, receptors and metalloproteinases, then lead to EMT and delamination of NC cells from the dorsal neural tube.

The dawn of the -omics era has contributed significantly to the study of GRNs. Coupled with an increasing catalogue of sequenced genomes, genome-wide approaches heralded an explosion of exploratory studies that not only confirmed previous knowledge but also increased the rate of identification of previously unknown developmental players. In the avian NC, a transcriptome study by Simões-Costa *et al.* (Simões-Costa et al., 2014) led to the identification of multiple genes not known previously to be expressed in the migratory CNC such as *Lmo4*, *RxrG*, *Ltk* and *Col9a3*. This study also highlighted a potential role for post-translational regulation during CNC development, as a significant number of RNA processing molecules were found to be enriched in this cell population. Using RNA-seq to compare the transcriptome of cranial and trunk NC populations followed by functional assays led to the identification of a cranial-specific module consisting of - *Brn3c*, *Lhx5*, *Dmbx1* in the neural plate border; *Sox8*, *Tfap2*β in premigratory NC; and *Ets1* (activated by *Tfap2*β) in the migratory NC (Simoes-Costa and Bronner, 2016). Overexpression of these factors in the trunk NC resulted in reprogramming of their identity, highlighted by ectopic activity of a cranial neural crest enhancer *Sox10E2* and increased expression of chondrocyte-related genes *Runx2* and *Alx1*.

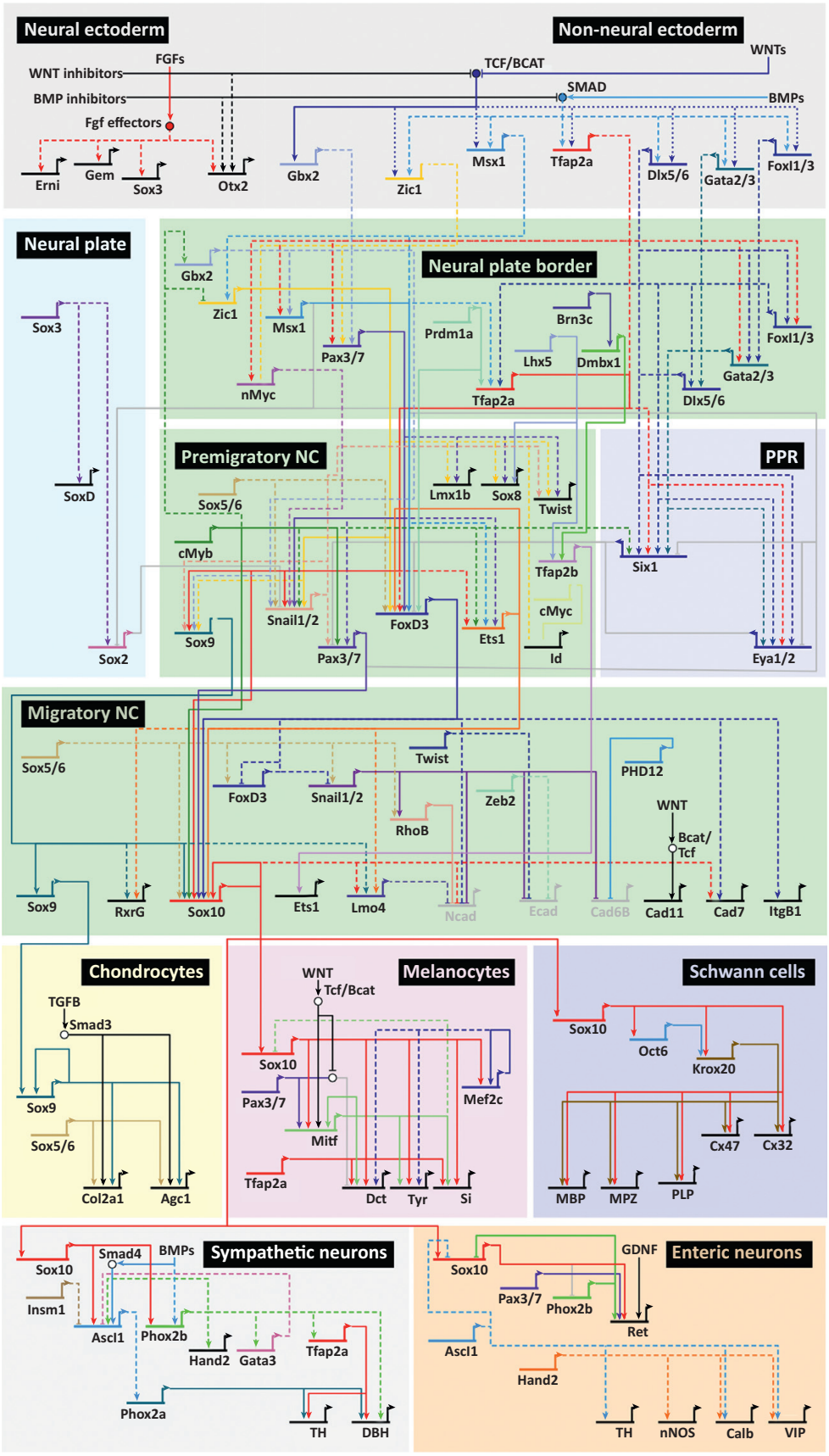


FIGURE 1.2: The neural crest gene regulatory network (GRN) in its most current form (Martik and Bronner, 2017).

1.1.3 Enhancers are non-coding players in the neural crest GRN

As mentioned at the start of the previous section, *cis*-regulatory elements serve as important "switches" within GRN modules by integrating inputs from upstream factors in order to coordinate output of downstream targets. SOX10E1 and SOX10E2 enhancers, situated 1 kb downstream of the coding region for the *Sox10* gene, a neural crest master regulator (reviewed in (Kelsh, 2006; Sauka-Spengler and Bronner-Fraser, 2008)) have been shown to control the expression of this gene in the chicken embryo (Betancur et al., 2010). Both enhancers demonstrated distinct spatiotemporal activity, where SOX10E1 was active in the migrating vagal and trunk neural crest but not in the CNC, unlike SOX10E2 which was active in early delaminating CNC cells reminiscent of endogenous *Sox10* expression. Mutations at key binding motifs identified in SOX10E2 followed by TF knockdown and ChIP experiments confirmed *Sox9*, *Ets1* and *cMyb* proteins as transcriptional inputs for *Sox10* gene expression. Similarly, two enhancers, NC1 and NC2, residing 20 kb and 44 kb upstream of *FoxD3* gene have been shown to control the expression of this NC specifier in the avian embryo (Simões-Costa et al., 2012). NC1 was found to be active in the premigratory CNC, but its activity diminished during migration and no activity could be detected caudal to rhombomere 3. On the other hand, NC2 is active in premigratory, delaminating and migratory NC at the vagal and trunk axial levels. While NC1 and NC2 display distinct and mutually exclusive spatiotemporal activities, in both cases these activities recapitulated endogenous *FoxD3* expression. Knockdown of upstream factors such as *Pax7*, *Msx1*, *Ets1* and *Zic1* resulted in loss of either one or both enhancers, thus confirming their participation in the NC *FoxD3* module underlying differential regulation between trunk and CNC.

From a genome-wide perspective, large-scale epigenomic mapping using p300, H3K27Ac, H3K4me1 and H3K4me3 enrichment profiles have facilitated the identification of enhancer elements in human NC cells (Rada-Iglesias et al., 2012). To

demonstrate the potential of their approach as a resource for NC biology, Rada-Iglesias *et al.* exploited their dataset for sequence analysis to identify major TF-binding events, therefore uncovering the association between a key neural crest specifier, *TFAP2A*, and permissive chromatin landscape at functionally annotated enhancers. From an evolutionary point-of-view, neural crest enhancers present yet another cohort of subjects that can be moulded under evolutionary pressures within the neural crest GRN to contribute to species divergence of craniofacial structures (Prescott et al., 2015).

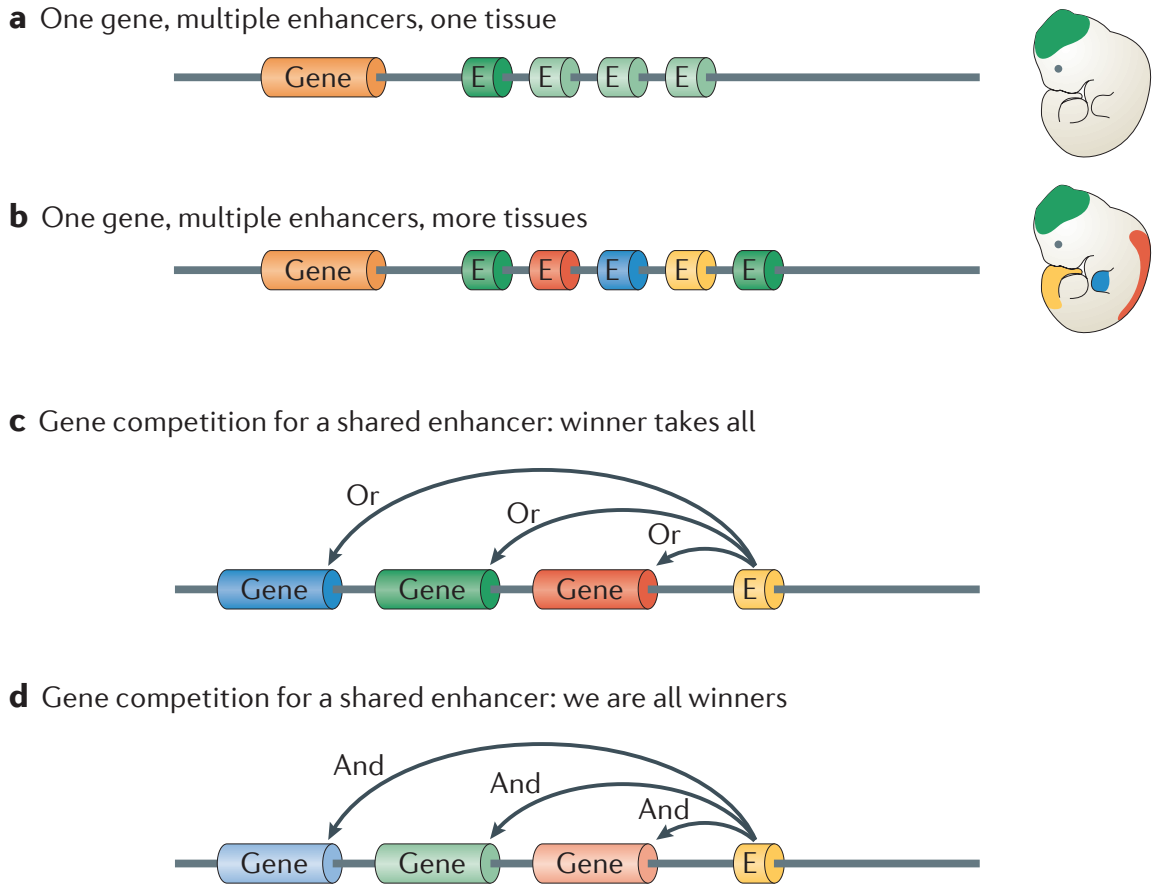


FIGURE 1.3: Enhancers are *cis*-regulatory elements that act cooperatively or promiscuously to regulate expression of their target genes (Krijger and de Laat, 2016).

1.2 The non-coding genome in development and disease

The classical view of gene regulation respects the central dogma of molecular biology dictating that "DNA makes RNA makes protein" (e.g. TFs), forming the basis of how biological processes are orchestrated. In the recent decade this simplistic view has been challenged with the conceptual rise of the "RNA regulatory world", representing a paradigm shift in the understanding of how gene expression is regulated. This shift was facilitated by the advent of deep sequencing technology such as RNA-seq (Wang et al., 2009), where whole transcriptome analyses by the ENCODE consortium uncovered pervasive transcription of ∼60 to 70% of the human genome that substantially overlaps non-coding regions (The ENCODE Project Consortium, 2007; Djebali et al., 2012), thus throwing non-coding RNAs (ncRNAs) into the spotlight. Numerous studies that followed attempted to present ncRNAs as new game-players of gene regulation in vertebrates, proposing they would act as fine-tuners of the gene regulatory machinery. Collectively, these hypotheses and their corresponding analyses have been driving the rapidly evolving field of the "regulatory RNAs", now mainly focused in uncovering their potential function. However, like with any other emerging field, this rapid progress and the "abundance of the unknown" constantly breeds emerging questions, as more discoveries were unfurled. It is important to note, that the production of functional transcripts from non-coding genomic regions (traditionally regarded as "junk DNA") was not an entirely new phenomenon and has been known for many years prior to reports from the ENCODE Project. Extensive research over the past few decades has led to the description of the various classes of non-coding RNAs such as small nucleolar RNA (snoRNA) (reviewed in (Tollervey and Kiss, 1997)), small interfering RNA (siRNA) in plants (Hamilton and Baulcombe, 1999), Piwi-interacting RNA (piRNA) (reviewed in (Iwasaki et al., 2015)) and micro RNA (miRNA) (reviewed in (He and

Hannon, 2004). More recently, long non-coding RNA (lncRNA) (including long intergenic/intervening non-coding RNA, *lincRNA*), enhancer-derived RNA (eRNA) and circular RNA (circRNA) have gained prominence as research groups raced to unfold mysteries surrounding their biogenesis and biological importance. For the purpose of this introduction, only the latter three "newcomer" classes of ncRNAs will be discussed.

The FANTOM consortium aims to "*assign functional annotations to full-length cDNAs that were collected during the Mouse Encyclopaedia Project at RIKEN*" (from the official FANTOM website at fantom.gsc.riken.jp). Characterisation of the mouse transcriptional landscape by genome-wide analysis of transcription start and termination sites revealed that non-protein-coding RNAs were slightly less conserved on average than their mRNA "counterparts" (5′ and 3′ UTRs), but their promoters appeared to be more conserved than that of protein-coding mRNAs, as well as contained binding sites for known TFs (Carninci et al., 2005). The ncRNAs studied were shown to also display cross-species positional conservation. Furthermore, non-coding RNAs were also uncovered by screening highly conserved non-coding regions of the human genome for thermostable structural RNAs (Washietl et al., 2005a). As well as recovering many then-curated ncRNAs including microRNAs and snoRNAs, the authors reported that nearly half of the high-confidence conserved structural RNAs identified were in fact unannotated and located >10 kb away from a protein-coding gene in the human genome, while the majority of the remaining half that overlapped known genes was located within the introns. This was remarkable considering the near-complete annotation of the human genome at the time of publication and in hindsight presented a tantalising premonition of the future boom in lncRNA and eRNA research.

1.2.1 Enhancer-derived RNAs (eRNAs)

ERNAs respond to biological stimuli and may affect chromatin looping

Enhancer-derived RNAs, or eRNAs, came into the limelight with a study describing developmentally regulated enhancers in mouse cortical neurons (Kim et al., 2010). Of the tens of thousands of functional enhancers defined by p300/(CREB Binding Protein) CBP-binding (both belong to the same family of transcriptional co-activators), a quarter were revealed to also be bound by RNA Polymerase II (RNAPII). Expression of eRNAs from this cohort was dynamic and responsive to membrane depolarisation with transcription proceeding bidirectionally on both strands. The spread of eRNA transcription also appeared to be controlled to a certain extent, diminishing at the ends of their respective H3K4me1-demarcated enhancer domains. In this study the authors also reported that eRNAs-of-interest were undetectable in poly(A) fractions, but detected at higher levels by randomly primed qPCR reactions, suggesting that these eRNAs were mostly non-polyadenylated.

Further studies demonstrated the sensitivity of eRNA induction as a hallmark of cellular response to biological stimuli (Wang et al., 2011a; Lam et al., 2013; Li et al., 2013), suggesting, in addition, that eRNA transcription can sometimes be correlated to regulation of chromatin looping (Hsieh et al., 2014; Melo et al., 2013). Macrophage-specific Rev-Erb response elements showed inhibition of eRNA transcription following Rev-Erb binding, functioning upstream of Rev-Erb-mediated repression of gene expression in bone marrow-derived macrophages (Lam et al., 2013). Pre-mRNA expression at a nearby locus was shown to be reduced following siRNA knockdown of its adjacent eRNA. eRNA transcription following androgen treatment of LNCaP cells (derived from human prostate adenocarcinoma) recapitulated patterns of androgen receptor (AR)-binding events, where eRNAs were upregulated at ectopic AR-binding regions and diminished in AR-binding regions lost following FoxA1 RNA interference, mimicking FoxA1 downregulation preceding prostate tumour progression (Wang et al., 2011a). This “phenotypic correlation” surpassed

that of a comparable approach using binding patterns of known proteins for enhancer function, p300b and Med12, as a readout. Another study in the same cell type reported that knocking down an eRNA of interest impaired chromatin looping between the enhancer with its target promoter, contributing to loss-of-expression of the target gene (Hsieh et al., 2014). Similar findings were also reported in a breast cancer-derived cell line, MCF-7. eRNAs were produced at p53-target enhancers following p53 binding and this phenomenon was required for p53-mediated activation of gene expression at their target loci, contributing to cell-cycle arrest (Melo et al., 2013). In a separate study, E2-ligand binding to oestrogen receptor α (ER-α) also induced rapid bidirectional activation of transcription of eRNAs at target enhancers (Li et al., 2013). Knocking down of an eRNA transcript also led to inhibition of gene expression at the adjacent protein-coding locus, and this effect was correlated with a decrease in cohesin recruitment to the enhancer for chromosomal looping. While these studies painted an early picture of eRNA expression and their potential function in regulating enhancer-mediated gene expression, mechanistic details surrounding these observations remained elusive. Later studies attempted to bridge this gap by focusing on eRNA crosstalk with the chromatin landscape, molecular partners during gene regulation, as well as attempting to distinguish between functionality of eRNA transcription or their RNA transcripts - as discussed below.

ERNAs as regulators of chromatin remodelling

The link between enhancer transcription and chromatin remodelling was shown in macrophages, where TLR4 signalling led to formation of *de novo* enhancers via the action of methyltransferases that deposited H3K4me1/2 histone marks at these regions (Kaikkonen et al., 2013). Interestingly, at these TLR4-activated enhancer loci, chromatin remodelling events were significantly reduced when RNAPII elongation was inhibited, but not via antisense oligonucleotide (ASO) knockdown of the eRNA transcript. Contrary to earlier findings, this study presented an example scenario where transcription at enhancers (rather than the transcripts themselves) could be

functionally important in exerting their biological effects. RNAi knockdown of two different eRNAs has led to decreased RNAPII-binding at the MYOD1 promoter and lower DNaseI sensitivity at a MYOG enhancer, respectively, illuminating an indirect role of eRNAs in regulating chromatin accessibility of enhancers and promoters during the myogenic differentiation programme (Mousavi et al., 2013).

ERNAs as molecular tethers that regulate transcription of genes

An in-depth study into RNA-mediated binding of a ubiquitiously-expressed TF, Yin-Yang (YY1,) to active promoters and enhancers in embryonic stem cells (ESCs) revealed key findings demonstrating another potential mode-of-action for eRNA-mediated regulation of gene expression (Sigova et al., 2015). Temporarily inhibiting RNAPII elongation at YY1 binding sites to decrease eRNA production led to a modest but significant loss of YY1 binding at these sites, which was rescued when the inhibition was removed. This phenomenon was also shown to be sensitive to steady-state levels of eRNA, as inhibiting post-transcriptional clearance of the eRNAs by exosome knockdown and allowing eRNAs to accumulate exhibited a similar negative effect on YY1-occupancy at its genomic target regions. eRNAs were also demonstrated to be involved in negative elongation factor (NELF) release from poised promoters, a process required for RNAPII elongation to proceed, which was blocked at the *Arc* and *Gadd45b* loci in neuronal cells after being subjected to shRNA-mediated knockdown of their corresponding eRNAs (Schaukowitch et al., 2014).

Global eRNA transcription has specific spatiotemporal patterns

While dissecting eRNA mechanisms at specific loci has contributed significantly to the current understanding of this ncRNA species, these studies did not address if their expression was developmentally regulated in a tissue-specific manner. Indeed, it was not until 2014 that Wu *et al.* demonstrated *in vivo* that global eRNA expression profiles correlated with tissue-specific enhancer activity in developmental mouse tissues, with eRNAs forming the majority proportion of ncRNAs with tissue-specific

expression (Wu et al., 2014). A study by the FANTOM consortium later showed that genome-wide eRNA transcription also appeared to be temporally regulated, often preceding transcription of associated protein-coding genes (Arner et al., 2015). Widespread eRNA transcription in lung fibroblasts were also shown to be more indicative of enhancer activity compared to H3K27Ac ChIP-seq profiles (Zhu et al., 2013). Similarly, we have also shown *in vivo* that bidirectional transcription could be used to identify clusters of putative enhancers that were functionally associated with NC ontogeny, thus recapitulating the molecular signature of their "host" cell type (Trinh et al., 2017). In short, eRNA transcription serves as a biological indicator during gene-switching events and acts as a qualitative readout of activated gene expression programmes.

Given the highly correlative nature of studies involving eRNAs, their importance is currently being actively debated. Another study systematically and globally dissected the phenomenon of bidirectional transcription at enhancers across a spectrum of human cell lines and concluded that transcription simply flagged up regions of accessible chromatin, and functional enhancers were still more accurately predicted using histone marks (Young et al., 2016). Hence, the debate remains if eRNAs are truly functional, and if yes, whether it is the eRNA transcripts or the act of their transcription that mediates their biological effects.

1.2.2 RNA Polymerase II pausing at promoters

Differences between eRNAs and promoter RNAs

Like eRNAs, nascent RNA sequencing had also revealed pervasive pausing and divergent initiation of RNAPII at human promoters (Core et al., 2008) and promoters in other metazoans (reviewed in (Adelman and Lis, 2014)). Following identification of eRNAs, Kim *et al.* presented arguments against the possibility that their neuronal eRNAs were actually unannotated promoters (Kim et al., 2010). Unlike in the case of eRNAs, this study found that promoters with bidirectional transcription

have low H3K4me1 and high H3K4me3 histone mark signal (and *vice versa* for enhancers with eRNAs). The report also pointed out that eRNAs tended to be shorter (<2 kb), and did not produce detectable levels of polyadenylated transcripts. However, it must be noted that the latter claim may be locus-dependent, as a separate study had reported the discovery of Ccl5-related eRNAs that were polyadenylated, but still supported the divergence in chromatin signature between eRNAs and promoter RNAs (de Santa et al., 2010). Kim *et al.* also noted that sense transcription is more prevalent than antisense transcription at promoters, while eRNA production seemed less biased towards a particular strand. This observation was similar to our observations of zebrafish promoter and enhancer-associated RNAs from NC cells that were analysed globally (Trinh et al., 2017).

Potential function of RNAPII pausing at promoters

Poised RNAPII at promoters lacking H3K4me3 were reported to be developmentally dynamic in *Drosophila* in a stage-specific fashion, suggesting that these H3K4me3-lacking promoters were being prepared (or "primed") to establish a permissive state for future activation (Gaertner et al., 2012). This phenomenon was also observed and described in the zebrafish during embryogenesis (Nepal et al., 2013). Possible reasons for RNAPII pausing included it being part of a mechanism to integrate regulatory signals before gene activation, triggering exosome degradation of the nascent transcripts (Henriques et al., 2013), as well as contributing to the demarcation of nucleosome-free regions at promoters and selectively "exposing" specific TF-binding motifs in order to establish a permissive chromatin microenvironment for gene expression to proceed (Scruggs et al., 2015).

1.2.3 Long non-coding RNAs (lncRNAs)

Long non-coding RNAs (lncRNAs) resemble protein-coding transcripts in that they are often polyadenylated, are at least 200 bp in length, they undergo splicing, and

have chromatin marks associated with actively transcribed genes (Guttman et al., 2009). In 2012, the GENCODE v7 catalogue of ∼15,000 lncRNAs from ∼10,000 genes in the human genome (most complete human lncRNA annotation at the time) was published by the ENCODE Consortium (Derrien et al., 2012). Although the biological functions of the vast majority of these lncRNAs were unknown, the study presented a classification of lncRNAs based on their location relative to the annotated protein-coding genes; antisense RNAs (directly overlapping coding transcripts), lincRNAs (long *intergenic/intervening* non-coding RNA), sense overlapping transcripts (containing coding genes within their introns), sense intronic transcripts (contained within introns of coding genes) and processed transcripts without open reading frames (ORFs) that did not fall into any of the other categories. Another class of lncRNAs, long upstream antisense transcripts (LUATs), described later, are transcribed divergently from promoters shared with protein-coding mRNA (Lepoivre et al., 2013). Lepoivre *et al.* noted that functional groups of genes involved in transcriptional and developmental regulation were highly enriched amongst LUAT-associated genes.

LncRNAs regulate biological processes

The roles of lncRNAs were largely under-explored, only coming to the forefront following landmark studies describing lncRNAs in the HOX locus. The discovery of HOTAIR lncRNA kickstarted the dawn of modern lncRNA biology (Rinn et al., 2007). Measuring 2.2 kb in length and residing within the human HOXC locus, HOTAIR silenced gene expression *in trans* across 40 kb region of the HOXD locus by interacting with Suz12, an RNA-binding protein containing a zinc-finger domain, and Enhancer of zeste homolog 2 (EZH2), a histone methyltransferase - two components that form part of Polycomb Repressive Complex 2 (PRC2). HOTAIR was shown to mediate Suz12 occupancy as well as H3K27 trimethylation at the HOXD locus. Mechanistically, it has been shown that HOTAIR lncRNA contained 5′ and 3′ functional domains that bind PRC2 and LSD1, respectively, where their modes of action

were significantly concentrated at the proximal promoter regions within the HOXD locus (Tsai et al., 2010). This activity was demonstrated to be important for the chromatin landscape reprogramming during induction of cancer metastatis (Gupta et al., 2010; Alves et al., 2013), although *in vivo*, HOTAIR transcript was shown to be dispensable during mouse development (Amândio et al., 2016). Similar to the findings of HOTAIR and its mechanism of action, HOTTIP, another lncRNA, transcribed in this case from the HOXA locus, was shown to target WDR5/MLL complexes to this locus and drive H3K4 trimethylation to activate gene transcription (Wang et al., 2011b). Another cancer-related candidate described was lincRNA-p21, was demonstrated to mediate transcriptional repression of target genes in the p53 pathway via hnRNP-K-binding and therefore inducing apoptosis (Huarte et al., 2010). This led to an explosion of candidate-gene approach studies describing the roles of lncRNAs in other disease models apart from cancer, including cardiovascular disease (Han et al., 2014), neurological disorders (reviewed in (Briggs et al., 2015)) and immunity (Hu et al., 2013; IIott et al., 2014).

On a broader scale, it is of great scientific interest to dissect the role of lncRNAs in gene regulatory networks underlying developmental programmes. This is in order to understand the native roles of lncRNAs within a complex environment where intricate and finely tuned gene regulation are the main driving forces behind success. Although the presence of lncRNAs in loci of imprinted genes had been known for awhile, it was not until recently that mechanisms of their action were studied in detail and elucidated. As more uncharacterised lncRNAs are being discovered, these "classical" lncRNAs have gathered a renewed interest as they form a pre-existing scientific scaffold for researchers to build upon. Indeed, the "lncRNA-mediated chromatin regulation" model presented by HOTAIR and HOTTIP were echoed in recent studies involving *Airn*, *H19* and *Xist* (Santoro et al., 2013; Monnier et al., 2013; Engreitz et al., 2013), thus providing a certain level of assurance to current findings involving newer lncRNAs.

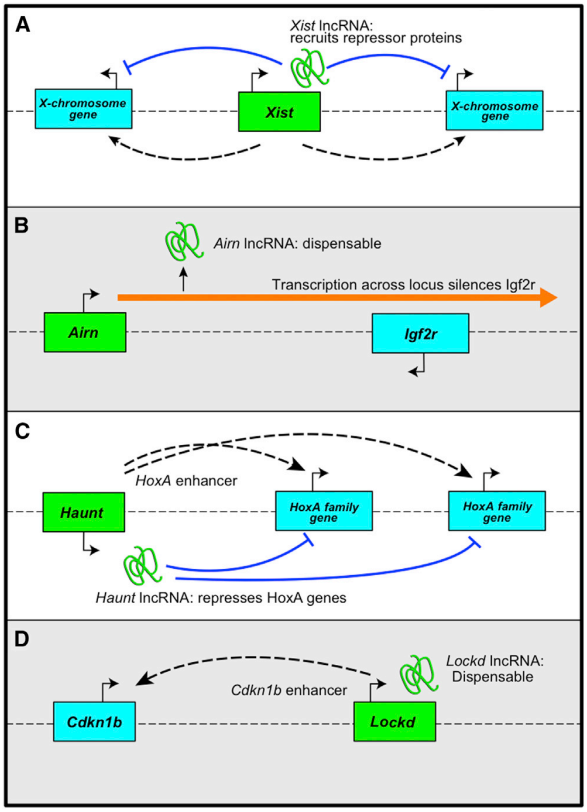


FIGURE 1.4: Long non-coding RNAs regulate expression of neighbouring genes (Paralkar et al., 2016).

Four examples of how long non-coding RNAs (lncRNAs) have been proposed to regulate expression of neighbouring genes. *Xist* lncRNA is transcribed from its locus and silences neighbouring X-chromosome genes by recruiting repressor proteins. The *Airn* lncRNA transcript is dispensable but its transcription by RNA Pol II extends to the neighbouring *Igf2r* locus, which silences expression of the latter. Similar to *Xist*, *Haunt* lncRNA transcript facilitates repression of neighbouring *HoxA* genes. Transcription within a *Cdkn1b* enhancer locus called *Lockd* is necessary for *Cdkn1b* gene expression, although the associated *Lockd* lncRNA itself is dispensable.

Genome-wide identification and characterisation of lncRNAs

Spurred by these early successes tackling lncRNA function, it became desirable to employ a systems-level approach in order to deduce general mechanistic details of lncRNA form and function. Khalil *et al.* went on to identify additional lncRNAs (but only focusing on lincRNAs) that presented a similar mode of action to HOTAIR and reported a cohort of human lincRNAs that associated with PRC2 (Khalil et al., 2009). It was then proposed that lincRNA expression in mESCs (a cohort of around 100 candidates) was regulated by pluripotency factors such as Oct4, Nanog and Sox2, and about a third of this cohort could be participating in pluripotency maintenance by forming scaffolds to recruit chromatin regulators to target genes (Guttman et al., 2011). These findings fed into the working hypothetical model where lncRNAs (or in this case, lincRNAs) guided chromatin-modifying complexes to specific genomic loci, in order to regulate gene expression (reviewed in (Guttman and Rinn, 2012)). Another such example is that of *Dali*, a lncRNA which binds directly to DNA

methyltransferase 1 (Dnmt1) to regulate gene expression during neural differentiation of mouse N2A cells (Chalei et al., 2014).

To mine for lncRNA candidates, Guttman *et al.* presented an approach using chromatin signatures to uncover mouse lncRNAs intervening protein-coding genes in an intergenic fashion (lincRNAs) (Guttman et al., 2009). With RNA-seq raising in popularity as an alternative approach to microarrays, Guttman then presented a method to identify novel transcripts, including lncRNAs, using genome-guided *de novo* assembly of RNA-seq reads (Guttman et al., 2010). However, parallel studies attempting to tackle the function of lncRNAs on a global scale utilising RNA-seq and ribosome-profiling have painted a less than clear picture of lncRNA expression and function (Pauli et al., 2012; Ulitsky et al., 2012; Chew et al., 2013; van Heesch et al., 2014). The labs of Alex Schier and David Bartel independently published whole embryo datasets of zebrafish lncRNAs (Pauli et al., 2012; Ulitsky et al., 2012), but there was very little overlap between the two studies in terms of identified lncRNAs (Chew et al., 2013). Ribosome-profiling datasets (Ingolia et al., 2012) attempting to clarify this discrepancy only led to more confounding questions as researchers discovered that translation is pervasive, occurring outside of protein-coding loci and even on annotated lncRNA transcripts. This raised the possibility that lncRNAs datasets available at the time contained a high number of non-annotated protein-coding genes (Ingolia et al., 2014). On the other side of the divide, Guttman asserted that ribosomal occupancy on a transcript alone did not necessarily indicate that the transcript encoded a protein (Guttman et al., 2013). These findings clearly illustrated the need for re-examination of the methods used in genome-wide identification of lncRNAs, in order to further minimise the likelihood of false-positives that may lead to data misinterpretation and obstacles in downstream functional analyses.

A FANTOM study cataloguing lncRNAs based on their transcription initiation regions further divided lncRNAs based on their origin of transcription (Hon et al., 2017). lncRNAs expressed from enhancers (e-lncRNA i.e. eRNAs - see previous

section) formed the majority (∼70%) of the catalogue, followed by lncRNAs transcribed divergently from promoters (divergent p-lncRNAs) while only a small portion of transcripts were intergenic (intergenic p-lncRNAs). Collectively, ∼2000 of these lncRNAs were found to be cell type-specific, with higher sequence conservation as well as being associated with human traits identified from genome-wide association studies (GWAS). This highlighted a novel approach in narrowing down lncRNAs that were more likely to be functional based on their connection to known phenotypes. Divergent p-lncRNAs were later shown to regulate gene expression of adjacent genes in pluripotent cells and thus mediate mesendodermal lineage differentiation (Luo et al., 2016) and gene activation during mouse preimplantation development (Hamazaki et al., 2015).

Clearly, a set of rules was needed to functionally segregate lncRNAs from protein-coding mRNAs. Mele *et al.* undertook this task and demonstrated that unlike promoters of protein coding genes yielding mRNAs, lincRNA promoters were generally depleted of histone marks but enriched in H3K9 trimethylation (a mark typically associated with transcriptional repression) and had an enrichment of conserved binding sites for specific TFs (GATA family, JUN and FOS) (Melé et al., 2017). The most distinctive feature of lincRNAs was lower splicing efficiency compared to mRNAs, despite having similar stability and features of the U1-PAS axis (enrichment of U1 binding motifs and depletion of polyadenylation motifs downstream from the TSS (Almada et al., 2013)).

Debates surrounding lncRNA form and function

A common debate surrounding the lncRNA-mediated gene regulation model is the physical nature of their action. Previous studies have shown that lncRNAs can silence gene expression *in trans* (e.g. Xist and HOTAIR), further supported by 26-30 lincRNAs in mESCs which appeared to affect gene expression *in trans* (Guttman et al., 2011). On the other hand, knockdown of *lncRNA-a7* specifically decreased expression of its adjacent target gene Snai1 *in cis* while leaving expression of other

neighbouring genes unchanged (Ørom et al., 2010). To further complicate matters, there was a conundrum of whether the lncRNA transcripts themselves, or the process of lncRNA transcription, is of functional importance. While this had led to divided opinions for or against each possibility, the truth is likely to be locus-dependent. Given the highly correlative nature of studies involving lncRNAs to date, additional studies directly addressing this question are needed before more definitive answers can be obtained. One example of such a study approached the subject by examining 12 genomic loci that produced lncRNAs (Engreitz et al., 2016a). Out of those 12, five were found to influence expression of neighbouring genes *in cis*. The authors established that these effects were independent of the lncRNA transcripts but reliant on their transcription, given that neighbouring gene activation responded to changes introduced artificially to control the length or amount of transcriptional activity within the lncRNA locus, but did not appear to require any specific sequence elements in the mature lncRNA transcript itself. Similarly, transcription of another lncRNA called *upperhand*, but not the *upperhand* transcript itself, controls *Hand2* expression in mice during heart development (Anderson et al., 2016). Two genetic studies in *Drosophila* also presented evidence that transcription across non-coding genomic loci, the Polycomb group response elements (PREs) (Schmitt et al., 2005) in one case and the *bithoraxoid* locus (Petruk et al., 2006) in another, was required to regulate neighbouring genes *in cis*. Interestingly, lncRNAs have been observed globally to have highly conserved promoters at levels comparable to the promoters of protein-coding genes (Carninci et al., 2005; Alam et al., 2014). This finding further supports the notion that the act of transcription rather than the transcripts themselves contributes towards biological function of lncRNAs. Recently it has also been shown that lncRNAs could exert their function by associating with super-enhancers (Ounzain et al., 2015), or even by forming micropeptides (Anderson et al., 2015).

Unlike smaller ncRNAs such as miRNAs and snoRNAs, the apparent lack of sequence conservation in lncRNA transcripts is often presented as damning evidence

by sceptics that they lack function. However, this was countered by the argument that lncRNAs tended to contain patches of high conservation within their overall sequence indicative of rapid evolution rate (Pang et al., 2006), unlike in the case of mature protein-coding mRNAs where sequence conservation usually spans large portions of their open reading frame (ORF). A solid argument against such use of protein-coding transcript conservation metric to assess the functional importance of lncRNAs was made in a large study characterising lincRNAs in zebrafish (Ulitsky et al., 2012). Another study involving phenotypic analysis of multiple mouse lncRNA knockout strains (Sauvageau et al., 2013), although initially promising, had also raised doubts regarding the validity of the produced library due to limitations in distinguishing between lncRNA function and that of putative *cis*-regulatory elements within the deleted locus (Bassett et al., 2014). *In vivo* results also appeared to be inconsistent; *Malat1*, which has been shown to be highly expressed in metastatic cancer cells (Ji et al., 2003), lacked any obvious phenotype when knocked out in the mouse (Zhang et al., 2012).

lncRNAs and metazoan evolution

lncRNAs have been hypothesised to drive evolution of vertebrates, where the complexity and scale of lncRNA expression correlated with the level of "complexity" of an organism. This attractive premise was highlighted in a study that explored the evolutionary history of lncRNAs in tetrapods (Necsulea et al., 2014), as well as by another study demonstrating that the human brain expressed a higher cohort of lncRNAs compared to the macaque brain (He et al., 2014). Unsurprisingly, a generous amount of lncRNA research is currently geared towards neuroscience and this aspect was reviewed extensively in (Roberts et al., 2014).

1.2.4 eRNAs and lncRNAs - different, same, or both?

Most early studies suggested that eRNAs and lncRNAs were two separate and different classes of non-coding RNA species; eRNAs were generally non-polyadenylated with enhancer-associated histone marks i.e. H3K4me1 and H3K27Ac (Kim et al., 2010) whereas lncRNAs tended to have defined promoters, performed RNA-mediated functions, and were marked with protein-coding like chromatin signatures i.e. H3K4me3 at the promoter and H3K36me3 at the body of the non-coding gene (Guttman et al., 2010; Ørom et al., 2010). HOTTIP transcribed from the 5′ tip of the HOXA locus coordinated activation of several 5′ HOXA genes *in vivo* downstream of chromosome looping, as close proximity was needed for HOTTIP to bind and target WDR5/MLL complexes to the HOXA locus and add the H3K4me3 mark (Wang et al., 2011b). This suggested "*bona fide*" lncRNA function independent of enhancer-like function, again reiterating the model where lincRNAs function as transmitters of information from chromosome looping (while not being involved in looping itself) to chromatin landscape in the form of modifications leading to gene expression.

This argument was perhaps too simplistic in nature; Orom had pointed out in his 2010 paper that the orientation-independent mode of action of his lncRNA to activate neighbouring gene expression was reminiscent of what is commonly understood to be a feature of enhancer action (Ørom et al., 2010) . This observation came to fruition in a study describing the cooperation between a lncRNA and the Mediator protein complex to promote chromatin looping (Lai et al., 2013). Knocking down this lncRNA or Mediator negatively affected chromosomal looping events between the *lncRNA-a7* and its interacting Snai1 promoter, demonstrating the functional overlap between these two classes of ncRNA regulatory elements (reviewed in (Ørom and Shiekhattar, 2011)).

The lines between eRNAs and lncRNAs were further blurred by a genome-wide

study that identified candidates from both classes that were p53-regulated. Intriguingly, one p53-regulated lncRNA (previously annotated as linc00086) was demonstrated to bind to enhancers and contribute to p53-regulated eRNA production, and was therefore christened as LED (LncRNA activator of Enhancer Domains) (Léveillé et al., 2015). Recently, Paralkar *et al.* presented an example of a "lncRNA" locus that was actually an enhancer, where its DNA sequence (and not the polyadenylated RNA transcript being produced) was required for looping and normal transcription of its target gene (Paralkar et al., 2016). This highlighted a very important point, often dismissed by lncRNA optimists in the quest to "discover" biological functions of novel lncRNAs; given the highly pervasive nature of transcription in the genome, it is imperative to precisely define the hypothesis to test and the appropriate experimental approaches in order to distinguish between enhancer-, enhancer transcription- or RNA-mediated regulation of gene expression.

1.2.5 Circular RNAs

Circular RNAs (circRNAs) are transcripts with an unusual structural organisation. Unlike linear mRNA transcripts, circRNAs have the positions of their 3′ and 5′ sequences swapped, and appear to undergo different co- or post-transcriptional modifications to produce their characteristic uncapped and non-polyadenylated circular transcripts. The first circRNA to be described, *circ-Sry*, was found to be the most abundant *Sry* RNA molecule in the testis except within the genital ridge, where the linear *Sry* transcript was predominantly expressed and important for sex determination in mammals (Capel et al., 1993). Capel *et al.* suggested that circular transcripts could represent an additional mechanism to prevent translation, similar to splicing variants that introduced alternative stop codons, as initially observed in *Drosophila* (Samuels et al., 1991). This finding revisited enigmatic questions relating to our understanding of how and why genes are expressed; in short, what does a cell have to gain by transcribing (and circularising) a gene that cannot be translated? The case

of circRNAs was then put to rest, until a resurgence of interest nearly two decades later following the advent of next generation sequencing (NGS).

It was then discovered that a significantly high proportion of spliced transcripts in the human genome were indeed circular RNA isoforms (Salzman et al., 2012) and circRNAs were enriched in neural tissues during human foetal development (Szabo et al., 2015). This led to a slew of studies attempting to identify more circRNAs (Memczak et al., 2013; Guo et al., 2014; Westholm et al., 2014; You et al., 2015) and characterise their biological functions. It has been suggested that circRNAs function as miRNA sponges (Hansen et al., 2013; Memczak et al., 2013; Xu et al., 2015), or regulate transcription of their parental genes in the nucleus (Li et al., 2015), and in certain cases may be translated into proteins (Wang and Wang, 2015).

Biogenesis of circRNAs

CircRNAs are largely processed from pre-mRNA post-transcriptionally, although this process is inefficient and therefore likely to rely on additional controls for them to accumulate to a steady state (Zhang et al., 2016). This could be achieved by modulation of the splicing machinery, as shown in a study that demonstrated a negative correlation between splicing efficiency and biogenesis of circRNAs, suggesting strong competition between circularisation and formation of mature mRNA transcripts (Ashwal-Fluss et al., 2014). Ashwal-Fluss *et al.* went on to describe a possible mechanism for this relationship; their circRNA-of-interest contained binding sites for the protein product of the parental gene, therefore forming a feedback loop between the two isoforms. It is currently still debated if alternative or canonical splicing events are predominant in the biogenesis of circRNAs with the evidence being presented to support both scenarios (Gao et al., 2016; Starke et al., 2015). Circularisation signals, such as complementary sequences and/or secondary structures, also appeared to be embedded within the introns flanking circularised exons (Liang and Wilusz, 2014; Zhang et al., 2014b; Starke et al., 2015). This observation may be genus-specific, as characterisation of circRNAs identified in adult zebrafish tissues

revealed that only 2% of exonic circRNAs had the characteristic reverse complementary sequences in flanking regions, reported in mammalian circRNAs (Shen et al., 2016). Last but not least, RNA-editing mechanisms could also contribute to regulation of circRNA formation. A highly-characterised A-to-I RNA editing enzyme, ADAR1, was found to inhibit circRNA formation by disrupting complementary base-pairing between introns brought together in close proximity during circularisation (Ivanov et al., 2015). In contrast, the RNA-binding protein Quaking (QKI) promoted circRNA formation, as introducing QKI binding sites into introns was demonstrated to be sufficient for generation of circRNAs (Conn et al., 2015).

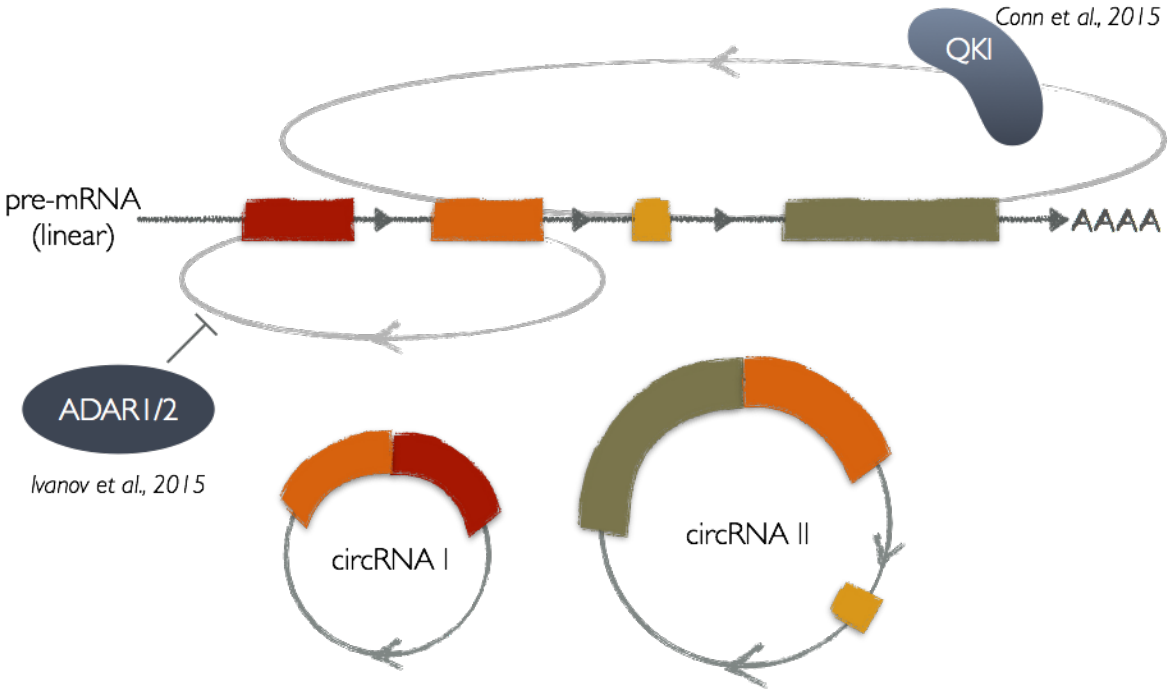


FIGURE 1.5: Formation of circular RNAs (circRNAs).

Circular RNAs (circRNAs) are formed either co- or post-transcriptionally from linear pre-mRNA that undergo back-splicing of its exons. This process has been shown to be regulated by Quaking (QKI) (Conn et al., 2015) and inhibited by ADAR1/2 (Ivanov et al., 2015).

1.2.6 Transposable elements

A FANTOM study surveying both mouse and human transcriptomes reported the high contribution of transposable elements (TEs) to transcriptional activity while

displaying tissue-specific patterns (Faulkner et al., 2009). Although the study did not fully delve into the link between TE transcription and biological activity, TE transcription was later shown to be a participant in transcriptional programmes driving embryonic development (Göke et al., 2015).

It has also already been known that a cohort of elements from the short interspersed repetitive element (SINE) retroposon TE family is highly conserved in tetrapods, where an ancestral SINE element was shown to have been co-opted over the course of evolution into a protein-coding exon and a distal enhancer (Bejerano et al., 2006). TE co-option as regulatory elements has also been reported to be important for dosage compensation of sex chromosomes (Ellison and Bachtrog, 2013) and the evolution of species-specific TF binding landscapes (Sundaram et al., 2014). MER130 repeats were co-opted and recruited during evolution of the neocortex, as their sequences were enriched among enhancers that are active during neocortex development in the mouse (Notwell et al., 2015). In humans, correlation of expression patterns across different tissue types revealed the associations of particular TEs with particular tissues in a systematic fashion, suggesting that TEs may principally function as enhancers (Pavlicev et al., 2015). Consistent with this hypothesis, histone hypomethylation (a known chromatin hallmark of enhancers) has been shown to mark specific TE families that were associated with tissue-specific enhancer activity, and their presence correlated with expression of genes important for the tissues they appeared to regulate (Xie et al., 2013).

1.3 Current methods and limitations in studying non-coding RNA

Early identification of non-coding elements discussed above was largely a serendipitous event, resulting from the efforts to annotate mammalian transcriptomes.

Genome-wide methods employed to actively produce lncRNA datasets have included microarrays, RNA-seq, *de novo* assembly of transcriptomes, histone-ChIP-seq and searches for functional structural RNAs using RNAz (Washietl et al., 2005b). Recently, single cell transcriptomics was also used for lncRNA identification (Zhang et al., 2014a). Furthermore, integration of multiple RNA-seq and ChIP-seq datasets can be used to computationally identify high-confidence candidates (Zhao et al., 2015). Structure-seq was recently developed for *in vivo* genome-wide profiling of RNA secondary structure (Ding et al., 2013) and could potentially be paired with RNA-seq as a powerful method for identification of multiple classes of ncRNAs.

While a 2005 FANTOM paper focused on polyadenylated mRNAs that were processed and exported to the cytoplasm, the authors highlighted that this excluded potentially functional non-polyadenylated nuclear RNAs which formed a large proportion of ncRNAs from intergenic regions (Carninci et al., 2005). As ncRNAs were increasingly reported to be functional in modulating chromatin changes, nuclear fractionation became a logical approach to enrich for candidates that were more likely to be functional. Using this method, Werner *et al.* produced a dataset of chromatin-tethered lncRNAs representing ∼60% of annotated lncRNAs (Werner and Ruthenburg, 2015). Furthermore, a paper characterising RNAPII transcription at sites overlapping enhancers demonstrated enrichment of these eRNAs in the nuclear compartment (de Santa et al., 2010). Technologies to genetically attain subcellular resolution currently exists with methods employing *in vivo* biotinylation (de Boer et al., 2003; Deal and Henikoff, 2010; Trinh et al., 2017) and those are readily suited to ncRNA identification. However, this may exclude potentially functional lncRNAs in the cytoplasm with currently unknown roles and modes of action (van Heesch et al., 2014).

Functional characterisation of non-coding RNA species is not quite as straightforward. Special considerations have to be taken into account (Bassett et al., 2014) due to their physical organisation within the genome, where they often overlap

protein-coding genes and often contain *cis*-regulatory elements in the form of transposable elements (Kelley and Rinn, 2012). While mouse knockout models have been employed in the past, this has largely been succeeded by methods that can distinguish between RNA- and DNA-mediated effects. RNA interference is a popular method-of-choice used in mammalian cell systems forming the bulk of the current literature. However this approach has been shown to be unsuitable for use in the zebrafish *in vivo* due to extensive off-target effects (Oates et al., 2000). Recently the successful use of antisense oligonucleotides (ASOs) were demonstrated in the zebrafish, presenting a promising alternative and tool for studying ncRNAs in this model (Pauli et al., 2015).

To dissect molecular mechanisms of candidate ncRNAs, RNA-centric methods were developed. CHART (**C**apture **H**ybridisation **A**nalysis of **R**NA **T**argets) method uses a hybridisation-based technique to specifically enrich endogenous RNAs along with their targets from crosslinked chromatin extracts, hence also revealing potential genomic sites of action of identified ncRNAs (Simon et al., 2011). Similarly, the ChIRP (**Ch**romatin **I**mmunoprecipitation by **R**NA **P**urification) method can also be used to isolate chromatin targets of defined RNA candidates (Chu et al., 2011). An extension of this method, dChIRP, was developed to identify functional domains within RNA transcripts if a higher resolution description of a favourite candidate ncRNA is desired (Quinn et al., 2014).

CRISPR/Cas9 technology has recently come to the forefront due to its flexibility and efficient utility in genome engineering. This led to multiple genome-wide screens of non-coding elements (Sanjana et al., 2016; Liu et al., 2016; Zhu et al., 2016), highlighting the high potential for this new technology to drive future scientific investigations into ncRNA function *in vivo*.

1.4 The non-coding genome in embryonic development

We conclude this introduction by presenting current examples of *in vivo* evidence of ncRNA function during embryonic development (for an authoritative review, please refer to (Pauli et al., 2011)). lncRNAs in particular were heavily featured in metazoan embryogenesis, including zebrafish (Ulitsky et al., 2011; Pauli et al., 2012; Dhiman et al., 2015). Tissue-specific examples of developmental lncRNAs in the mouse include *Fendrr* (heart and body wall development) (Grote et al., 2013) and *lincGET* (a nuclear lincRNA promoting *cis*-regulatory activity between its endogenous retrovirus (ERV) repeats and target genes during embryo cleavage) (Wang et al., 2016). Epigenetic coordination of embryonic heart transcription was shown to be dynamically regulated by lncRNAs (Matkovich et al., 2014). Furthermore, a lncRNA named *Braveheart* was shown to be necessary for the activation of a cardiovascular gene network that drives progression of nascent mesoderm to a cardiac fate in the mouse (Klattenhoff et al., 2013) by interacting with a zinc finger TF (Xue et al., 2016). In fact, both lincRNAs and eRNAs appeared to be involved in cardiac development (Ounzain et al., 2014). The importance of lncRNAs in heart development was further reiterated with the systematic identification and characterisation of cardiac lincRNAs in zebrafish (Wang et al., 2017). Last but not least, *TUNA* (in mouse) or *megamind* (in zebrafish) is a conserved lncRNA with tissue-specific expression in the central nervous system (CNS) of both species, proposed to control pluripotency and neural lineage commitment by recruiting specific RNA and chromatin-binding proteins to the Sox2 promoter (Lin et al., 2014).

In the subsequent chapters, we will demonstrate how we built a genetic toolkit in the zebrafish to allow analysis of RNA transcripts in subcellular compartments, yielding interesting and novel results linking non-coding RNA biology to zebrafish embryonic development.

Chapter 2

The *biotagging* toolkit for the analysis of subcellular compartments and molecular interactions in specific cell types

2.1 An *in vivo* biotinylation toolkit for zebrafish

The zebrafish is by many considered the rising star of vertebrate animal systems, due to its genetic amenability and high fecundity. While numerous embryology, histo- and cytochemistry as well as molecular biology methods are well-established in the zebrafish, the system lacks a purely genetic approach to isolate and study small/rare cell populations in isolation from their environing tissues. In this chapter, we will discuss the building of the *biotagging* toolkit, as well as describe the meticulous design of transgenic zebrafish lines as its components and optimisation of this system to enable genome-wide analysis of subcellular compartments in specific cell types. This body of work has resulted in a co-first author publication (Trinh et al., 2017).

2.1.1 Principles and history of *in vivo* biotinylation

Biotinylation in *E.coli* occurs when biotin holoenzyme synthetase, BirA, catalyses addition of a single biotin moeity onto a lysine residue within the biotin carboxyl carrier protein (BCCP) subunit of acetyl-CoA carboxylase (Fall, 1979). Elena de Boer *et al.* has adapted this naturally-occurring biochemical process to *in vivo* studies

and demonstrated that a transcription factor tagged with an artificial 23-aa acceptor peptide previously identified by Peter Schatz (Schatz, 1993) can be efficiently biotinylated in mammalian cells and transgenic mice (de Boer et al., 2003). Further work by Dorothy Beckett *et al.* in Schatz lab led to successful truncation of this 23-aa peptide to a 15-aa minimal substrate required for specific and efficient biotinylation by BirA, leading to the AviTag^{TM}, a peptide currently widely used for generation of biotinylation-ready fusion proteins (Beckett et al., 2008).

In vivo, biotin residues are recognised and bound with high affinity by avidin, a tetrameric protein enriched in the egg-whites of birds, reptiles and amphibians (Livnah et al., 1993). Although genetically-distinct from avidin, streptavidin (from *Streptomyces avidinii*) bears a highly similar crystal structure and is also able to bind biotin residues with high affinity. In fact, the streptavidin-biotin complex is currently the strongest known non-covalent interaction in nature, with dissociation constant of $K_d = 10^{-14}$ (Hendrickson et al., 1989; Weber et al., 1989) and is thus widely used for purification of biotinylated AviTagged proteins and their interacting molecular components (Kim et al., 2009).

2.1.2 The binary *in vivo* biotinylation system for zebrafish

To adapt the *in vivo* biotinylation to the zebrafish model, we designed a genetic binary system called *biotagging* consisting of two transgenic components. *BirA driver* transgenic lines carry a transgene enabling tissue-specific expression of *E. coli* BirA in selected cell types and *Avi effector* transgenic lines enable expression of the AviTagged protein-of-interest, expressed either ubiquitously or in selected cell populations. When heterozygote *BirA driver* and the *Avi effector* are crossed, biotinylation will only occur in 25% of the offspring carrying both alleles, and only in cells where both BirA and the AviTagged target are expressed. (Figure 2.1).

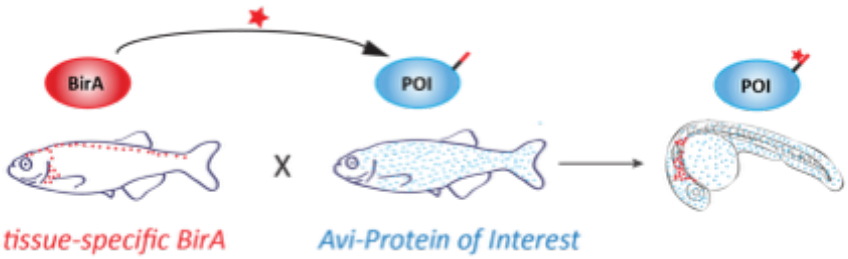


FIGURE 2.1: The *biotagging* toolkit.

Schematic diagram of the binary *in vivo* biotinylation system in zebrafish for *biotagging*. Biotin ligase is expressed in a tissue-specific (or ubiquitous, depending on usage) manner in a *BirA driver* transgenic line. Similarly, the AviTagged protein-of-interest (POI) is expressed in an *Avi effector* transgenic line. When heterozygote parents of both lines are crossed, biotinylation of the AviTagged POI only occurs in cells that also express BirA in ~25% of the offspring that have inherited both alleles.

The power of *biotagging* lies in the astute selection of *BirA drivers* and *Avi effectors*, allowing the flexibility needed to design experiments unique to each biological question addressing different developmental or disease processes. In this chapter, we will present our "starter" toolkit of predefined *BirA drivers* and *Avi effectors* (Table 2.1) by describing their design, followed by the discussion of key optimisation parameters allowing for maximal efficiency of the *biotagging* method. We will end the chapter by demonstrating the utility of our *biotagging* method in a proof-of-principle experiment characterising the nuclear transcriptome of cardiomyocytes.

2.2 Establishment of *Avi effector* transgenic lines

We designed *Avi effector* transgenics for multiple lines of scientific investigation. The first approach enables selection and purification of specific subcellular compartments (nuclear or polysomal). The second approach targets ubiquitously-expressed chromatin remodelling proteins for *in vivo* biotinylation, enabling antibody-free ChIP experiments ("biotin-ChIP") and facilitating tissue-specific epigenomic studies in the developing embryo. Similar to the second approach, the third method (not discussed in this thesis) is the AviTagging of specific TFs-of-interest for biotin-ChIP. All

Biotagging transgenic line	Type	Allele name	Expression
Tg(actb2:BirA-2a-membCherry)	BirA driver	ct709a	Ubiquitous
Tg(ubb:NLS-BirA-2a-mCherry)	BirA driver	ox114	Ubiquitous
Tg(ubb:BirA-2a-mCherry)	BirA driver	ox115	Ubiquitous
Tg(myl7:BirA-2a-membCherry)	BirA driver	ct704a	Myocardium
TgBAC(sox10:BirA-2a-mCherry)	BirA driver	ox104a	Neural crest
TgBAC(sox10:NLS-BirA-2a-Citrine)	BirA driver	ox108a	Neural crest
Tg(actb2:Avi-Cerulean-GgRanGAP)	Avi effector	ct700a	Ubiquitous
Tg(actb2:GgRanGAP-Cerulean-Avi)	Avi effector	ct701a	Ubiquitous
Tg(ubb:Avi-Cerulean-GgRanGAP)	Avi effector	ox113	Ubiquitous
Tg(actb2:Avi-Cerulean-DrRpl10a)	Avi effector	ox111	Ubiquitous
Tg(ubb:Avi-Cerulean-DrRp110a)	Avi effector	ox112	Ubiquitous

TABLE 2.1: *Biotagging* transgenic lines presented in this chapter.

The "ct" allele names denote constructs cloned and injected by collaborators at Caltech prior to this thesis, while "ox" allele names denote constructs cloned and injected in our lab as part of this thesis. ox104a and ox108a were cloned by U. Senanayake and injected by the author. All other "ox" lines were cloned and injected by the author.

three approaches capitalise on the efficiency and specificity of *in vivo* biotinylation coupled with fully-optimised purification protocols that we have developed, which will also be discussed in this chapter.

2.2.1 *Avi effectors* for subcellular compartmentalisation

To allow selection and purification of biotinylated nuclei, we adapted the **I**solation of **N**uclei **TA**gged in specific **C**ell **T**ypes (INTACT) method developed by Roger Deal and Steven Henikoff that targets the WPP domain of Ran-GTPase Activating Protein 1, RanGAP1, for biotinylation by BirA. RanGAP1 is localised to the nuclear pore complex of eukaryotic cells except in yeast (Matunis et al., 1998; Hopper et al., 1990). By co-expressing Avi-WPP (RanGAP1) and BirA, Deal and Henikoff successfully demonstrated that cell type-specific biotinylation enabled isolation of nuclei from hair and non-hair cells of the root epidermis in *A. thaliana* for genome-wide RNA- or ChIP-profiling (Deal and Henikoff, 2010).

We sought to build similar transgenic lines for INTACT experiments in zebrafish. We cloned two constructs encoding N- or C-terminally AviTagged carboxyl-terminal

domain of the avian RANGAP1 protein from *G. gallus* ("GgRanGAP"), that putatively associates with the nuclear pore in zebrafish cells under the control of the zebrafish *actb2* (also known as *bactin2*) ubiquitous promoter. Using *tol2*-mediated transgenesis (Kawakami, 2004), we generated stable transgenic lines that ubiquitously express these recombinant proteins - *Tg(actB2:Avi-Cerulean-GgRanGAP)*^{ct700a} and *Tg(actB2:GgRanGAP-Cerulean-Avi)*^{ct701a}. In these lines, GgRanGAP was fused to a blue variant fluorescent protein (Cerulean) to facilitate embryo screening and to help validate subcellular localisation of the GgRanGAP fusion proteins.

Interestingly, the GgRanGAP fusions were able to localise to the outer nuclear envelope while transgenes containing zebrafish RanGAP C-terminal domain did not (see Supplemental Information of (Trinh et al., 2017)). Although both N-term and C-term AviTagged GgRanGAP showed similar localisation of the transgene protein to the nuclei and were therefore interchangeable, we chose to perform all subsequent experiments using *Tg(actB2:Avi-Cerulean-GgRanGAP)*^{ct700a} (N-term Avi-GgRanGAP) in which the position of the
GgRanGAP domain was equivalent to that in the endogenous RanGAP protein, i.e. at the carboxyl end (Figure 2.2).

Ubb promoter analogues in human and *Drosophila* have been well-characterised and provide faithful ubiquitous expression of transgenes in their hosts (including in mice for the human promoter) (Lee et al., 1988; Schorpp et al., 1996). To remedy several limitations of ubiquitous promoters (including *actb2*) previously used in zebrafish research, Christian Mosimann *et al.* from Zon lab cloned and characterised the zebrafish *ubiquitin B* (*ubb*) promoter (Mosimann et al., 2011). As *ubb* promoter has since been accepted by the zebrafish community as the "gold standard" ubiquitous promoter, we have also generated *Tg(ubb:Avi-Cerulean-GgRanGAP)*^{ox113} and showed that expression as well as localisation of Avi-Cerulean-GgRanGAP was highly similar to previously described *actb2* versions (Figure 2.2).

Next, we expanded our toolkit to include the **T**ranslating **R**ibosome **A**ffinity **P**urification (TRAP) method by Myriam Heiman *et al.* for purification of polysomal

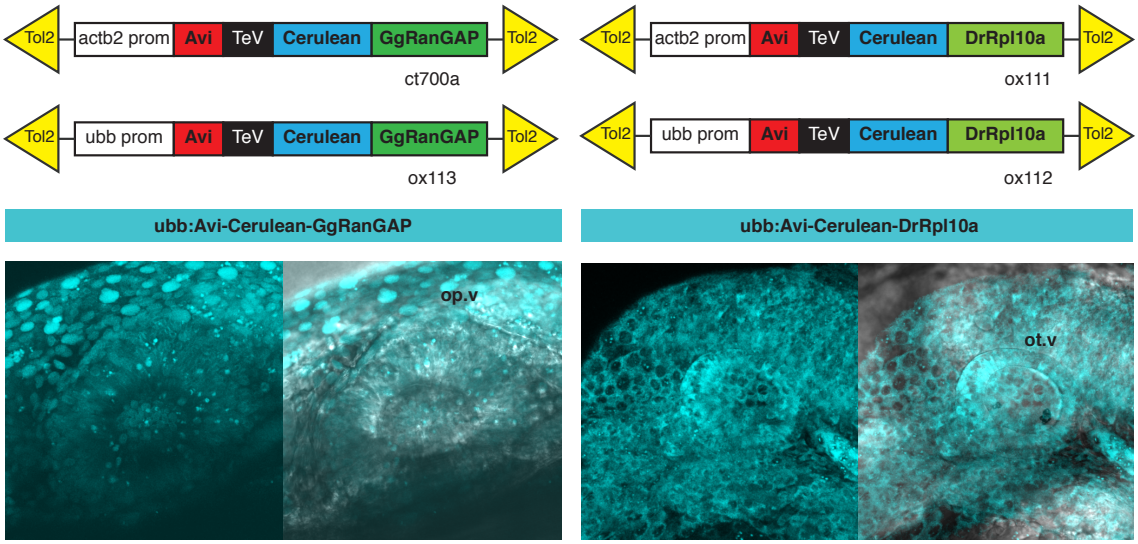


FIGURE 2.2: *Avi effectors* for subcellular compartmentalisation.

Schematic representation of *Avi effector* transgenes with corresponding *ox* or *ct* allele names, for expressing Avi-Cerulean-GgRanGAP and Avi-Cerulean-DrRpl10a fusion proteins in zebrafish. Each cassette is flanked with *tol2* sequences (yellow triangles) for genomic integration and generation of stable transgenic lines. Confocal imaging of F_1 embryos shows the distinctive pattern of expression for Avi-Cerulean-GgRanGAP (left; nuclear envelope) and Avi-Cerulean-DrRpl10a (right; cytosolic, nucleoplasm excluded except for nucleoli puncta within the nuclei). Labelled embryonic structures are optic vesicle (left; op.v) and otic vesicle (right; ot.v).

complexes from cytosolic lysates and their associated mRNA transcripts (Heiman et al., 2008). In TRAP, GFP antibody is used to target and isolate eGFP-Rpl10a fusion proteins expressed in transgenic animals or in cell culture (Heiman et al., 2014; Tryon et al., 2013). We adapted TRAP for *in vivo* biotinylation by AviTagging the zebrafish 60S large ribosomal subunit, *rpl10a* (Tryon et al., 2013), and building stable transgenic lines that ubiquitiously express this fusion protein under the control of either *actb2* or *ubb* promoter as previously described - *Tg(actB2:Avi-Cerulean-DrRpl10a)*^{ox111} and *Tg(ubb:Avi-Cerulean-DrRpl10a)*^{ox112}. Consistent with Tryon *et al*. 2013, confocal imaging confirmed expression of the Avi-Cerulean-Rpl10a fusion protein in the cytosol with discrete puncta within the nuclei, most likely representing accumulation of Avi-Cerulean-Rpl10a in the nucleoli where ribosome biogenesis occurs (Figure 2.2).

2.2.2 *Avi effector* chromatin regulators for epigenomic profiling

As part of a collaboration with U. Senanayake, we sought to exploit the power of *biotagging* to perform tissue-specific biotin-ChIP on selected chromatin regulators in zebrafish. Five chromatin regulators - *p300b*, *ring1b*, *dnmt3ab* (also known as *dnmt6*), *kdm1a* (also known as *lsd1*), *chd7* - were AviTagged and transgenic lines were generated by U. Senanayake. Citrine-Avi-TeV cassette was recombineered using 50 bp homology arms into the start codon of *p300b*-, *dnmt3ab*-, or *ring1b*-containing BACs for transgenesis (see Chapter 6 Materials and Methods). Using CRISPR/Cas9-mediated homology-directed repair (HDR), the endogenous *chd7* and *kdm1a* loci were AviTagged by replacing their start codons with an AviTag recombination cassette. *Kdm1a* was AviTagged using a donor vector containing a Citrine-Avi cassette with long homology arms (1 kb). Unlike *kdm1a*, *chd7* was AviTagged using a ssDNA oligo containing Avi without a Citrine reporter with microhomology arms (30 bp) (Figure 2.3A).

Previous studies have demonstrated the feasibility of transgene knock-in in the zebrafish using CRISPR/Cas9-mediated HDR (Hisano et al., 2015; Auer et al., 2014). We co-injected Cas9 protein, pre-loaded with the guide RNA targeting the knock-in genomic site, together with a donor construct or a single-stranded AviTag™ HDR oligo into one-cell stage zebrafish embryos. We identified F_0 founders producing F_1 transgenic embryos harbouring AviTagged targeted loci, thus demonstrating successful germline transmission of the desired genetic modification. Polymerase Chain Reaction (PCR) followed by Sanger sequencing of the modified genomic region, for both *kdm1a* and *chd7*, in genotyped F_1 embryos confirmed error-free and in-frame insertion of the AviTag™ which replaced the start codon of the first exon for each gene (data not shown).

Next, to insure that AviTag is transcribed in frame with the endogenously tagged chromatin regulator, we probed this genetic modification at the RNA level. Total

RNA was collected from genotyped F_1 embryos and poly(A)-selected libraries prepared for RNA-seq. The genetically-modified transcript was computationally extracted using a Trinity-based custom pipeline (see Chapter 6 Material and Methods) for further examination.

(Please refer to Figure 2.3 on the next page.)

A Citrine-Avi-TeV cassette was recombineered using 50 bp homology arms (bold black lines) into the start codon of *p300b-*, *dnmt3ab-*, or *ring1b*-containing BACs. Each BAC contains *tol2* sequences (yellow triangles) for genomic integration of recombineered loci and generation of stable transgenic lines. *AmpR* gene (blue) is used for bacterial propagation and selection of each BAC. Using CRISPR/Cas9-mediated homology-directed repair (HDR), the endogenous *chd7* and *kdm1a* loci were AviTagged by replacing their start codons with an AviTag recombination cassette. *Kdm1a* was AviTagged using a donor vector containing a Citrine-Avi cassette with long homology arms (1 kb). *Chd7* was AviTagged using a ssDNA oligo containing Avi without a Citrine reporter with microhomology arms (30 bp). **B** The genetically-modified transcript for *chd7* showing A to G substitutions (red bases) within the AviTag™. Similar substitutions were also present in the flanking 5'UTR and first exon. This phenomenon was not observed with the BAC AviTagged chromatin regulator Trinity-assembled transcripts (not shown). **C** Biotin-ChIP of p300b was performed in *tcf21*-expressing cells. Tcf21 NLS-*BirA driver* and Avi-p300b *effector* were crossed and 5 dpf embryos collected for the biotin-ChIP experiment. Genome Browser screenshot shows genomic locus containing *tcf21* gene (*tcf21* gene labelled in light purple) with following tracks - Tcf21 ATAC (ATAC-seq from 5 dpf FAC-sorted *tcf21*-expressing cells), Tcf21/p300b ChIP (genomic DNA-binding map of Streptavidin-pulled down Avi-p300b biotinylated by NLS-BirA in *tcf21*-expressing cells), Tcf21/p300b input (genomic DNA fragments before Streptavidin-pulldown of Avi-p300b) and Sox10 ATAC (ATAC-seq from non-*tcf21*-expressing cells, for comparison).

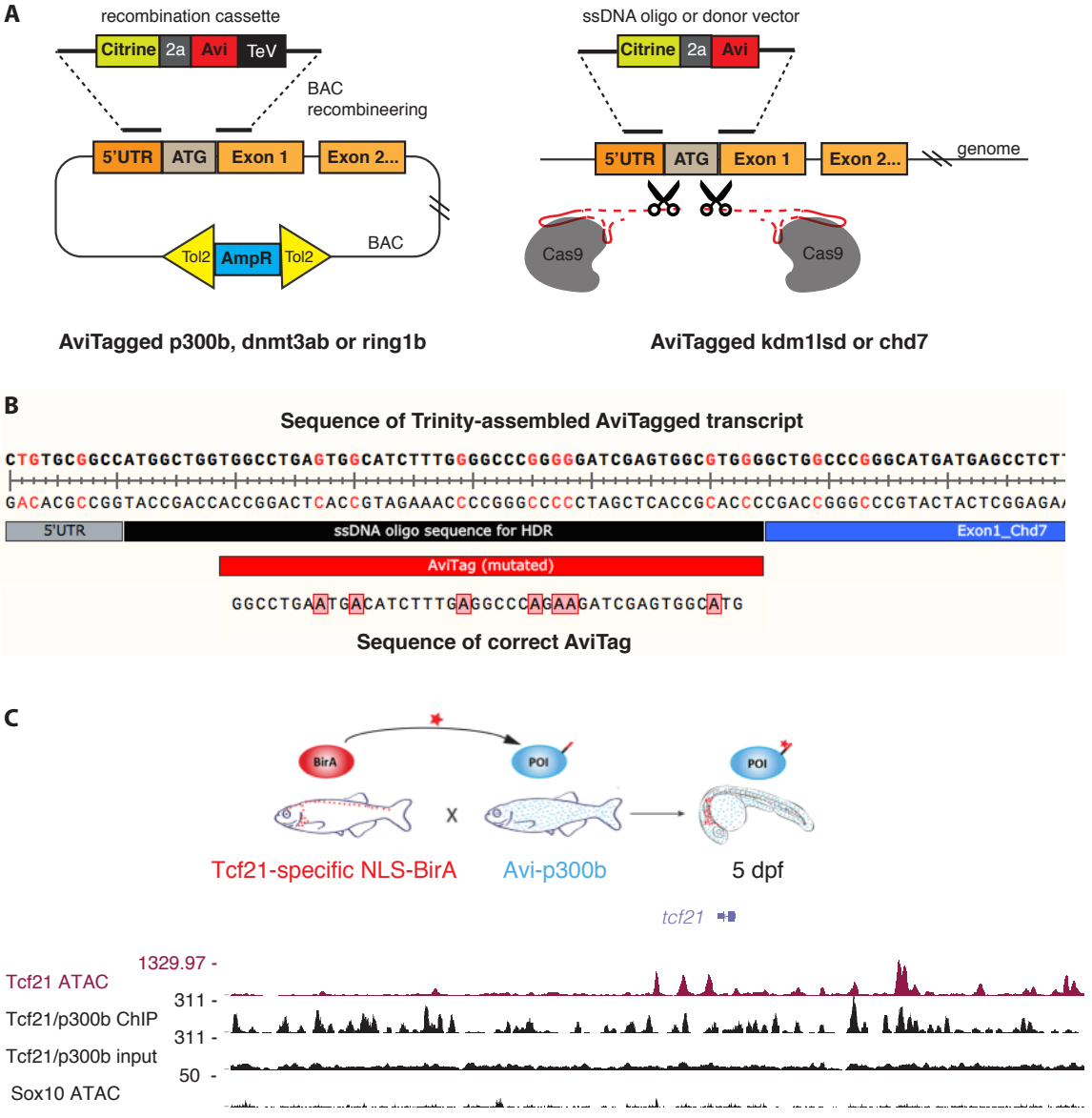


FIGURE 2.3: *Avi effector* chromatin regulators for epigenomic profiling. Full legend is shown on previous page.

We confirmed the presence of transgene mRNA transcripts with accurate sequence and splicing for all three BAC transgenic lines (*p300b*, *dnmt3ab*, *ring1b*), but could not detect the AviTagged *kdm1a* transcript generated by CRISPR/Cas9- mediated HDR. We reasoned this was likely due to low levels of expression for this gene, as only ~400 copies of the endogenous non-modified transcript could be detected. For CRISPR/Cas9-mediated HDR AviTagging of *chd7*, we successfully detected both the AviTagged and endogenous transcript. Interestingly, we observed

multiple G to A substitutions within the 5' untranslated region (UTR) and first exon flanking ± 40 bp of the AviTag™, as well as within the AviTag™ sequence itself (Figure 2.3B). Given that the BAC transcripts for *p300b*, *dnmt3ab* and *ring1b* assembled using the same pipeline were error-free, these observations suggest an RNA editing event occurring as a result of, or triggered by, the presence of an AviTag™ at the *chd7* start codon for reasons unknown. For practical considerations, this phenomenon would render the mutated AviTag™ protein to be non-functional for biotin-ChIP experiments. However, using our optimised biotin-ChiP protocol, we successfully obtained a genome-wide binding map of Avi-p300b in the developing epicardium, using a *tcf21* NLS-BirA (nuclear-localised BirA, see below) driver (Figure 2.3C).

2.3 Establishment of *BirA driver* transgenic lines

To complement our *Avi effector* transgenic lines described above, we designed two types of *BirA driver* transgenic lines - (1) tissue-specific *BirA drivers* which can either be crossed to ubiquitous or tissue-specific *Avi effector* lines to achieve biotinylation in a given cell type subpopulation and (2) ubiquitous *BirA driver* lines to be crossed with tissue-specific *Avi effector* transgenic lines (e.g. AviTagged TFs) when biotinylation of specific/endogenously driven regulatory factors is required.

2.3.1 Ubiquitous and tissue-specific *BirA drivers*

Two constructs were cloned and used to generate stable lines that ubiquitously express BirA, using *tol2*-mediated transgenesis (Kawakami, 2004) - *Tg(ubb:BirA-2a-mCherry)*^{ox115} and *Tg(ubb:NLS-BirA-2a-mCherry)*^{ox114} i.e. cytoplasmic BirA and nuclear-localised (via a **N**uclear **L**ocalisation **S**ignal peptide, NLS) BirA, respectively, under the control of the zebrafish *ubb* promoter. The NLS-*BirA driver* line is crossed to *Avi effector* chromatin regulator and transcription factor lines to enable biotinylation of corresponding AviTagged target proteins in the nucleus. Cytoplasmic *BirA driver*

is used to biotinylate *Avi effectors* for subcellular compartments as described in the previous section (Avi-Cerulean-GgRanGAP and Avi-Cerulean-Rpl10a).

Two other lines that have already existed in the lab and were used for experiments in this chapter are *Tg(actb2:BirA-2a-membCherry)*^{ct709a} (cytoplasmic BirA with membrane-localised mCherry fluorescent reporter expressed under the control of *actb2* promoter) and *Tg(sox10:BirA-2a-membCherry)*^{ct706a} (cytoplasmic BirA with membrane-localised mCherry fluorescent reporter expressed under the control of a *sox10* promoter) (Figure 2.4).

Using BAC transgenesis, we have also created two neural crest-specific *BirA driver*s that express BirA or NLS-BirA under the control of a *sox10* locus-containing BAC (DKEY-201F15) - *TgBAC(sox10:BirA-2a-mCherry)*^{ox104a} and *TgBAC(sox10:NLS-BirA-2a-Citrine)*^{ox108a} (Figure 2.4). BAC transgenesis has several advantages over conventional transgenesis that uses proximal regulatory elements to drive expression of a transgene. In most cases, a BAC recapitulates the entire endogenous genomic landscape of a predefined locus, and transgene expression is driven by all the regulatory elements present in the selected locus assuring expression in an endogenous-like fashion. Due to the larger size of BAC alleles, integration events are rarer and therefore more likely to be singular in identified founders while conventional transgenesis usually results in multiple integrations of the transgene construct.

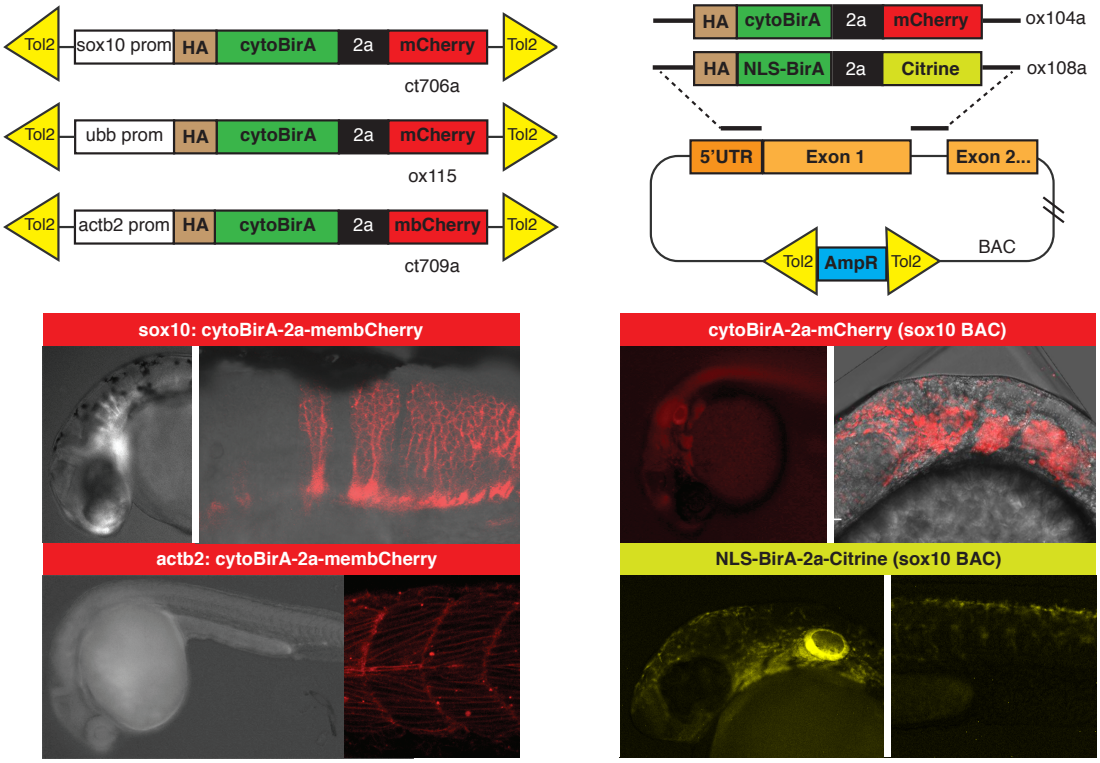


FIGURE 2.4: *BirA drivers* for biotinylation of *Avi effectors*.

Schematic representation of *BirA driver* plasmid or BAC-recombineered transgenes with corresponding *ox* or *ct* allele names, for expressing HA-tagged BirA (with fluorescent reporter) fusion proteins in zebrafish. Each cassette is flanked with *tol2* sequences (yellow triangles) for genomic integration and generation of stable transgenic lines. *Sox10*-BAC driven expression of BirA cassettes has a highly reproducible pattern in the cranial regions (right). Close-up of mCherry expression in the branchial arches is shown for cytoBirA-2a-mCherry. Citrine expression pattern in the trunk is shown for NLS-BirA-2a-Citrine.

2.4 Establishment and refinement of the *biotagging* method

To date, several established protocols for *in vivo* biotinylation have been published for samples from cell culture (Heiman et al., 2014), brain tissue from rodents (Heiman et al., 2008) and plant tissue (Deal and Henikoff, 2011). The use of INTACT for chromatin, proteomic and gene expression studies has been demonstrated in the *Xenopus*, *Drosophila* and *C. elegans* (Ooi et al., 2009; Steiner et al., 2012; Amin et al., 2014). Furthermore, a study from the Stainier lab utilising *in vivo* biotinylation TRAP in

zebrafish was also published while our *biotagging* paper was in revision, but the method employed was suboptimal as it required biotin supplementation of the embryos for efficient biotinylation of the Avi-Rpl10a fusion protein (Housley et al., 2014). Here, we have optimised protocols for robust isolation of *in vivo* biotinylated target proteins (Avi-Cerulean-DrRpl10a and Avi-Cerulean-GgRanGAP) in the zebrafish, without biotin supplementation but with high efficiency.

2.4.1 Specificity and efficiency of *biotagging*

To test the specificity and limits of our *biotagging* method, we performed a series of Western Blots (see Chapter 6 Materials and Methods). First, to compare the efficiency of biotinylation with varying levels of BirA expression, *Tg(ubb:Avi-Cerulean-DrRpl10a)^{ox112}* were crossed with *Tg(sox10:BirA-2a-membCherry)^{ct706a}* or *TgBAC(sox10:BirA-2a-mCherry)^{ox104a}* and protein lysates were isolated from 24 hpf (hours post fertilisation) embryos. At 24 hpf, membCherry expression in *Tg(sox10:BirA-2a-membCherry)^{ct706a}* was detectable by fluorescence only in the cranial regions of the embryo. In contrast, mCherry expression in *TgBAC(sox10:BirA-2a-mCherry)^{ox104a}* was detectable by fluorescence in a clear endogenous *sox10*-like fashion - not only in the cranial regions, but also in an anteroposterior pattern along the dorsal neural tube. Despite this difference in levels of fluorescent reporter (and BirA) expression, biotinylation levels were fairly equal in both conditions (Figure 2.5A).

Next, we varied the levels of *Avi effector* expression by crossing *TgBAC(sox10:BirA-2a-mCherry)^{ox104a}* with *Tg(ubb:Avi-Cerulean-DrRpl10a)^{ox112a}* (strong expression, bright fluorescence) and *Tg(ubb:Avi-Cerulean-DrRpl10a)^{ox112b}* (weak expression, faint fluorescence). Strikingly, we clearly detected a decrease in biotinylation with weak Avi-Cerulean-Rpl10a expression (Figure 2.5A). Taken together, these results demonstrated that biotinylation efficiency is sensitive to expression levels of the *Avi effector* but not the *BirA driver*, and should be taken into account when planning *biotagging* experiments (see Discussion). In all cases, however, biotinylation was not detected

when either *Avi effector* or *BirA driver* was present alone. This demonstrated the specificity of this system (Figure 2.5A).

To quantify the efficiency of *biotagging*, we compared the recovery of cell numbers per 100 embryos from *biotagging*-RanGAP, *biotagging*-Rpl10a and **F**luorescence **A**ctivated **C**ell **S**orting (FACS) by quantifying total RNA yields. These were quantified and expressed as mass per cell (nucleus in the case of INTACT biotagging) as follows; 1 pg/cell for nuclei (Deal and Henikoff, 2011), 0.05 pg/cell for polysomes (Heiman et al., 2014) and original cell counts from FACS experiments. Two biological replicates per sample type were used, and significance calculated using Student's *t*-test (one-tailed, two-sample equal variance). The replicates were also used to construct and sequence corresponding RNA-seq libraries (see Chapter 6 Materials and Methods). The results demonstrated a $\sim$5 to 7 fold increase in yields in *biotagging* experiments as compared to the FACS approach, highlighting a significant increase in efficiency using *biotagging* approach (Figure 2.5B).

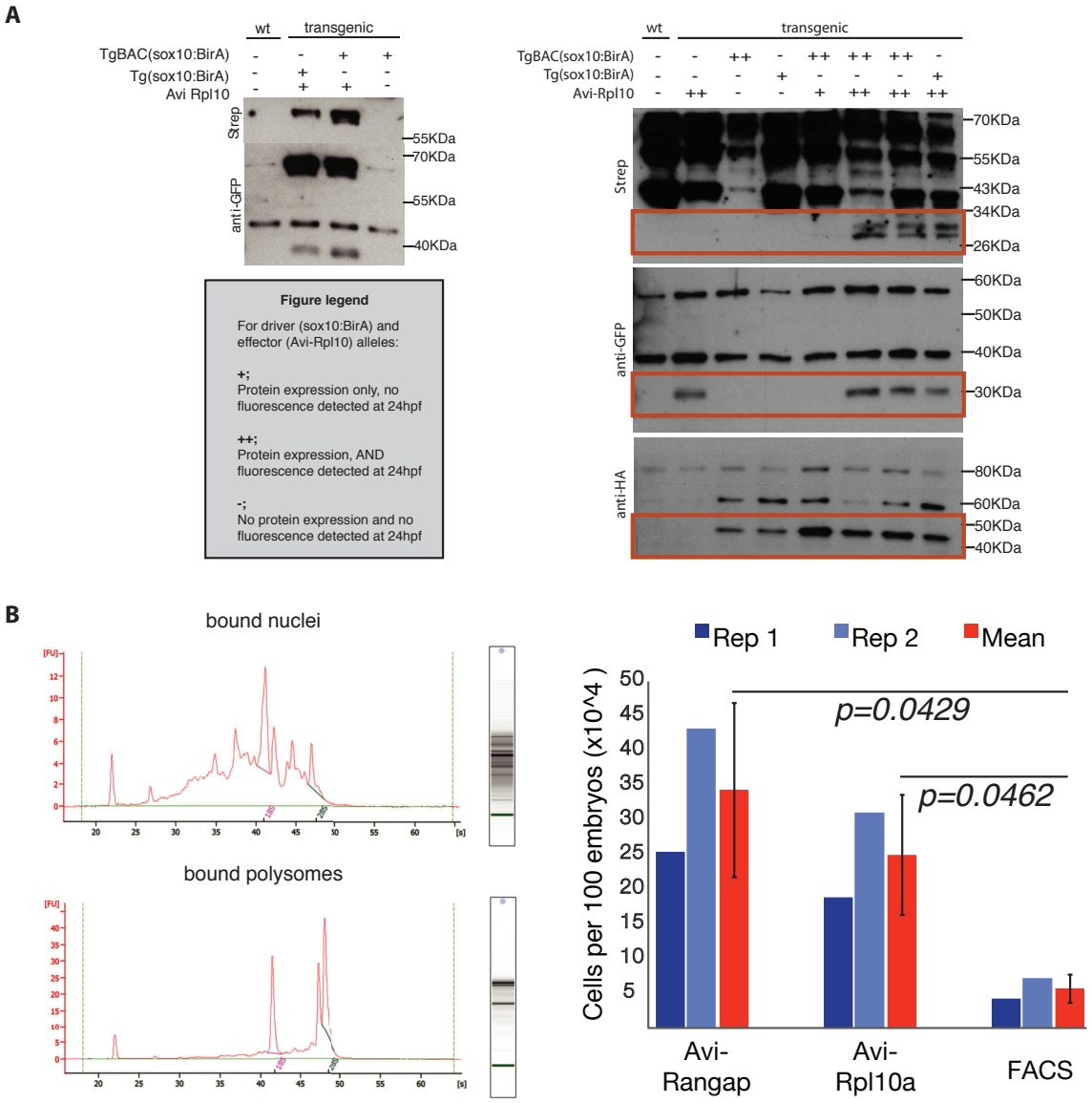


FIGURE 2.5: Specificity and efficiency of the *biotagging* method.

A Western blots demonstrating similar levels of biotinylation when *Tg(sox10:BirA-2a-membCherry)^{ct706a}* or *TgBAC(sox10:BirA-2a-mCherry)^{ox104a}* were used as weak or strong, respectively, *BirA drivers* (left). On the other hand, a decrease in biotinylation is detected with weak *Avi effector*, regardless of *BirA driver* strength (right; biotinylated protein is highlighted in red boxes).
B Total RNA extracted from Streptavidin-bound nuclei and polysomes show clearly distinguishable Bioanalyzer profiles, demonstrating subcellular resolution obtained via *biotagging* (left). Quantification of RNA yields from *biotagging* samples show statistically significant higher (∼5 to 7 fold) yields, highlighting increased efficiency compared to FACS samples. Cell numbers per 100 embryos were quantified as follows; 1 pg/cell for nuclei (Deal and Henikoff, 2011), 0.05 pg/cell for polysomes (Heiman et al., 2014) and original cell counts from FACS experiments, from two biological replicates per condition.

2.4.2 Experimental conditions for successful nuclear or polysomal isolation

To achieve maximal efficiency for the *biotagging* method, we optimised numerous variables in the original biotagging pull-down protocol (Trinh et al., 2017). The original protocol produced promising initial results, but we observed trace amounts of RNA even when wild type embryos, expressing no biotagging components, were subjected to the *biotagging* procedure. This suggested suboptimal lysis and pull-down conditions resulting in background material being bound to Streptavidin beads used in our experiments. We first reasoned that this was due to residual cell-to-cell contacts via connective tissue fibres still remaining after the physical homogenisation step, which led to cell surface proteins still persisting and binding to the Streptavidin beads. Unfortunately, treatment of the lysates with elastase did not address the background issue. Next, we increased the number of pestle strokes using Dounce homogeniser, both during the first (pestle A) step aimed at obtaining single cell suspension as well as during the second (pestle B) step aimed at disrupting cell membranes and releasing nuclei/cytoplasmic contents. This modification yielded a slight improvement, suggesting that the incomplete cell lysis could be at the origin of the observed non-specific binding to the Streptavidin beads. Thus, we next optimised a condition in which wild type embryos were incubated and homogenised with excess amount of hypotonic buffer (see Chapter 6 Materials and Methods), to account for the changes in the salt concentration of the buffer during the cell lysis process. This change insured that the cells remained in the hypotonic environment during the entire lysis and pull-down procedure, facilitating the membrane disruption/bursting and collection of intracellular "biotagged" organelles. When wild type embryos were lysed in this fashion and incubated with Streptavidin beads, we no longer detected any traces of RNA observed with the previous protocol, suggesting that we obtained complete lysis of the cellular membranes and the release of the cellular contents into the lysate. This determinant proved to be the simplest and

most effective change to our isolation protocol, producing the finalised version of our *biotagging* method (see Chapter 6 Materials and Methods). We were able to obtain consistent results with the subcellular resolution needed to clearly distinguish between nuclear and polysomal total RNA profiles, differentiated by their rRNA ratios (lower ratio of rRNA in nuclear RNA compared to ∼90% in polysomal RNA) (Figure 2.5B).

2.5 Genome-wide profiling of the cardiomyocyte nuclear transcriptome

As proof-of-principle of the *biotagging* method, we profiled the nuclear transcriptome of developing cardiomyocytes at 26 hpf. A myocardial *BirA driver*, *Tg(myl7:BirA-2a-membCherry)*^{ct704a}, was crossed to the nuclear envelope effector *Tg(actB2:Avi-Cerulean-GgRanGAP)*^{ct700a}. Biotinylated nuclei were isolated as per optimised procedure and the total RNA extracted was used for preparation of NGS libraries using ribo-depletion (RiboZero) which were subsequently sequenced using Illumina platform (see Chapter 6 Material and Methods). This dataset was compared to a stage-matched control obtained by crossing a ubiquitous *BirA driver*, *Tg(actb2:BirA-2a-membCherry)*^{ct709a}, to the same *Avi effector* i.e. *Tg(actB2:Avi-Cerulean-GgRanGAP)*^{ct700a}. These datasets are referred from henceforth as "*myl7* nuclei" and "*actb2* nuclei".

2.5.1 Enrichment of myocardial molecular signature by *biotagging*

Differential expression analysis revealed 3,715 significantly more abundant and 3,035 less abundant genes in *myl7* nuclear transcriptomes when compared to *actb2* nuclei. ZFIN-annotated myocardial markers such as *acta1b*, *tbx1*,*sox9b*,*tbx20*, *gata3*, *myh6*, and *gata4* were amongst the genes found more abundant in *myl* nuclei (p<0.05) (Figure 2.6A). Statistical over-representation analysis of the higher abundance gene set using PANTHER functional annotation tool revealed enrichment (p<0.001) of **G**ene

Ontology (GO) terms for biological processes related to myocardium development - muscle contraction, muscle organ development and mesoderm development (Figure 2.6B).

Furthermore, gene set analysis (GSA) of the higher abundance genes using **P**latform for **i**ntegrative **an**alysis of **o**mics data (Piano) (Väremo et al., 2013) revealed the presence of signalling pathways implicated in myocardium development and function. The largest and most statistically significant GSA node ($p\sim0$) consisted of 76 Wnt pathway genes, consistent with a large body of evidence in the literature on the importance of Wnt signalling during cardiac development and differentiation (Gessert and Kühl, 2010). A smaller, but also statistically highly significant GSA node ($p\sim0$) consisted of 6 Hedgehog (Hh) signalling genes, also known to be important for cardiac development (Thomas et al., 2008) (Figure 2.6C). Last but not least, DESeq2 differential expression analysis between *myl7* nuclei and *sox10* nuclei (using TgBAC *sox10 BirA driver*) highlighted the clear divergence between these two cell populations, with myocardial markers clearly less abundant in *sox10* nuclei (Figure 2.6D).

This proof-of-concept experiment using the developing myocardium demonstrated both the utility of *biotagging* in obtaining tissue-specific molecular signature, as well the exquisite specificity of its approach. Our results highlight *biotagging*'s prospects in uncovering new biology during embryonic development, which will be discussed in the next chapter.

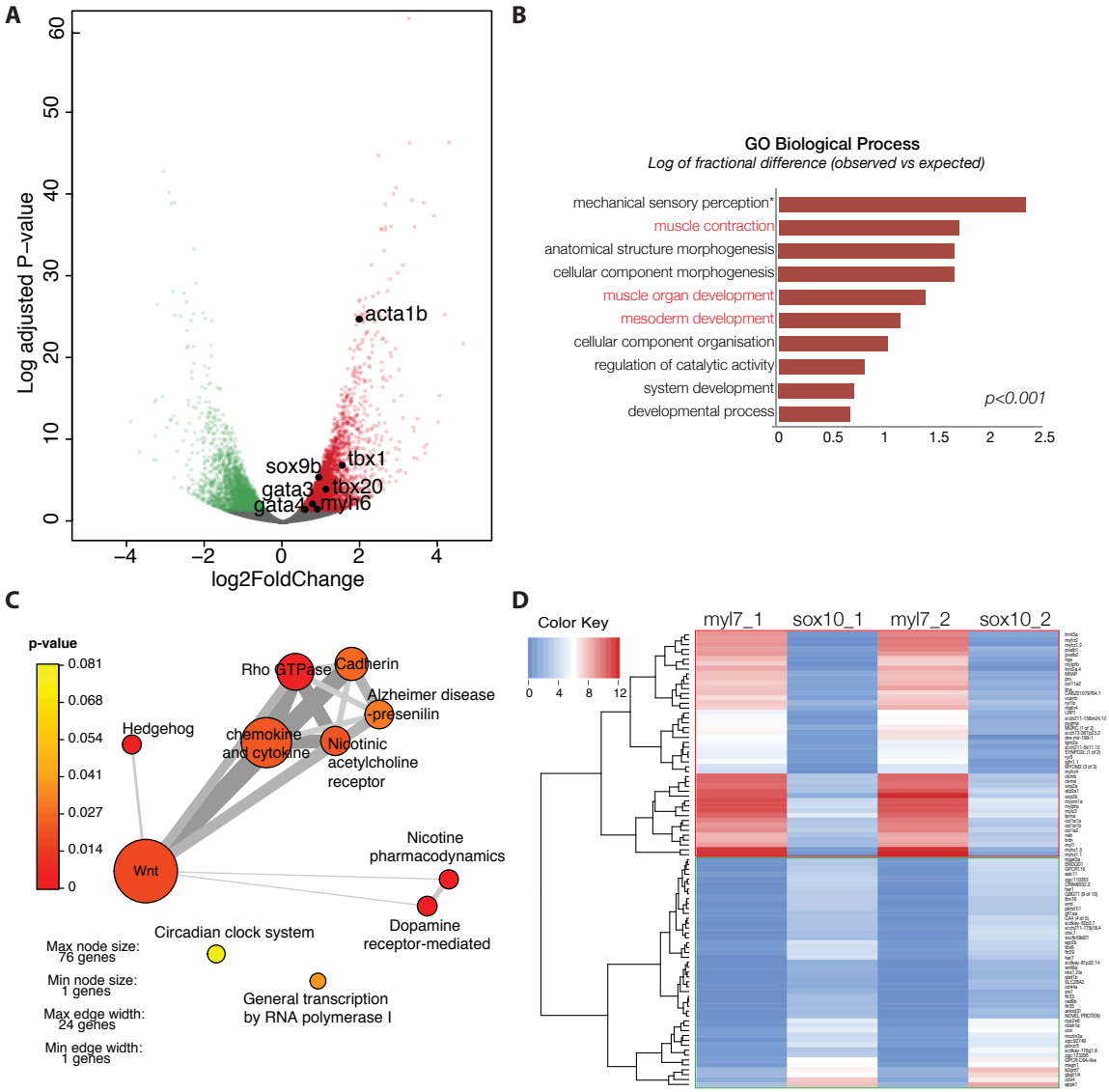


FIGURE 2.6: Enrichment of myocardial molecular signature by *biotagging*.

A Scatterplot of upregulated (red) and downregulated (green) genes in *myl7* nuclei compared to stage-matched *actb2* nuclei.
B Statistical over-representation analysis by Panther of the upregulated gene list. GO terms related to myocardium development highlighted in red.
C Network plot from gene set analysis (GSA) using Piano showing the presence of signalling pathways implicated in myocardium development and function. *P*-value of each GSA node is presented in yellow-orange colour key.
D FPKM (**F**ragments **P**er **K**ilobase of transcript per **M**illion mapped reads) heatmap of top 50 differentially upregulated genes in *myl7* nuclei (top half) and *sox10* nuclei (bottom half) demonstrates stark divergence of these two cell types from each other. Log2-fold enrichment is presented in blue-red color key.

2.6 Discussion

To meet the need in zebrafish studies that aim to analyse specific cell populations at the genome-wide level, we have introduced the *biotagging* method. In this chapter, we have described the successful establishment of the *biotagging* toolkit with plasmid constructs, transgenic lines and optimised biochemical protocols. *Biotagging* not only offers tissue-specificity, but also enables the analysis of transcriptomes at the subcellular resolution.

2.6.1 The finer details of *biotagging*

Balance between *Avi effector* and *BirA* levels. Binding of streptavidin to biotin is highly efficient, although the total number of interactions in a given space may be limited due to steric hindrance (Buranda et al., 1999). In theory, it is essential to have a critical mass of biotinylated AviTagged cellular compartments for efficient binding and subsequent pulldown by streptavidin. We demonstrated that biotinylation efficiency is sensitive to *Avi effector* protein levels but not to *BirA*, highlighting a point of consideration for determination of Avi-BirA combinations in *biotagging* experiments. The biotinylation reaction was shown to be rate-limited by the protein-protein interaction between biotin ligase and its substrate (in our case, the AviTag^{TM}) (Ingaramo and Beckett, 2011). This presented a possible explanation for our observation - given the size difference between the two proteins ($\sim$40 kDA BirA and $\sim$2 kDa AviTag^{TM}), multiple fold quantity of AviTag^{TM} is needed for every molecule of BirA to maximise chances of physical interaction, at least within the context of nuclei and polysomal biotinylation by cytoplasmic BirA. Our finding that low levels of BirA are sufficient to efficiently biotinylate its target is an added strength of the *biotagging* system since it is not uncommon for cell type-specific markers to have low levels of expression. Therefore, BAC *BirA driver* lines can be preferentially generated to exploit their characteristic to reproducibly express transgenes (e.g. expressing BirA in macrophages using *mpeg* BAC (Kenyon et al., 2017)). However, we have

also found that biotin-ChIP utilising AviTagged chromatin regulators (see below) was more tolerant to lower levels of *Avi effector* protein, probably reflecting differences in turnover and/or intracellular trafficking between nuclear-localised proteins and organelle proteins as well as interaction dynamics with nucleus-targeted BirA (NLS-BirA; see below) that we have yet to completely elucidate.

Subcellular tethering of BirA. We have built two versions of BirA - cytoplasmic (non-targeted) (cytoBirA) and nuclear-targeted (NLS-BirA). Given that active polysomal complexes are located in the cytosol and that RanGAP1 is localised to the cytoplasmic face of the RanBP2/Nup358 nucleoporin complex (Matunis et al., 1996), we reasoned that cytoBirA would be the logical choice for their efficient biotinylation. NLS-BirA would be an unfavourable choice, given the high efficiency of the nuclear localisation signal protein to mediate translocation of its tethered protein into the nucleus (Kalderon et al., 1984). Conversely, to obtain sufficient critical quantity of biotinylated target proteins that exert their functions within the nucleus (TFs, chromatin regulators), NLS-BirA is the ideal choice to maximise levels of biotinylation. Hence to meet the needs for biotinylation in both nuclear and cytoplasmic sites using a single *BirA driver*, we have built a third version of BirA tethered to the endoplasmic reticulum (BirA-ER) (Howarth et al., 2008). This ubiquitious *BirA driver* line *Tg(ubb:IgK-BirA-ER-2a-mCherry)*^{ox116} biotinylates target proteins undergoing protein transport and folding within the ER network before translocation to their respective subcellular compartments.

Implications of subcellular movements of RanGAP1. We have cloned and tested two constructs encoding N- or C-terminally AviTagged carboxyl-terminal domain of the avian RanGAP1 protein from *G. gallus* ("GgRanGAP"). In vertebrates, RanGAP1 associates with RanBP2/Nup358 nucleoporin of the nuclear pore complex (NPC), where this interaction is dependent on ubiquitin-like SUMO-1 post-translational modification at its carboxyl-terminal by RanBP2 (Matunis et al., 1996; Mahajan et al.,

1997; Matunis et al., 1998; Pichler et al., 2002). During mitosis, RanGAP1 is phosphorylated (also at the carboxy-terminal) by Cdk1 prior to nuclear envelope breakdown (Swaminathan et al., 2004). RanGAP1/RanBP2 then co-localise at kinetochores of the mitotic spindle, an interaction that also requires SUMO-1 modification of RanGAP1 (Joseph et al., 2002). The importance of this close relationship between the carboxyl-terminal domain for RanGAP1 and its "SUMOlator" RanBP2 for appropriate RanGAP1 localisation throughout the cell cycle cannot be discounted. By using only the carboxyl-terminal of RanGAP1, we bypassed the need to use the full-length RanGAP1 protein which would interfere with development, given that overexpressing Avi-RanGAP1 led to developmental defects and embryonic lethality (data not shown). Using a truncated RanGAP1 protein could have also interfered with its localisation to kinetochores if phosphorylation (but not SUMO-1 modification) was affected, thus favouring NPC localisation.

2.6.2 Biotin-ChIP of chromatin regulators in defined cell types

We sought to exploit the power of *biotagging* to perform tissue-specific biotin-ChIP on selected chromatin regulators in zebrafish - *p300b*, *ring1b*, *dnmt3ab* (also known as *dnmt6*), *kdm1a* (also known as *lsd1*), and *chd7*. This allows us to study chromatin dynamics of developmental processes in specific cell populations, by looking at different types of regulation mediated by histone/DNA modifiers. Recently this method was demonstrated in the mouse as "bioChIP-seq", where AviTagged Ep300 was used to map enhancers in a tissue-specific manner within the developing heart and the forebrain (Zhou et al., 2017). However, the authors reported that several million cells were needed as starting material, which is similar to "standard" antibody-based ChIP. Our version requires only ∼100,000 cells, thus surpassing their efficiency by at least 10-fold.

The molecules selected for AviTagging were chosen due to their implications in modulating the chromatin landscape within the context of gene regulatory elements,

and in particular at enhancers. LSD1 associates with both enhancers and promoters and plays a "repressive" role by "erasing" H3K4 mono-/dimethylation at enhancers. DNMT6 is a *de novo* DNA methyltransferase involved in DNA methylation that leads to gene repression. P300 is implicated in enhancer activation by "writing" (adding) H3K27 acetyl mark at enhancers. CHD7, a DNA helicase, binds promoters but has also been shown to co-localise with P300 and PBAF at enhancer regions, where its interaction with PBAF is of importance in regulating NC gene expression (Bajpai et al., 2010; Schnetz et al., 2010). Last but note least, Ring1b is a tethering protein and catalytically active component of Polycome Repressive Complex 1 (PRC1) (Calo and Wysocka, 2013; Chen and Dent, 2014). Taken together, genome-wide binding maps of these remodellers will report on poised/repressive state transitions in regulatory developmental processes being studied. We are currently characterising how these different chromatin regulators associate with the epigenome in migrating NC, by using *TgBAC(sox10:NLS-BirA-2a-Citrine)*^{ox108a} as the *BirA driver*, at the same developmental stage and from the same cell type used to characterise nuclear and polysomal transcriptomes (see next chapter). This will provide a comprehensive picture of both gene expression and epigenome dynamics within a spatiotemporal snapshot, thus elucidating the potential mechanisms of crosstalk between the two biological processes (manuscript in preparation).

Chapter 3

Systems level analysis of the neural crest translatome and nuclear transcriptome

3.1 Current approaches to studying translatomes

3.1.1 Ribosome profiling

The study of translatomes can be traced back to pioneering research by Joan Steitz and Marilyn Kozak from ∼40 years ago. In her seminal paper, Steitz defined the polypeptide codon sequences required for translation initiation by isolating and sequencing ribosome-protected fragments from bacteriophage (Steitz, 1969). Using the same approach from different organisms, Kozak went on to publish a series of papers e.g. (Kozak and Shatkin, 1977) to further elucidate the mechanistic details of translation initiation. This method was recently resurrected by Jonathan Weissman and Nicholas Ingolia. Dubbed *ribosome profiling*, the method involves isolation of ribosome-protected fragments (RPFs) by digestion with *E.coli* RNaseI followed by sucrose-gradient centrifugation or similar techniques to select the ribosome-RPF complexes (Ingolia et al., 2012). When paired with NGS, this approach allows single-nucleotide resolution of RNA sequences directly bound by ribosomes at a given time and space, hence contributing novel insights into the translation process, such as characterisation of translation speed, control of gene expression, codon usage as well as refining the annotation of translated sequences (Ingolia, 2014).

3.1.2 Polysome profiling

Polysome profiling presents the translatome at a different resolution to *ribosome profiling*. The key difference that separates the two is elimination of the RNase digestion step, thus enabling the isolation of full-length mRNA transcripts bound and actively translated by two or more ribosomes in a polysomal complex. While earlier methods also employed a sucrose gradient approach to isolate polysomes and associated mRNAs, the recently-developed TRAP method discussed in the previous chapter was conceived to simplify this step by directly isolating polysomes via affinity purification (Heiman et al., 2014). When last surveyed in 2013, both methods were broadly equal in terms of popularity judged by the number of published studies employing either method (Piccirillo et al., 2014).

In this section, we present the results from the genome-wide *polysome profiling* of active NC translatome at 16-18 ss using the *biotagging* TRAP method we have developed in zebrafish. Current knowledge of migratory NC identity is mainly based on studies focusing on mRNA (*in situ* hybridisation-, qPCR-based studies) or protein (immunohistochemistry-, morpholino knockdown-based studies) expression. Genome-wide profiling of the transcriptome from FAC-sorted neural crest cells provided a more comprehensive view of the gene regulatory network underlying neural crest function by quantifying mRNA transcripts from whole cell lysates in the cranial NC (Simões-Costa et al., 2014). To our knowledge, this is the first comprehensive study that directs its focus at the NC translatome, aimed at providing novel insights into the current understanding of NC function.

3.2 *biotagging*-TRAP-seq of the migrating neural crest

Sox10 [SRY (sex determining region Y)-box 10] is a well-characterised TF with key roles in NC development. In the zebrafish *sox10* mutant, *colourless*, NC cells fail to

migrate and differentation of non-ectomesenchymal derivatives was impaired despite relatively normal NC specification (Dutton et al., 2001). Further studies elucidated the downstream gene regulatory circuits controlled by sox10, showing that sox10 actively regulates differentiation of derivatives including melanocytes (Elworthy et al., 2003), Schwann cells (Finzsch et al., 2010), glial cells (Stolt et al., 2002) and sensory neurons (Carney et al., 2006) (reviewed by (Kelsh, 2006)). At 16-18 ss (somite stage) in the zebrafish, *sox10* prominent expression is detected in the premigratory (cranial and trunk) and migrating (cranial) NC cells, as well as in the otic vesicle (Figure 3.1), where Sox10 plays a role in patterning of otic placode derivatives (Dutton et al., 2009). To investigate the utility of *biotagging*-TRAP-seq for purifying translating mRNA transcripts from migrating NC, we crossed our *sox10* BAC *BirA driver*, *TgBAC(sox10:BirA-2a-mCherry)*^{ox104a}, to *rpl10a Avi effector*, *Tg(ubb:Avi-Cerulean-DrRpl10a)*^{ox112} described in the previous chapter. Biotinylated polysomes from double positive offspring at 16-18 ss were isolated using Streptavidin beads for downstream analyses.

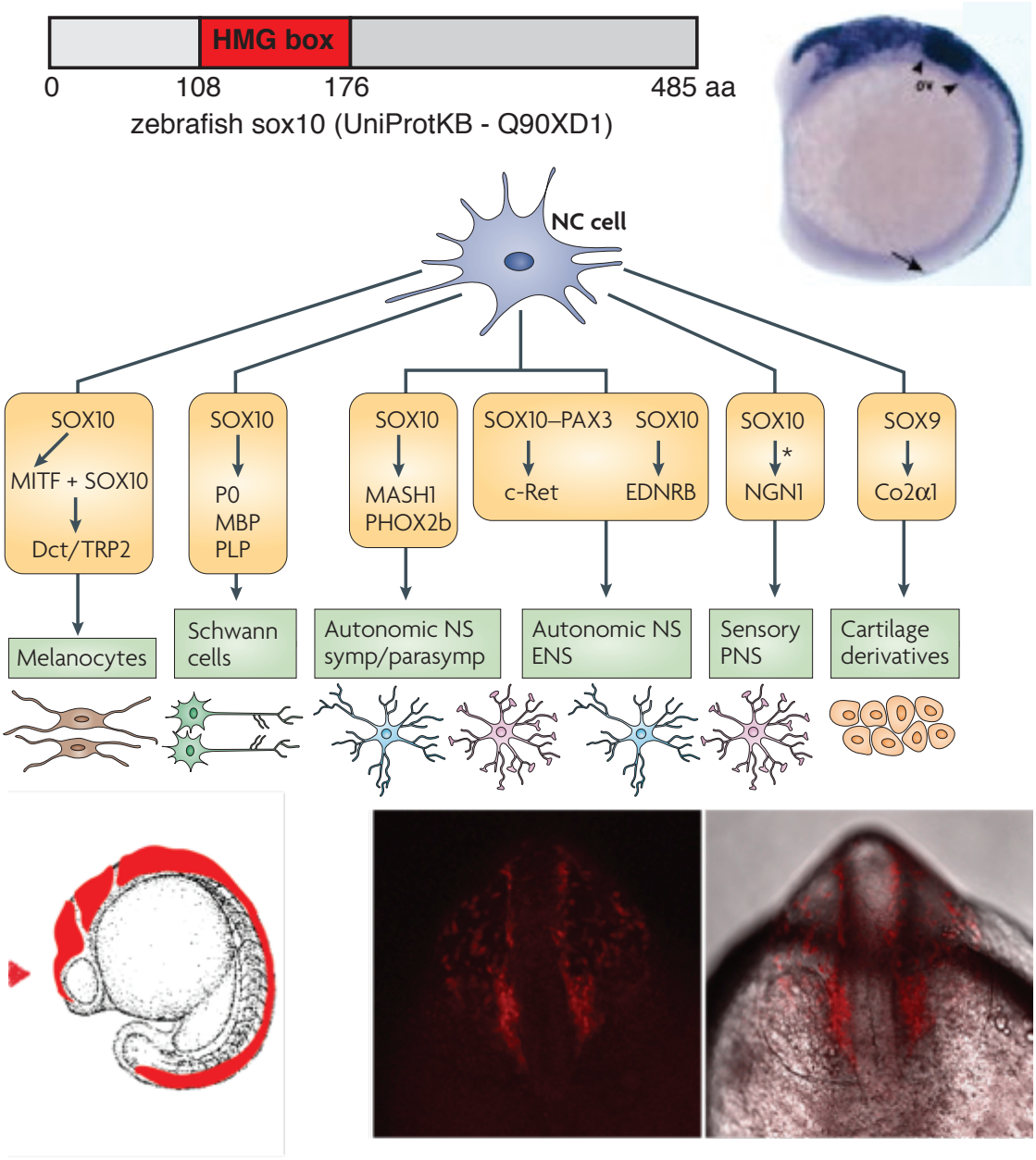


FIGURE 3.1: *Sox10* is a neural crest marker.

The 485-aa zebrafish sox10 protein contains a HMG (High Mobility Group) DNA-binding domain. In the embryo at 10-14 ss, it is expressed in otic vesicle (ov), premigratory neural crest in the trunk and cranial regions, as well as in the migratory neural crest in the cranial regions (*in situ* image from (Montero-Balaguer et al., 2006)). Sox10 functions upstream of non-ectomesenchymal neuroglial derivatives (schematic diagram from (Sauka-Spengler and Bronner-Fraser, 2008)). Migrating neural crest in the cranial region at 16 ss can be visualised with an mCherry reporter in *TgBAC(sox10:BirA-2a-mCherry)*^{ox104a}.

3.2.1 Exploratory data analysis of the neural crest translatome

Total mRNA samples were collected from Streptavidin-isolated *sox10* polysomes (*sox10 bound*) and the remaining lysate after Streptavidin beads incubation (*sox10 lysate*) with two biological replicates each. Libraries were prepared using poly(A) selection for strand-specific deep sequencing (see Chapter 6 Materials and Methods).

Kallisto (Bray et al., 2015) was used to quantify transcript levels of 58,549 Ensembl cDNA sequences annotated in the zebrafish genome (GRCz10/danRer10). Pairwise comparison scatterplots of **T**ranscript **P**er **M**illion (TPM) values between biological replicates of *sox10 bound* demonstrated the high reproducibility of our optimised *biotagging*-TRAP protocol (ρ = 0.86) to isolate polysomal-bound mRNA transcripts. *Sox10 lysate* replicates demonstrated higher variability (ρ = 0.67) which is similar to comparing *sox10 bound* and *sox10 lysate* averaged TPM values (ρ = 0.71) (Figure 3.2A). To arbitrarily determine translated transcripts that were present at the level or higher than *sox10*, we first identified *sox10*'s expression ranking within *sox10 bound* dataset. Since *sox10* was found to be the 6786th most expressed gene, we extracted the top 7000 expressed transcripts of *sox10 bound*. For comparison, the top 7000 expressed transcripts of *sox10 lysate* were identified. We obtained both overlapping (82.1%) and mutually exclusive (17.9%) transcripts between the two populations, demonstrating their divergence and preliminarily suggesting successful isolation of polysome-bound transcripts from the lysate (Figure 3.2B).

Next, we extracted transcript levels of a subset of 261 Ensembl cDNA sequences annotated by Zebrafish Information Network (ZFIN) as being expressed in the NC at 14-19 ss. Heatmap of TPM values successfully confirmed the divergence between these two populations at the level of known NC markers, with *sox10 bound* containing 221/261 transcripts and *sox10 lysate* 167/261 transcripts expressed at TPM>2 (Figure 3.2C). Comparing the two populations revealed that a high proportion (159/261) of NC transcripts were found in both *sox10 bound* and *sox10 lysate*,

while 62/261 NC transcripts were specific to the *sox10 bound* sample and only 8/261 to the *sox10 lysate* sample. Interestingly, the 8 *sox10 lysate*-specific NC transcripts were isoforms of genes present in the shared population between *sox10 bound* and *sox10 lysate*. Moreover, the lack of *sox10 lysate*-specific candidates was also confirmed when the comparison was performed using gene IDs instead of transcript IDs (Figure 3.2C′ and C″).

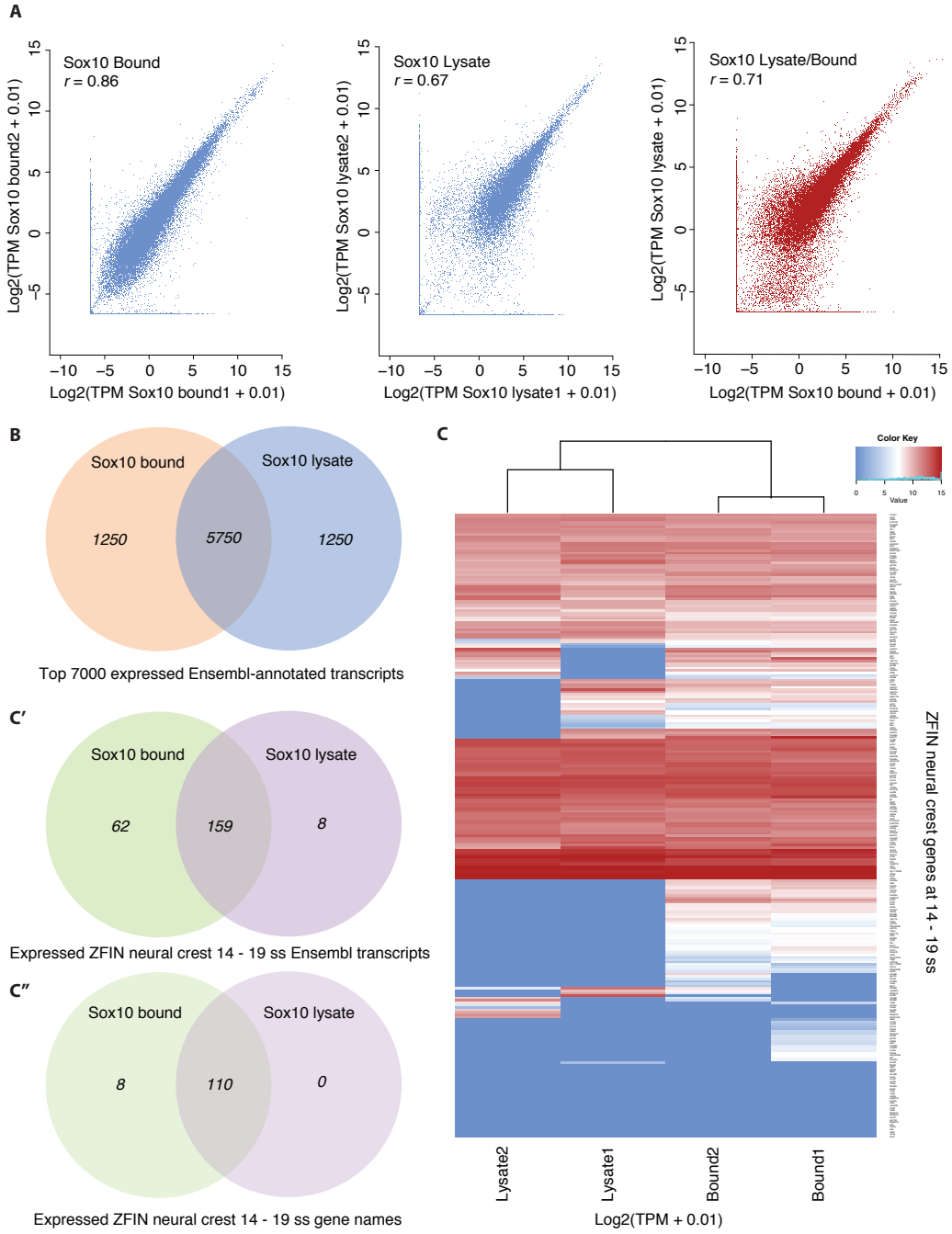


FIGURE 3.2: Exploratory data analysis of the neural crest translatome.

A Pairwise comparison scatterplots with Spearman correlation ρ values demonstrate reproducibility between replicates of *sox10 bound* translatome.
B Venn diagram showing overlap and divergence of top 7000 expressed Ensembl-annotated transcripts in *sox10 bound* vs *sox10 lysate*.
C Heatmap of ZFIN-annotated genes found to be expressed in the neural crest at 14-19 ss. **C′** Venn diagram showing overlap and divergence of ZFIN neural crest Ensembl-annotated transcripts (TPM>2) in *sox10 bound* vs *sox10 lysate*. **C″** Venn diagram showing overlap and divergence of ZFIN neural crest gene names in *sox10 bound* vs *sox10 lysate*.

The initial analysis yielded promising results, but given that our approach was entirely exploratory in nature we sought to apply a statistical method to deduce candidates with potential functional significance in migratory NC biology. Unlike previous studies using TRAP that analysed their samples using microarray, our experiments were performed using RNA-seq to survey the translatome in an unbiased fashion and with greater sensitivity thus maximising information attained. While RNA-seq and microarrays can produce data that tightly correlate each other, RNA-seq has been shown to be more sensitive at quantifying transcripts at the lower extremes of abundance (Wang et al., 2014) as well as differentiating between isoforms (Zhao et al., 2014). Two previously published TRAP studies identified statistically enriched transcripts by either comparing the bound fraction to the unbound lysate after isolation (Heiman et al., 2008) or by using the input lysate before isolation as a control (Housley et al., 2014).

Given that TRAP experiments were performed by RNA-seq for the first time, there was no well-established and validated protocol for the differential analysis of this type of data. We therefore tested a number of RNA-seq analysis pipelines in order to produce meaningful results. Using two widely-used differential expression methods, DESeq2 and edgeR with unbound lysate (*sox10 lysate*) as a control, we failed to produce results with satisfactory distribution of *p*-values, indicating variation factors between the samples that cannot be accurately nor sufficiently normalised in the analysis.

A possible contributing factor to this observation is the low number of replicates (two biological replicates per sample), due to low RNA yields and high amounts of starting material required for TRAP as limiting experimental factors. Another possible reason is that the two fractions are inherently different in composition - theoretically, *sox10 bound* would contain transcripts bound to *sox10* polysomes, isolated from a specific cell type, while *sox10 lysate* would contain transcripts bound to non-*sox10* polysomes as well as cytoplasmic transcripts from both *sox10* and non-*sox10* cells.

Taken together, we deduced that the *sox10 lysate* is a poor control for unbiased RNA-seq differential expression analysis and subsequently used an alternative "background" translatome experimental approach.

3.2.2 Differential expression analysis of the neural crest translatome

We obtained a "background" translatome by using stage-matched double positive offspring from crosses between a ubiquitious *Avi effector*, *Tg(actB2:Avi-Cerulean-DrRpl10a)*^{ox111}, and a ubiquitious *BirA driver*, *Tg(actb2:BirA-2a-membCherry)*^{ct709a}. Similar to the previous section, total mRNA samples were collected from Streptavidin-isolated polysomes (*sox10* translatome and *bactin2* translatome) with two biological replicates each. Libraries were prepared using ribo-depletion (RiboZero) for deep sequencing (see Chapter 6 Materials and Methods).

For differential expression analysis, we employed a novel approach developed by Lior Pachter's lab for fast and accurate RNA-seq quantification without the need for genomic alignments. Reference transcripts (58,549 Ensembl cDNA sequences annotated in GRCz10/danRer10, used in previous section) were directly quantified by *pseudoalignment* of rRNA-filtered sequencing reads (see Chapter 6 Material and Methods) using *kallisto* (Bray et al., 2015). Bootstrapping estimates from *kallisto* quantification were subsequently used as input for differential expression analysis using *sleuth* (Pimentel et al., 2016).

Here, we report 78 significantly more abundant and 150 less abundant genes (81 and 156 Ensembl-annotated transcripts, respectively) in the *sox10* translatome vs *bactin2* translatome at FDR<0.1 (Benjamini-Hochberg adjusted *p*-value) (Figure 3.3A). The relatively low number of more and less abundant candidates is unsurprising and most likely reflects the higher statistical rigour of *sleuth* compared to current methods such as DESeq2 (Pimentel et al., 2016). A complete table of more and less abundant transcripts with FDR<0.1 is provided in Appendix A.

Genes with higher abundance in the *sox10* translatome

We first classified the higher abundance genes at FDR<0.05 according to PANTHER (http://pantherdb.org/) Protein Class annotations to gain a broad insight into their nature. 41 genes were successfully assigned, with an overwhelming majority (39%) annotated as *nucleic acid-binding* (Panther ID:PC00171) (Figure 3.3B). Remarkably, genes related to the translational machinery were found to occur at high frequency in this classification (6/16) with the presence of ribosomal proteins, *rpl27*, *rps21*, *rps26l*, *rplp2*, *rps11* and *rpl39* as well a translation initiation factor binding protein, *eif4ebp2*. *Eif4ebp2* has been characterised as a translational repressor (Mader et al., 1995), acting directly on *eif4e*, a key mediator of translation of oncogene mRNAs leading to tumour formation and metastasis (Mamane et al., 2004).

We noted that *elavl3* was the most abundant gene (β = 6.12; FDR = 0.0075). *Elavl3* (ELAV-like neuron-specific RNA binding protein 3, also known as *HuC*) is characterised in the zebrafish as an early neuronal marker, whose expression is first detected in the neural plate following gastrulation. As embryogenesis proceeds, *elavl3* becomes restricted to expression clusters within trigeminal ganglia neurons and along the spinal cord (Kim et al., 1996). Another gene found to be more abundant at levels equivalent to *elavl3* in the *sox10* translatome was *rbck1* (RanBP-type and C3HC4-type zinc finger containing 1) (β = 5.32; FDR = 0.0446). The role of *Rbck1* in zebrafish embryonic development is poorly understood. *Rbck1* was identified *in vitro* as a binding partner of *eya1* (EYA transcriptional coactivator and phosphatase 1) (Landgraf et al., 2010). *Eya1* mutants, *dog-eared* or *dog*, display abnormal phenotypes relating to ear, branchial arch and lateral line development (Whitfield et al., 1995). Similarly, *npm1a* (nucleophosmin 1a) was strongly more abundant in the *sox10* translatome compared to *bactin2* translatome (β = 4.87; FDR = 0.01). *Npm1a* is also poorly characterised at this stage but ZFIN-curated *in situ* data shows it is basally expressed in a ubiquitous manner but with specific, stronger expression in

embryonic structures such as the branchial arches, cranial ganglia and otic vesicle. Last but not least, *Egr1* (early growth response 1) (β = 1.96; FDR = 0.004) has been shown to form part of a genetic cascade with *runx3* (runt-related transcription factor 3) and *sox9b* [SRY (sex determining region Y)-box 9b] within the pharyngeal endoderm that controls differentiation of cranial NC cells into cartilage (Dalcq et al., 2012).

Genes with lower abundance in the *sox10* translatome

78 out of 99 lower abundance genes at FDR<0.05 were successfully classified according to PANTHER Protein Class annotations. *Receptor* (Panther ID:PC00197), *extracellular matrix protein* (Panther ID:PC00102), *signalling molecule* (Panther ID:PC00207) and cell-cell/matrix adhesion proteins [*cell adhesion molecule* (Panther ID:PC00069) and *cell junction protein* (Panther ID:PC00070)] represented 21.8%, 12.8%, 7.7% and 7.7% of classified genes, respectively (Figure 3.3B').

A triad of laminin genes - *lama1* (laminin, alpha 1, *bashful*), *lamb1a* (laminin, beta 1a, *grumpy*) and *lamc1* (laminin, gamma 1, *sleepy*) - were all found to be very significantly less abundant in the *sox10* translatome and thus more abundant in the *bactin2* translatome (β_{mean} = -0.66; FDR = 2.18 x 10^{-8} to 3.14 x 10^{-7}). All three belonged to the same mutant class affecting retinotectal axon pathfinding (Karlstrom et al., 1996), brain (Schier et al., 1996) and notochord development (Stemple et al., 1996). Another laminin gene that was also found to be as less abundant compared to the whole embryo translatome was *lama5* (laminin, alpha 5), previously shown to be essential for formation of the apical ectodermal fold during fin development (Webb et al., 2007). Cadherins are cell adhesion molecules with wide-ranging roles in NC development, with several cadherins undergoing dynamic changes in expression throughout the course of NC ontogeny (reviewed in (Taneyhill, 2008)). We found *cdh6* (cadherin 6) to be less abundant in our *sox10* translatome (β = -0.58; FDR = 0.007). Slit/Robo signalling has also been shown to be involved in guidance of the trunk neural crest migration via the ventral-medial pathway (Jia et al., 2005). Our

data revealed that both *slit3* and *robo1* were less abundant in the *sox10* translatome (β = -0.55, -0.64; FDR = 0.009, 0.021). Other key genes that we found to be significantly less abundant in the *sox10* translatome, i.e. enriched in the *bactin2* translatome included *fzd7b* (frizzled-7b), which is important for mesoderm formation (Nikaido et al., 2013), *celsr2* (*off-road*), a neuroepithelium marker (Wada et al., 2006) and *col8a1a* (*leviathan*), implicated in notochord development (Stemple et al., 1996).

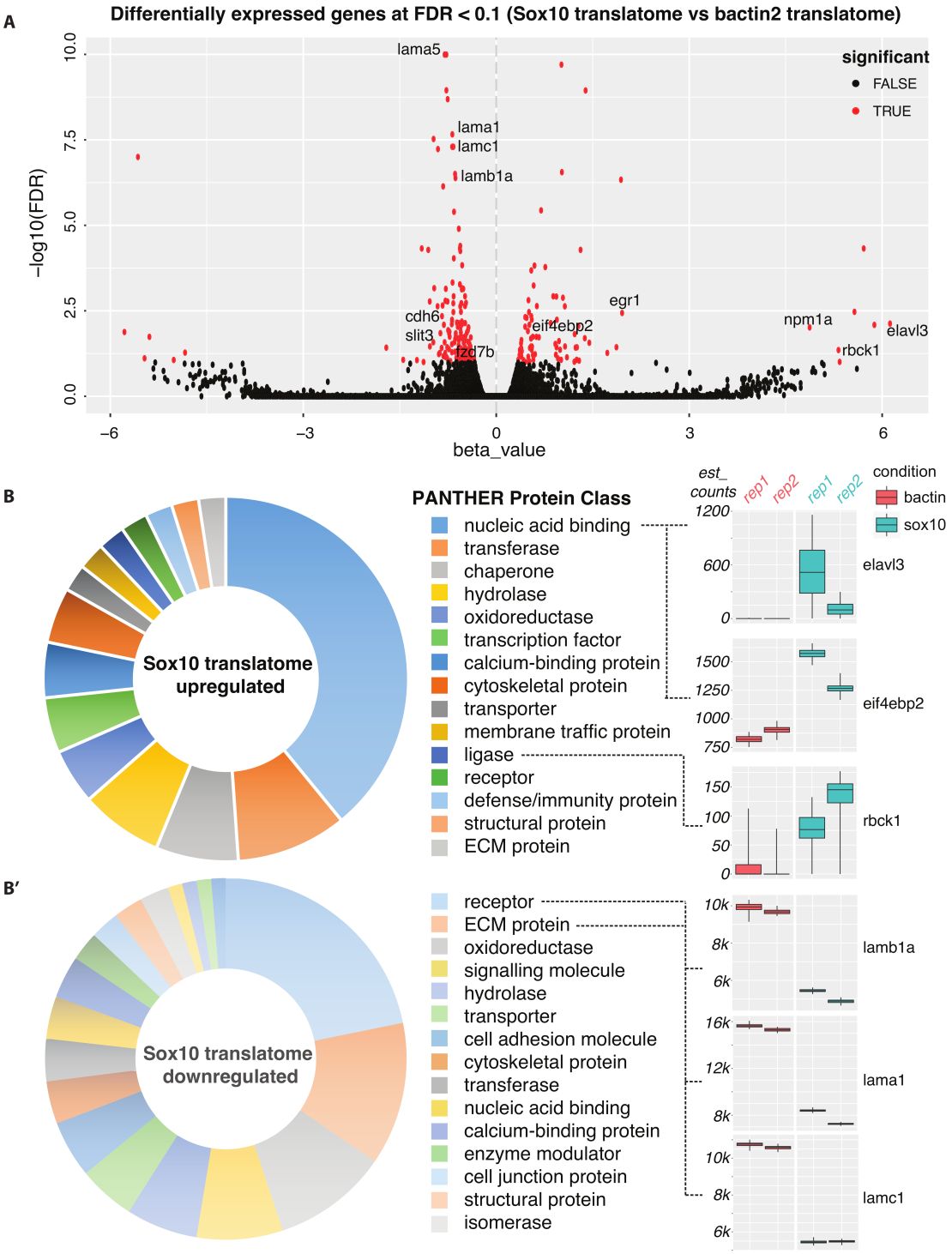


FIGURE 3.3: Differential expression analysis of the neural crest translatome.

A Volcano plot demonstrating differentially expressed genes more and less abundant in the *sox10 translatome* at FDR<0.1. Genes discussed in the main text are labelled. **B** PANTHER Protein Class classification of higher abundance genes at FDR<0.05. Boxplots with bootstrap estimated counts of three candidate genes, *elavl3*, *eif4ebp2*, *rbck1*, are shown. **B'** PANTHER Protein Class classification of less abundant genes at FDR<0.05. Boxplots with bootstrap estimated counts of three candidate genes, *lamb1a*, *lama1*, *lamc1*, are shown.

3.3 *biotagging*-RanGAP-seq of the migratory neural crest

Although analysing the *sox10* translatome was instrumental in obtaining a high resolution view of proteins actively being produced within a defined spatiotemporal window, this approach excluded details of the genetic programme underlying their translation. By moving "upwards" of the central dogma of molecular biology, we sought to retrieve information relating to the onset of gene activation, at the level of transcription at both coding and non-coding regions. Hence, we now shifted our focus to the NC nuclear (i.e. nascent) transcriptome in order to recover clues relating to their current, and possibly future, genetic programmes upstream of events occurring at the level of the translatome.

The nuclear transcriptome was obtained by crossing the same *sox10* BAC driver from the previous section to *Tg(actB2:Avi-Cerulean-GgRanGAP)*^{ct700a}. Similarly, a "background" nuclear transcriptome was obtained by using stage-matched (16-18 ss) double positive offsprings from crosses of the *Avi effector* to a ubiquitious *BirA driver*, *Tg(actb2:BirA-2a-membCherry)*^{ct709a}. RNA samples were collected from Streptavidin-isolated biotinylated nuclei (*sox10 nuclear* and *bactin2 nuclear*). To retain non-coding RNA species, libraries were prepared using ribo-depletion (RiboZero) for strand-specific deep sequencing (see Chapter 6 Materials and Methods). The findings discussed in this section have been published in (Trinh et al., 2017).

3.3.1 Bidirectional transcription across open promoters is enriched in the neural crest

Divergent transcription initiated within the same region, but in opposite directions, and the subsequent presence of short bidirectional transcripts is a known hallmark of active promoters (Core et al., 2008). We used our strand-specific sequencing datasets to evaluate differences in divergent transcription at the active promoters in NC nuclei compared to whole embryo nuclei. Open promoters were defined as **A**ssay for **T**ransposase **A**ccessible **C**hromatin (ATAC)-seq (Buenrostro et al., 2013)

positive regions at the 5'-end of the Ensembl-annotated zebrafish genes, extended ±100 bp to account for gene misannotation (**ATAC_TSS** set), and consisted of a total of 16,600 elements. Quantification of stranded transcription from two RNA-seq datasets - *sox10 nuclear* and *bactin2 nuclear*, revealed that bidirectional transcription was enriched in *sox10 nuclear* compared to *bactin2 nuclear* context. Using scatterplot comparison of **ATAC_TSS** reads from each strand, we demonstrated that *sox10 nuclear* had 86% (14,295/16,600) of **ATAC_TSS** peaks which were transcribed bidirectionally (read count>0 on both strands). On the other hand, *bactin2 nuclear* only had 32% (5383/16600) of **ATAC_TSS** peaks that were bidirectionally transcribed (Figure 3.4A)

To probe this phenomenon further, κ-means linear clustering (Ye et al., 2011) was used to analyse **ATAC_TSS** peaks based on their underlying transcription pattern in an extended dataset consisting of not only *sox10 nuclear* and *bactin2 nuclear*, but also included the *sox10 translatome* datasets from the previous section ("*sox10 ribo*"). Heat maps of the sense and antisense transcription at active promoters (including ± 1.5 kb window surrounding the TSS) for these three samples were generated using a seqMINER pipeline (Ye et al., 2011). This analysis identified short bidirectional transcripts at the majority of open promoters in *sox10 nuclear*, clustered according to patterns reflecting coding strand direction and structural organisation of genes (length and number of exons/introns within the analysed region). We found that the loci of interest, including known *bona fide* NC regulators, exhibited the characteristic pattern of short antisense transcripts within TSS/promoter region only in *sox10 nuclear*, but not in *bactin2 nuclear* or *sox10 ribo*, and were distributed across five clusters (Figure 3.4B). Scatterplot quantification of normalised counts of each strand from these clusters showed that 55% of the loci (1884/3391 in C1-3, 1657/2986 in C4-5) were specific to *sox10 nuclear* and only 5% (93/1600 and 68/1397) to *bactin2 nuclear* (Figure 3.4C). Similarly, scatterplot comparison of these *sox10 nuclear* enriched clusters (C1-5) highlighted clear differences in their Pearson correlation coefficients (ρ = 0.60, 0.80, 0.66 on the positive strand and 0.56, 0.75 on the negative strand) from the

background clusters (ρ = 0.92, 0.90) (not shown).

We investigated the nature of gene loci found in *sox10 nuclear* enriched clusters C1-5 compared to loci without bidirectional transcription at their promoters. Statistical overrepresentation tests demonstrated that the top enriched terms associating (p<0.01) to loci with bidirectionally transcribed promoters included developmental processes such as eye and sensory organ morphogenesis, neurogenesis, and cellular differentiation. In contrast, for loci not exhibiting bidirectional transcription, enriched GO terms reflected multiple metabolic processes (Figure 3.4D).

(Please refer to Figure 3.4 on the next page.)

A Scatterplot pairwise comparison of RNA-seq read counts overlapping **ATAC_TSS** peaks from each strand in *sox10 nuclear* and *bactin2 nuclear* demonstrate higher proportion of peaks are bidirectionally transcribed in *sox10 nuclear* (∼86%) compared to *bactin2 nuclear* (∼32%). **B** κ-means liner clustering heatmap of RNA-seq reads overlapping **ATAC_TSS** peaks. Five clusters (C1-5) reveal **ATAC_TSS** peaks where bidirectional transcription can be detected in *sox10 nuclear* (red box) while almost absent in *bactin2 nuclear* and *sox10 ribo* (red arrows). Clustering of reads visualised from low (yellow, 'low' transcription) to high (blue, 'high' transcription). **C** Scatterplot quantification of normalised counts of each strand from C1-5 showed that ∼55% loci (1884/3391 in C1-3, 1657/2986 in C4-5) were specific to *sox10 nuclear* (red dots on blue dots) and only ∼5% (93/1600 and 68/1397) to *bactin2 nuclear* (red dots on black dots). **D** Statistical overrepresentation tests of top enriched terms associating (p<0.01) to loci with bidirectionally transcribed promoters included developmental processes such as eye and sensory organ morphogenesis, neurogenesis, and cellular differentiation. Loci not exhibiting bidirectional transcription demonstrate enriched GO terms reflecting multiple metabolic processes.

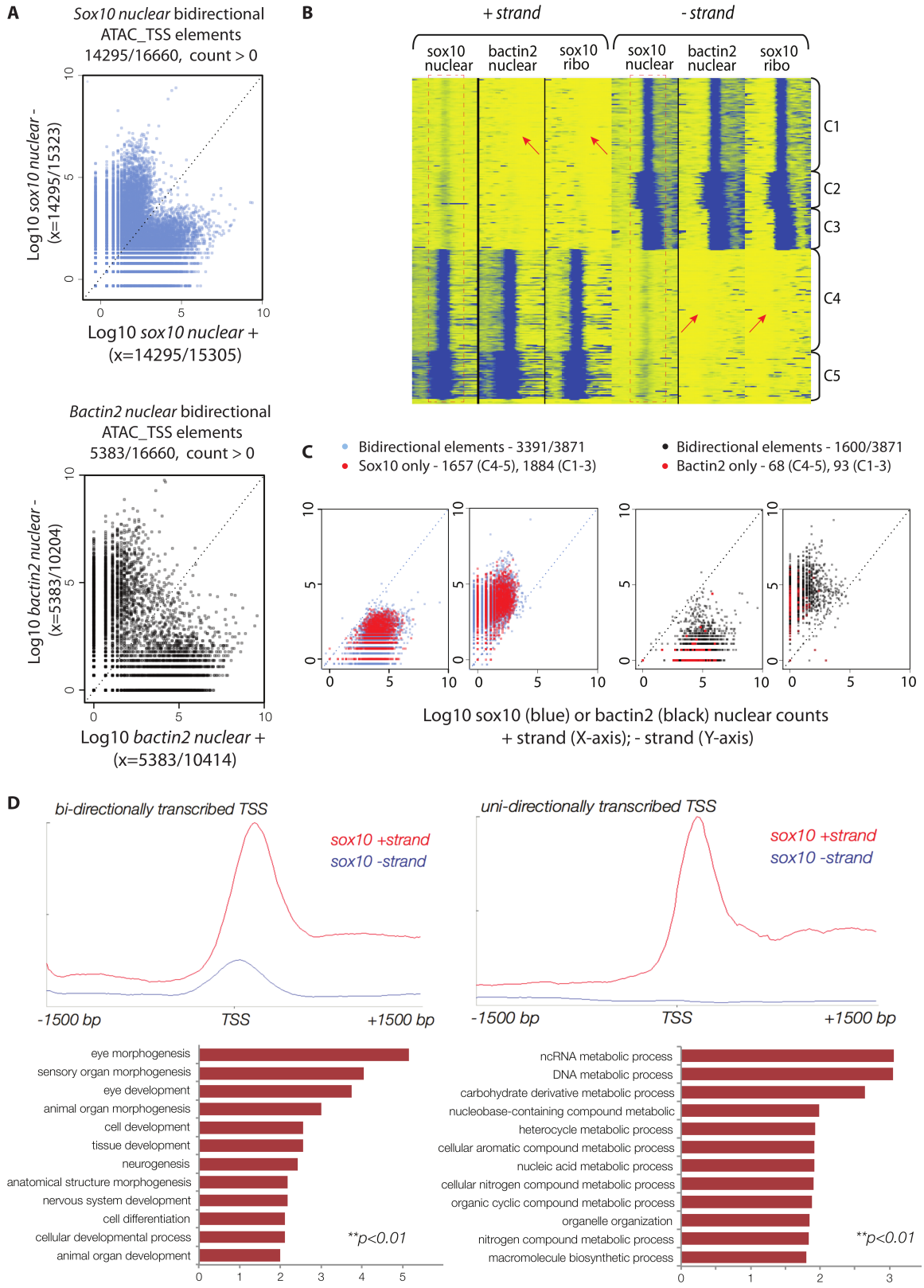


FIGURE 3.4: Bidirectional transcription across open promoters is enriched in the neural crest. Full legend is shown on previous page.

3.3.2 Neural crest molecular signature is encoded in the *cis*-regulatory transcriptome

Like active promoters, associated distal *cis*-regulatory elements are pervasively bidirectionally transcribed, resulting in nuclear-enriched enhancer RNAs (eRNAs) (Andersson et al., 2014; Kim et al., 2010).

We used our bidirectional *sox10 nuclear* and *bactin2 nuclear* transcriptome datasets at 16-18 ss to identify the ensemble of active enhancers that coordinate NC regulatory programme. We defined a set of putative distal regulatory elements to be used in further analyses (**ATAC_enhancer** set, 65,458 elements) as extragenic open chromatin ATAC peaks that did not overlap with Ensembl-annotated promoter regions and exons. To determine whether NC nuclear transcriptional profiles harbour tissue-specific patterns of enhancer transcription and identify distinct groups of *cis*-regulatory modules (CRMs) that define NC regulatory signature, we applied κ-means clustering algorithm to strand-split datasets obtained from *sox10 nuclear*, *bactin2 nuclear*, *sox10 ribo* and *bactin2 ribo* datasets, using seqMINER platform (Ye et al., 2011).

We identified two distinct clusters of CRMs that showed clear pattern of short eRNA bidirectional transcription only in *sox10 nuclear*, but not in *bactin2 nuclear* or in translatome samples (C1-2; total 17,071 CRMs). Clusters 1 and 2 showed a similar characteristic pattern but on different strands (Figure 3.5A). The merged profile obtained by calculating the mean for both clusters on each strand indicated clear enrichment in divergent transcription of **ATAC_enhancer** elements in *sox10 nuclear* over *bactin2 nuclear* (Figure 3.5A'). A third cluster (C3; 2,561 elements) showed a Cl1-like "architecture" with divergent transcription in *sox10 nuclear* only but included elements transcribed across longer regions surrounding the ATAC-peaks. These likely correspond to long intergenic non-coding RNAs (lincRNAs) and transcribed transposable elements, in addition to enhancers (Figure 3.5A). All three clusters showed enrichment of RNA-seq read counts in *sox10 nuclear* compared to *bactin2 nuclear* with

Pearson correlation values skewed from the background clusters (Figure 3.5A'). Of the ten clusters totalling 65,458 distal open chromatin regions, a large proportion (C4; 30,669 elements) were not transcribed. CRMs that show transcriptional signal in both *sox10 nuclear* and *bactin2 nuclear* formed two groups of reciprocal clusters each (C5-6 and 7-8; total 12,230 CRMs), likely revealing ubiquitous enhancers. Interestingly, two clusters contained elements transcribed ubiquitously and were detected in both nuclear and ribosomal compartments. These possibly represent novel non-annotated lncRNA loci that are exported from the nucleus (C9-10; total 2,927 elements) (Figure 3.5B).

To quantify tissue-specific *sox10 nuclear* enrichment of CRMs in clusters 1 and 2, we calculated the fold change (FC) as a ratio of transcriptional output (total FPKM over **ATAC_enhancer** peaks) in the *sox10 nuclear* and *bactin2 nuclear* samples. All valid CRMs (11,655 CRMs with FPKM>1 in both samples) were ranked according to the FC value. We defined three brackets of CRM activity, with low (FC<1), intermediate (1<FC<5) and high (FC>5) FC values, corresponding to different levels of *sox10 nuclear*-specific enrichment in the putative enhancer signal. When annotated using bedtools, we found that CRMs proximal to known factors in the NC gene regulatory network fall within the intermediate to high FC levels (Figure 3.5C).

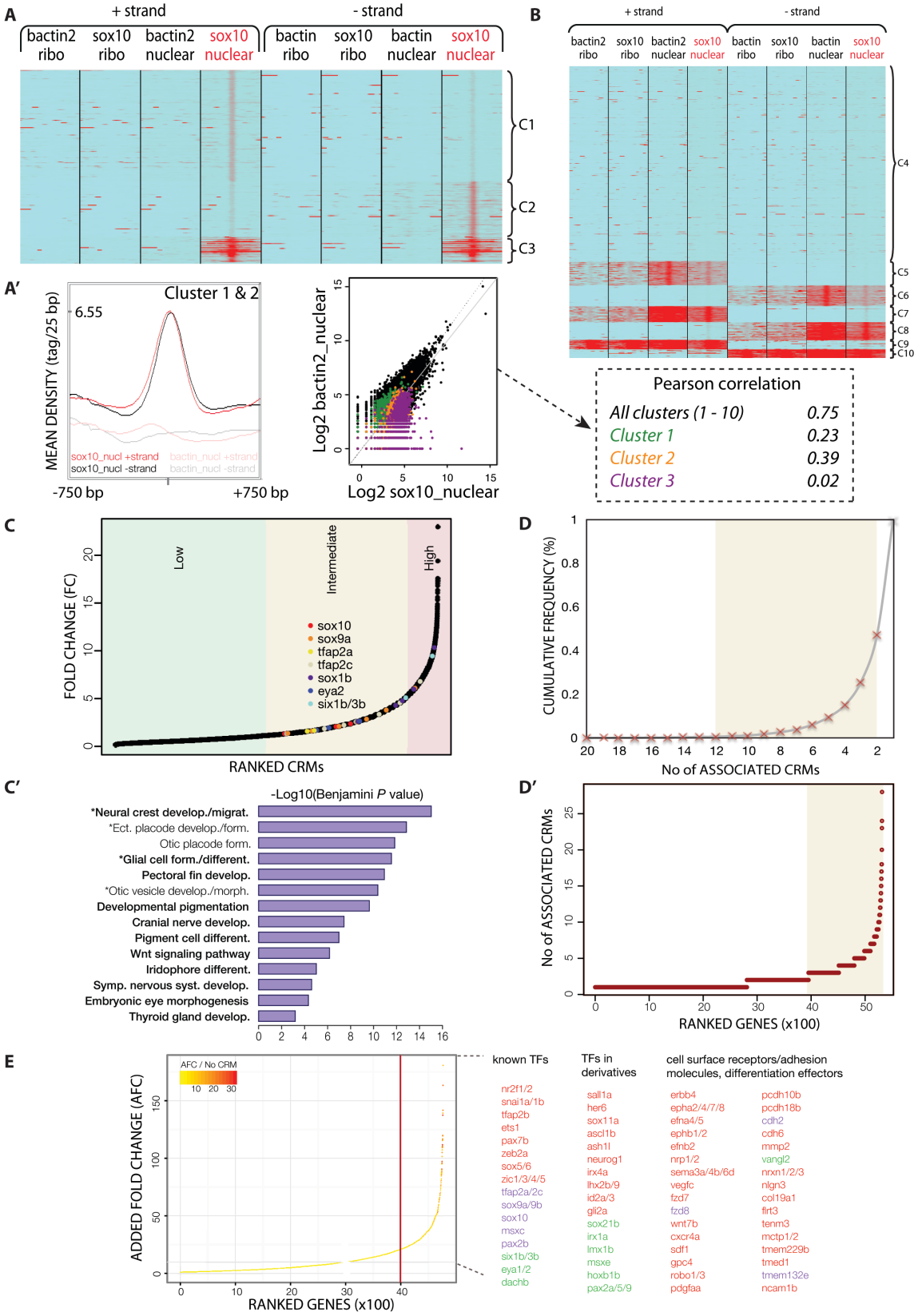


FIGURE 3.5: Neural crest molecular signature is encoded in the *cis*-regulatory transcriptome. Full legend is shown on the next page.

(Please refer to Figure 3.5 on previous page.)

A κ-means liner clustering heatmap of RNA-seq reads overlapping **ATAC_enhancer** peaks. Three clusters (C1-3) reveal **ATAC_enhancer** peaks where bidirectional transcription can be detected in *sox10 nuclear* while almost absent in *bactin2 nuclear*, *sox10 ribo* and *bactin2 ribo*. Clustering of reads visualised from low (cyan, 'low' transcription) to high (red, 'high' transcription). **A'** Merged profile of the mean for C1 and C2 on each strand show clear enrichment in bidirectional transcription of **ATAC_enhancer** elements in *sox10 nuclear* compared to *bactin2 nuclear*. Scatterplot comparison shows enrichment of RNA-seq read counts for C1-3 in *sox10 nuclear* compared to *bactin2 nuclear*. **B** Remaining clusters with **ATAC_enhancer** peaks where bidirectional transcription is not detected (C4), detected in both *sox10 nuclear* and *bactin2 nuclear* (C5-6 and C7-8), and detected in both nuclear and ribosomal compartments (C9-10). **C** CRMs in C1 and 2 ranked according to fold change (FC) as a ratio of transcriptional output (total FPKM over **ATAC_enhancer** peaks) in the *sox10 nuclear* vs *bactin nuclear*. **C'** Enrichment of GO terms related to neural crest and otic placode formation from CRMs in C1 and 2 with 1<FC<5 (5,087 elements). **D** Cumulative frequency graph of annotated genes based on number of associated CRMs per gene. **D'** Annotated genes ranked by the number of CRMs associated with them. **E** Gene loci associated by GREAT analysis to CRMs ranked according to Additive **F**old **C**hange (AFC) (sum of FCs of all CRMs assigned to a given regulated gene locus). Genes ranked beyond inflexion point with known involvement in neural crest development (red), otic placode/vesicle development (green), or both (purple) are listed.

To verify whether the collection of CRMs, identified above whose transcription is enriched in *sox10 nuclear*, harboured a neural crest molecular signature, we used **G**enomic **R**egions **E**nrichment of **A**nnotations **T**ool (GREAT) (McLean et al., 2010). GREAT allows assignment of functional significance to a set of non-coding genomic regions by analysing the annotations of nearby genes and integrating statistically significant distal regulatory elements. GREAT analysis of CRMs from cluster 1 and 2 (5,087 elements with 1<FC<5) resulted in enrichment of GO terms specific for biological processes closely related to NC and otic placode formation. This enrichment of NC-specific GO terms was statistically significant by both binomial and hypergeometric tests (Benjamini-Hochberg p<0.01), using the whole genome as background. Highlighted GO terms included NC development and migration, as well as biological processes covering the entire complement of NC derivatives (glial cells, pigment cells, melanocytes and iridophores, sympathetic nervous system, cranial nerve components, pectoral fin mesenchyme, as well as adrenal gland neural crest

contributions) (Figure 3.5C'). To our knowledge, this was the first comprehensive and genome-wide *in vivo cis*-regulatory signature obtained from migrating and differentiating NC.

We next used GREAT to annotate expressed genes associated with CRMs from cluster 1 and 2, resulting in a total of 6332 genes. The number of associated enhancers per gene was calculated. Cumulative frequency graph showed that ∼47% of the annotated gene were associated with at least two CRMs, whereas 25% of genes were associated with three or more elements (15% for four elements) (Figure 3.5D). This supports the current school of thought in enhancer dynamics, where tight tissue-specific patterns and gene expression of key developmental regulators are likely to be achieved by the combinatorial activity of multiple *cis*-regulatory elements working in a concerted fashion to regulate one locus. We ranked the annotated genes by the number of CRMs associated with them to identify key NC regulators likely to have higher activity. Highly regulated NC loci, defined as those falling beyond the inflexion point on the ranking plot, were found to be regulated by at least three elements (Figure 3.5D').

Although the use of multiple enhancers to control the same gene locus often leads to redundancy of transcriptional patterns, the action of multiple enhancers on expression level appears to be additive (Arner et al., 2015). To extract information about neural crest factors within our datasets, we calculated the **A**dditive **F**old **C**hange (AFC) as a sum of FCs of all CRMs assigned to a given regulated gene locus by GREAT analysis. These gene loci (a total of 4,767 genes) were then ranked according to their AFC value. A critical set of highly regulated loci were defined as genes whose AFC value falls beyond the inflexion point (Figure 3.5E). These included genes coding for known TFs involved in specification of NC and ectodermal placodes (Sauka-Spengler and Bronner-Fraser, 2008), and a number of TFs involved

in control of specific NC derivative fates (*ascl* and *ash* in sympathetic neurons, *neurogenin/lhx2b/her6/sox11a/irx4a* in sensory neurons; *sall1a/irx1* in pectoral fin mesenchyme or *gli2a* in adrenal lineage). Predominant categories included not only previously described, but also novel signalling and cell adhesion molecules involved as NC guidance and migration cues (*eph/ephrin*, *neuropilin/sema*, *wnt*, *sdf1/cxcr4* etc.). Given that the analysed developmental stage (16-18 ss) marks both migration and differentiation steps in NC ontogeny, a large number of highly regulated loci encode for downstream effectors involved in terminal differentiation of NC derivatives, such as *neurexins* and *neuroligins*, presynaptic cell-adhesion molecules secreted by sympathetic neurons, as well as *erbb4*, the *neuregulin* receptor which controls differentiation of NC-derived glia. We also discovered a high number of highly regulated novel transmembrane proteins (e.g. *tmed1*, *tmem229*, *bmctp1/2*, *flrt3*, *tmem132*, *tenm3*). This is consistent with the fact that NC cells rely heavily, at this stage in their development, on the cell-cell interactions with each other and their environment to migrate, stop and differentiate.

3.4 *In vivo* tools for studying gene regulatory networks

To validate putative enhancers identified in our *biotagging*-RanGAP-seq study, we sought a reliable system to screen *cis*-regulatory elements in the zebrafish embryo. Several methods currently exist and are widely-used, with *tol2*-mediated integration being a popular choice for vertebrates (Kawakami, 2007). In the zebrafish, *tol2* integration was established and is reliably efficient for somatic and germline integration, and is almost always the method-of-choice for creating transgenic reporter lines. For transient DNA integration experiments, however, *tol2*-mediated integration often produces variable results with high background (in the muscle, in particular) and mosaicity, and we therefore often rely on the generation of F_1 offspring carrying enhancer-reporter constructs for proper interpretation of results. Hence,

we built a new system for transient enhancer assay experiments in the zebrafish embryo, using *Ac/Ds*-mediated integration enhancer reporter cassettes.

3.4.1 *Ac/Ds*-mediated transposition of genetic elements

The *Ac/Ds* dissociation element was first identified in maize by Barbara McClintock in the late 1940s (McClintock, 1950), leading to her 1983 Nobel Prize in Medicine. This led to the molecular isolation and characterisation of the components required for transposition to occur - two *Ds* (*Dissociation*) genetic elements and an *Ac* (*Activator*) transposase (Fedoroff et al., 1983). The system was subsequently shown to exist in other plants (plantae), but its true utility become apparent when Weil and Kunze demonstrated that *Ac/Ds* elements were able to transpose in *Saccharomyces cerevisiae* (fungi), without the presence of plant proteins (Weil and Kunze, 2000), presenting the first evidence of *Ac/Ds* trans-kingdom versatility. Buoyed by this important finding, *Ac/Ds* elements were then shown to facilitate integration of a reporter construct in the zebrafish with high efficiency leading to remarkable germline transmission rates (Emelyanov et al., 2006) and this approach was the integration method of choice used to generate zebrafish transgenic lines for the chemical-inducible LexPR transactivation system (Emelyanov and Parinov, 2008).

The *Ac/Ds* toolkit for assaying enhancer activity in the embryo

We expanded the use of *Ac/Ds*-mediated integration in the zebrafish by adapting the system for screening *cis*-regulatory elements as an alternative to the more commonly-used *tol2*-integration-based approach. We cloned a plasmid construct that combined a cassette with eGFP driven by the zebrafish E1b minimal promoter (gift from Sarah de Val) (Becker et al., 2016) flanked by *Ds* elements in the correct orientation for integration into the genome (Figure 3.6). Putative enhancers/*cis*-regulatory elements were cloned in the *Nhe*I restriction site upstream of the minimal promoter using

InFusion™ or Gibson Assembly cloning (Gibson et al., 2009). With our microinjection conditions (see Chapter 6 Materials and Methods), this approach demonstrated low mosaicity thanks to a low load requirement for efficient integration (30 pg per embryo, compared to *tol2* constructs at up to 200 pg per embryo) and very high somatic integration rate, hence providing an efficient binary yes/no tool for transient screening of enhancer activity. Germline transmission rates were comparable to those obtained using the *tol2*-integration system.

3.4.2 Transcribed *cis*-regulatory elements have biological activity

Using our *Ac/Ds* enhancer assay tookit, we screened putative enhancers from *k*-means clusters 1 to 3, that were enriched in bidirectional transcriptional in *sox10 nuclear* vs *bactin2 nuclear* and associated with NC genes. Seven putative enhancers were cloned and microinjected according to our protocol. Of these, we demonstrated a positive identification rate of ∼71% with 2, 2 and 1 positive enhancer(s) for *sox10*, *pax3a* and *ets1*, respectively. These clearly showed different expression patterns that were specific to NC related regions within the embryo with high specificity and low background (Figure 3.6), therefore presenting us with a set of novel enhancers for further characterisation that may be important for neural crest ontogeny.

Enhancer	Position (danRer7)	Forward primer	Reverse primer
Sox10 enh 3	chr3:2520055-2520923	GAGAGTGTATGGGTGCCTCC	CGCCCTCAAACTCAATATTCAGC
Sox10 enh 5	chr3:2538101-2538698	TGTGGAATGAACCGTCAAAA	GTGAACAGAAATGCGTCACC
Pax3a enh 4	chr2:47557230-47558100	TATTCACCTGCCCTTTCTCG	GCACGTGATCCTCTCGAAAT
Pax3a enh 5	chr2:47574905-47575347	CCACACAATTGTCCCAACAC	CAGCCATCTCACCTTTACCC
Pax3a enh 2	chr2:47609437-47609939	GGTTTTGGGAAGCTTTTATCAC	TTCTGTCATTACCACAGCAAAAA
Ets1 enh 5	chr18:46901069-46901697	GGAGCTAAGCAGGCAGTGAT	AACGATCAAACGATCAAAAGAAA

TABLE 3.1: Chromosome positions and primer sequences of new enhancers identified from C1, 2 and 3 (except for Pax3a enh 2).

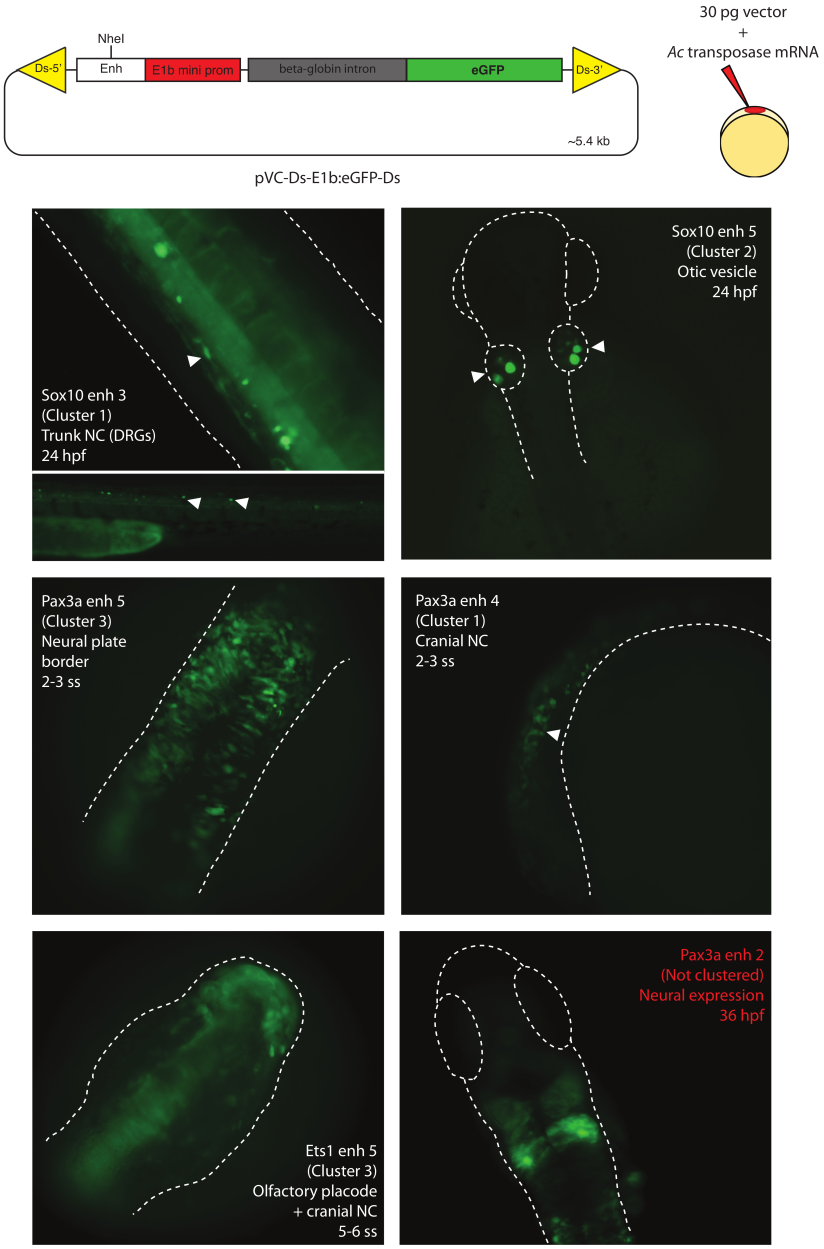


FIGURE 3.6: The *Ac/Ds* toolkit for assaying enhancer activity in the embryo.

The pVC-Ds-E1b:eGFP-Ds construct contains an expression cassette flanked by *Ds* integration elements. The expression cassette contains eGFP driven by a *cis*-regulatory element-of-interest via the zebrafish E1b minimal promoter. 30 pg of this construct is co-microinjected with *in vitro* transcribed *Ac* transposase into the blastomere of early one-cell stage embryos within 20 minutes of spawning. Sox10 enh 2 (C1) activity is detected in trunk as discrete puncta, possibly neural crest forming dorsal root ganglia (DRGs). Sox10 enh 5 (C2) activity is detected very clearly in a bilateral pattern within the otic vesicles. Pax3a enh 5 (C3) is active in the neural plate border. Pax3a enh 4 (C1) signal is present in the cranial neural crest. Ets1 enh 5 activity is detected in the neural tube, early migrating cranial neural crest and in the olfactory placode region. An enhancer that was not clustered (Pax3a enh 2) shows strong activity in the neural regions - midbrain, hindbrain and neural tube (not shown).

3.5 Discussion

We probed the NC GRN at the level of subcellular compartments, enabled by our *biotagging* methodology. Our results showed that subcellular consideration when probing cellular identity is far from trivial, as the results obtained between the two compartments were markedly different thus revealing a sharper resolution of gene expression at the genome-wide level.

3.5.1 Current insights from the *sox10* translatome at 16-18 ss

Higher abundance genes. We report a cohort of ribosomal proteins more abundant in the neural crest translatome. While initially surprising, this observation is not entirely novel as tissue-specific expression of ribosomal proteins in human tissues has been observed and also shown to be dysregulated in cancer (Guimaraes and Zavolan, 2016). In fact, hints pointing at the link between ribosomal proteins and cancer surfaced nearly a decade ago, when a screen for tumour suppressor genes in the zebrafish yielded a cohort of tumour-forming lines that carried mutations in different ribosomal protein genes (Amsterdam et al., 2004). Although none the ribosomal proteins listed in the study by Amsterdam *et al.* were found in our higher abundance list, this does not discount the possibility that taken together with previous findings, our observation may point towards an unexplored facet of NC ontogeny driven by regulation of the translational machinery.

We found *elavl3* to be the most abundant gene. Interestingly, Kim *et al.* had noted that *elavl3* expression was observed in all cranial ganglia by 32 hpf although data was not shown (Kim et al., 1996). Using *elavl3* as a genetic marker, two studies went on to describe the mechanisms underlying the establishment, neurogenesis and chemokine-regulated assembly of trigeminal sensory neurons into trigeminal sensory ganglia (Knaut et al., 2005; Caron et al., 2008). As cranial sensory ganglia (including trigeminal sensory ganglia) are derived from both cranial placodes (which

are ectodermal) and the NC, this suggests *elavl3*'s involvement in the GRN underlying differentiation of NC cells into trigeminal sensory neurons. While *sox10* has been shown to directly regulate *neurogenin1* in the specification of sensory neurons of the dorsal root ganglia (DRG) (Carney et al., 2006), it is currently unknown if *sox10* similarly regulates *elavl3* in the specification of trigeminal sensory neurons, given that it is highly translated in *sox10*-expressing cells at this developmental stage.

Landgraf *et al.* demonstrated that *rbck1* morphants mimicked *eya1* mutants with pharyngeal arch cartilage defects and abnormal ear development, therefore presenting functional significance of their *in vitro* results presenting *rbck1* as a binding parter of *eya1* (Landgraf et al., 2010). As the phenotypic characterisations of *rbck1* morphants were performed at 5 dpf onwards, it is currently still unknown if *rbck1* plays any role in NC function and/or patterning of the craniofacial placodes in *sox10*-expressing cells during early embryogenesis that could explain the reason(s) behind its enrichment in our dataset.

We also found *egr1* to be more abundant in the *sox10* translatome, although Dalcq *et al.* reported no co-localisation of *egr1* with *sox9a*, a marker of cranial NC cells, at 48 hpf (Dalcq et al., 2012). In addition, a study using *egr1* morphants demonstrated its function in morphogenesis of neuroepithelium-derived tissues (retinal ganglion and retina) in the eye (Hu et al., 2006). Both of these developmental processes occur much later than the developmental stage of our *sox10* translatome. At earlier stages, *egr1* has been shown to be co-expressed with *snai1a* (snail family zinc finger 1a) in adaxial cells of the posterior segmental plate mesoderm (Close et al., 2002) at 12 ss and in rhombomeres 3 and 5 at 14-19 ss (Rauch *et al.* 2003, Submission and Curation of Gene Expression Data directly submitted to ZFIN). Due to the differences in developmental stage and tissues studied, these findings neither contradict not support our findings, and it remains to be validated if *egr1* is important for migrating NC function and/or differentiation of its derivatives.

Lower abundance genes. Laminin genes have been described in the form of early mutants from the Nüsslein-Volhard screens, highlighting their significance in early development. Three of these, *lama1*, *lamb1a* and *lamc1*, were found to be less abundant (and therefore more abundant in the environment of the migratory NC) in our dataset. Morphological analysis later characterised the three laminins to be important for retinal differentiation and maintenance, with mutants showing defective lens (ectoderm-derived), optic nerve and ganglion cell layer (Biehlmaier et al., 2007) (both neuroepithelium-derived). Furthermore, *lamb1a* was shown to regulate left-right asymmetry within the lateral plate mesoderm, with *grumpy* mutants displaying abnormal symmetry of the liver and pancreas (Hochgreb-Hägele et al., 2013). Collectively, these findings demonstrate the important role of laminins in non-NC-derived structures. On a side note, migrating cranial NC cells have been shown *in vitro* to be more sensitive to surrounding laminin concentration than trunk NC cells, requiring higher concentrations of laminin for persistence and velocity during their migration (Strachan and Condic, 2003).

In the avian embryo, CDH6B (*cdh6 orthologue*) was shown to be a direct target of the transcriptional repressor Snail2 and important for NC delamination from the neural tube. This study suggested that downregulation of *Cadherin 6b* by Snail2 enabled delamination of trunk and midbrain-level NC cells from the neural tube (Taneyhill et al., 2007). However at an earlier timepoint, Cdh6-mediated stimulation of BMP-signalling in the premigratory NC cells was found to be necessary for the epithelial-to-mesenchymal transition (EMT) and subsequent delamination (Park and Gumbiner, 2010). This dynamic process is further supported by a recent study in zebrafish where both *cdh6* and *sox10*-expressing cells were characterised during EMT of the cranial NC (Clay and Halloran, 2014). As embryonic development progressed from 10 ss to 18 ss, the expression of *cdh6* and *sox10* diverged - initially broadly co-expressed in the premigratory NC within the neuroepithelium at the peak of *cdh6* expression levels, after delamination NC cells only expressed *sox10*, but not *cdh6*, in line with our finding that this gene was depleted from the *sox10*

translatome at later migration stages.

During EMT, the roundabout guidance receptor 1 (*Robo1*) was shown to be expressed in the migrating NC while the chemorepellant ligands Slits (including *slit3*) are expressed in the surrounding dermomyotome (Jia et al., 2005). Over-expressing *Robo1* in the neural tube also impairs NC migration in the trunk (Giovannone et al., 2012). The less abundance of both *robo1* and *slit3* in our dataset suggests both that *slit3* might already be enriched within the environment, but also that the translation of *robo1* is precisely regulated and likely has not onset at the time of our analysis, when trunk NC is yet to migrate.

3.5.2 Current insights from the *sox10* nuclear transriptome at 16-18 ss

Promoter-associated RNAs. Our findings that promoter-associated RNAs are enriched in the NC nuclear transcriptome are in line with previous studies suggesting that antisense transcription is associated with promoters of transcriptional regulators (Lepoivre et al., 2013), where it arises as a consequence of RNA Polymerase II stalling. RNA Polymerase II pausing at promoters of important developmental regulators has been proposed to be crucial for the coordination of transcriptional events during development, as it allows for their dynamic and rapid activation (Boettiger and Levine, 2009; Gaertner et al., 2012; Henriques et al., 2013). This supports our results showing that *sox10 nuclear* enriched clusters with bidirectional transcription at promoters contained genes involved in various developmental processes, while loci not exhibiting bidirectional transcription at TSS had no developmental features.

Enhancer-derived RNAs. eRNAs are typically short, non-spliced and sensitive to degradation by the nuclear exosome complex, much like upstream antisense transcripts produced from divergent promoters of protein-coding genes discussed in above (Andersson et al., 2015). The shared features of promoters and enhancers

(similar frequencies of core promoter elements and modes of TF binding, and divergent transcription) have been used to suggest a unifying architecture for these regulatory elements. The only fundamental distinction between promoters and enhancers is the differential RNA stability of post-initiation sense RNA transcript (Andersson et al., 2015; Core et al., 2014; Kim and Shiekhattar, 2015). Recent studies demonstrated that enhancer transcription correlates with transcriptional outputs from the downstream coding genes, and is the earliest event in the transcriptional cascade, often preceding gene activation (Arner et al., 2015). In light of this, we demonstrated that we could recover the molecular signature of migratory neural crest cells by clustering *cis*-regulatory elements based on the presence of eRNAs, providing evidence that the phenomenon reported by Arner *et al.* is also likely to be regulated during the execution of developmental programmes.

3.5.3 Compartmentalising the identity of migratory neural crest

Our zebrafish *sox10* translatome at 16-18 ss both corroborates existing knowledge of NC activity at this developmental stage, when the cranial NC is actively migrating and beginning to differentiate and trunk NC migration is still in its infancy, as well as raises potential avenues for further investigation. For example, the translational control of developmental mRNAs has yet to be characterised within the context of the dynamic gene expression changes during EMT within the forming NC cells, despite being abundantly studied in cancer metastasis. Timing is also key - a tissue-specific translatome dataset provides a snapshot of types of proteins actively produced within a defined spatiotemporal window without the "noise" from cytoplasmic or nuclear transcripts. The results presented here suggest that at this time-point, specification and/or differentiation of trigeminal sensory neurons seem to dominate with *elavl3* being heavily translated while translation of ectodermal, mesodermal and neuroepithelial genes was stalled. A stark observation is the complete lack of differentially-expressed early cranial NC markers such as *foxd3*, *pax3a*,

snai1b, *tfap2*α/β, *ets1*. This is unsurprising, as the cranial NC is fully specified by this time and either migrating or differentiating while other *sox10*-expressing cells begin to pattern the cranial placodes.

To our surprise, differential expression analysis of Ensembl-annotated genes (danRer7/Zv9) did not reveal any clear NC signature between *sox10 nuclear* and *bactin2 nuclear* samples. We reasoned that a combination of factors contributed to the absence of a "typical" NC gene signature. First, in our *bactin2 nuclear* samples the detected transcriptional output integrated signals from all cell types, including NC cells. Second, analysis of total transcriptional activity at coding loci in the nucleus uncovered pervasive features of stem cell-like transcription in NC and the whole embryo nuclei at 16-18 ss. This is in clear contrast to the differential expression of nuclear transcription in cardiomyocytes discussed in Chapter 2, observed at a developmental stage more than ten hours later, which successfully revealed the transcriptional signature of a differentiated cell-type.

The NC ontogeny during embryonic development is a highly dynamic and context-dependent process. Therefore it is necessary to interpret the transcriptome and translatome profiling data with the spatiotemporal sensitivity within the organism being studied. Taken together, our findings revealed the NC migratory and differentiating regulatory signature at this stage that is consistent with previously published factors, as well as identification of novel players. We also provided the first global representation of upstream regulatory control and presented a genome-wide cohort of putative enhancers for downstream characterisation and analyses. We anticipate that this resource will be instrumental to increasing our understanding of the NC GRN.

In the next chapter, we demonstrate the use of subcellular compartmentalisation to identify and characterise novel as well as previously annotated non-coding RNAs.

Chapter 4

Identification and characterisation of zebrafish non-coding RNA species

4.1 Non-coding RNAs are new game players of development

The advent of deep sequencing technology meant that genomes can be surveyed with a bird's eye view instead of from a loci-to-loci basis. In light of this, whole transcriptome analyses have revealed the complex regulatory landscape of the genome where protein-coding genes only represent a small fraction (∼2%) of the genome which is otherwise pervasively transcribed (Djebali et al., 2012). The "RNA-regulatory world" represents a paradigm shift in the understanding of how gene expression is controlled at the genomic level that has emerged following this observation, a rapid area of research whose concept is repeatedly being re-defined as more evidence emerged. Part of these findings have led to the argument that non-coding genomic regions, traditionally regarded as "junk DNA", may actually have important biological roles underlying complex regulation of gene expression and function via the activity of non-coding RNAs (ncRNAs). In this chapter, we describe our findings in probing the *biotagging* datasets for three classes of ncRNAs - circular RNAs (Salzman, 2016), long non-coding RNAs (Quinn and Chang, 2015) and transcribed transposable elements (Garcia-Perez et al., 2016).

4.2 Circular RNAs present a new paradigm in the transcription of genes

In this section, we present the identification of putative circRNAs in our compartment-specific *biotagging* datasets - a novel approach as all previous studies utilised and interpreted data obtained from total RNA from whole cells/tissues. We have also optimised a novel amplification-based method for screening and validating the sequence of putative circRNAs with low levels of abundance.

4.2.1 Putative zebrafish circular RNAs form genomic clusters and are compartment-specific

We used a published Python script, *find_circ.py* (Memczak et al., 2013), to identify putative circRNAs in six *biotagging* RNA-seq datasets/libraries prepared using ribo-depletion - *sox10 16-18 ss nuclear*, *bactin2 16-18 ss nuclear*, *sox10 16-18 ss ribo*, *bactin2 16-18 ss ribo*, *myl7 26 hpf nuclear* and *bactin2 24 hpf nuclear*. We identified hundreds of putative circRNAs within each cell type at a rate of ∼6-12 circRNAs per million reads (Table 4.1).

Sample	Average no. of reads (2 replicates/sample)	No. of circRNA candidates	% circRNAs predicted	No. of circRNAs per million reads	No. of reads assigned to circRNAs
Sox10 16ss nuclear	51,151,788	348	0.0007	6.803	6,349
Bactin16ss nuclear	31,511,738	242	0.0008	7.680	4,080
Myl7 26 hpf nuclear	49,748,086	382	0.0008	7.679	17,099
Bactin 24hpf nuclear	34,220,784	337	0.0010	9.848	6,368
Sox10 16ss ribo	46,661,325	556	0.0012	11.916	10,954
Bactin 16ss ribo	42,974,211	356	0.0008	8.284	7,847

TABLE 4.1: Identification of putative circRNAs in *biotagging* RNA-seq datasets.

Interestingly, we discovered that predicted circRNAs often "congregate" within a genomic locus overlapping annotated genes e.g. *acy3.1/2* (Figure 4.1A; top). Next, we quantified the occurrences of predicted circRNAs both at the level of these genomic loci (containing clusters of putative circRNAs) and separately as individual putative circRNAs. At both levels, the circRNAs appeared to be spatiotemporally regulated, with clear overlapping as well as mutually exclusive groups (Figure 4.1A;

bottom left). Individual putative circRNAs formed clusters consisting of 1 to 25 circularisation events within a particular genomic locus (Figure 4.1A; bottom right).

Our key observation from cell types where we have surveyed both nuclear and polysomal compartments (*sox10 16-18 ss* and *bactin2 16-18 ss*) was that many putative circRNAs were localised within the polysomal compartment. In particular, identified circRNAs were either common to both, or exclusive to either compartment (Figure 4.1B). This preliminary finding contradicts previous reports, based on candidate-based approaches that suggested that circRNA(s)-of-interest are always depleted in the polysomal-bound fraction (Capel et al., 1993; You et al., 2015), and proposing that circRNA either functioned in the nucleus (Li et al., 2015) or as miRNA sponges in the cytoplasm (Memczak et al., 2013; Hansen et al., 2013). However, our finding that circRNAs were associated with polysomes (Figure 4.1C) is in line with recent ribosome profiling studies which showed that ncRNAs, including 5'UTRs, were associated with ribosomes, suggesting that association with ribosomes does not necessarily equate to translation into proteins (Chew et al., 2013)

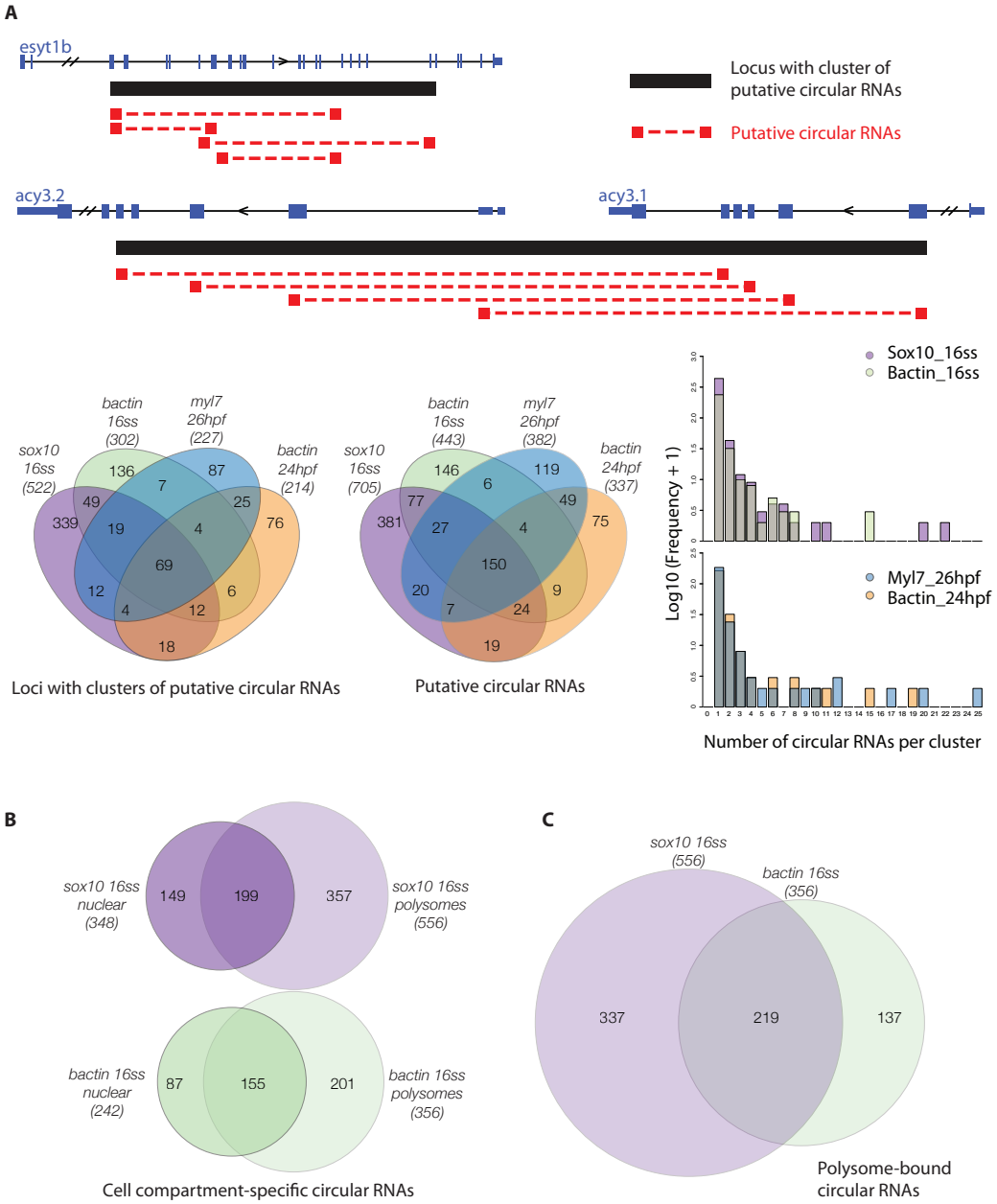


FIGURE 4.1: Putative zebrafish circular RNAs form genomic clusters and are compartment-specific.

A Schematic representation of a gene (*esyt1b*) and genomic locus with two tandem duplicated genes (*acy3.1/2*), both containing multiple overlapping circularisation events predicted. Venn diagrams show overlapping as well as mutually-exclusive genomic clusters of putative circRNAs (left) and individual putative circRNAs (right) according to cell type (*sox10 16ss* and *bactin 16ss* consist of both nuclear and ribosomal compartments, *myl7 26hpf* and *bactin 24hpf* consist only of nuclear compartment). Histogram shows frequency distribution of number of circRNAs predicted per genomic cluster of circRNAs (from 1 to 25 circRNAs/cluster), according to cell-type. **B** Venn diagram showing overlapping and mutually-exclusive putative circRNAs from each subcellular compartment (nuclear or ribosomal) from *sox10 16ss* and *bactin 16ss*. **C** Polysome-bound putative circRNAs are detected in both overlapping and cell type-specific fashions.

4.2.2 Predicted circularisation events are weakly correlated to underlying mRNA expression

To test the relationship between the expression of predicted circRNAs and their underlying linear mRNA transcripts, we performed a scatterplot comparison between the total read counts for each cluster of circRNAs in each cell type. We plotted the total **F**ragments **P**er **K**ilobase of transcript per **M**illion reads (FPKM) values of corresponding Ensembl-annotated mRNAs transcribed on the same strand (where more than one mRNA transcript were annotated, the sum of FPKM values was plotted) in function of total number of circRNA reads per locus and found a weak positive correlation between the two groups of transcripts, with Pearson correlation values of ρ = 0.297, 0.397, 0.397 and 0.447 for *sox10 16-18 ss nuclear and ribo*, *bactin2 16-18 ss nuclear and ribo*, *myl7 26 hpf nuclear* and *bactin2 24hpf nuclear*, respectively (Figure 4.2).

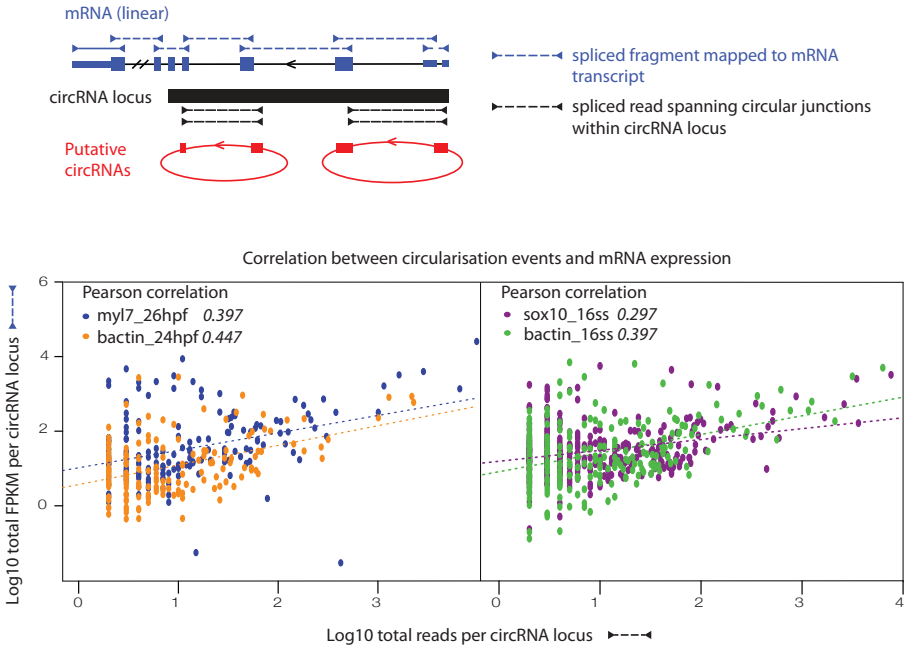


FIGURE 4.2: Correlation between expression of putative circRNAs and underlying mRNA expression.

Scatterplot comparison between the total read counts from each genomic cluster of circRNAs according to cell type and the total FPKM values of corresponding Ensembl-annotated mRNA transcripts found on the same strand. If more than one mRNA transcripts were annotated, the sum of FPKM values was plotted.

4.2.3 Circularisation "hotspots" reveal enrichment at tandem duplicated genes

Detected putative circRNAs in our datasets had up to ∼50-fold lower abundance compared to previous studies, most likely reflecting differences in species, RNA preparation, cell types, nature of sample (whole tissue versus compartment-specific, specific embryonic populations versus cultured cells), and choice of algorithm software for circRNA prediction (for a comprehensive comparison between currently available software, see (Hansen et al., 2015)). To obtain a high confidence set of circRNAs for validation, we reasoned that circRNAs detected across multiple samples were more likely to be functional and therefore of potential biological significance. Since we obtained a more extensive complement of cell types (NC, cytoskeleton and myocardium) and developmental stages (16-18 ss, 24-26 hpf) for the nuclear compartment, we chose to scrutinise circularisation events that were present in all four nuclear datasets - *sox10 nuclear 16-18 ss*, *bactin2 nuclear 16-18 ss*, *bactin2 nuclear 24 hpf* and *myl7 nuclear 26 hpf*, which we defined as "hotspots".

We noted that many of our "hotspot" circRNAs overlapped tandem duplicated genes in the zebrafish genome, where the predicted circularisation junctions were often present in neighbouring duplicated loci, thus suggesting that circRNAs would span several kilobases between tandem genes within a duplicated gene region. We quantified this observation and found that "hotspots" overlapping duplicated genes (Ouedraogo et al., 2012), which we now class as "tandem-circRNAs", were highly significantly enriched (55 out of 112; hypergeometric p-value of 3.43×10^{-5}) in this cohort of putative circRNAs. To explore co- or post-transcriptional splicing events that underlie tandem-circRNAs, we performed *de novo* transcript assembly for each dataset using StringTie (Pertea et al., 2015). Interestingly, for 20 out of 55 putative tandem-circRNAs, we observed assembled "chimaeric" transcripts that consisted of two or more exons from tandem duplicated genes (with predicted splicing) and one or more fusion transcripts with unannotated splicing junctions between one

tandem transcript to another (Figure 4.3). While these findings do not necessarily suggest biological significance of tandem-circRNAs themselves, we speculated that their enrichment at duplicated gene regions hints at an uncharacterised mechanism where closely-related tandem duplicated transcripts (but highly similar at the level of protein structure) undergo circularisation to produce only one functional full-length mRNA transcript per duplicated region and ensure translation of a correct protein. As first-pass test of our hypothesis, we sought to validate the presence of tandem-circRNAs by performing a screen to resolve full composition of the circRNA transcripts.

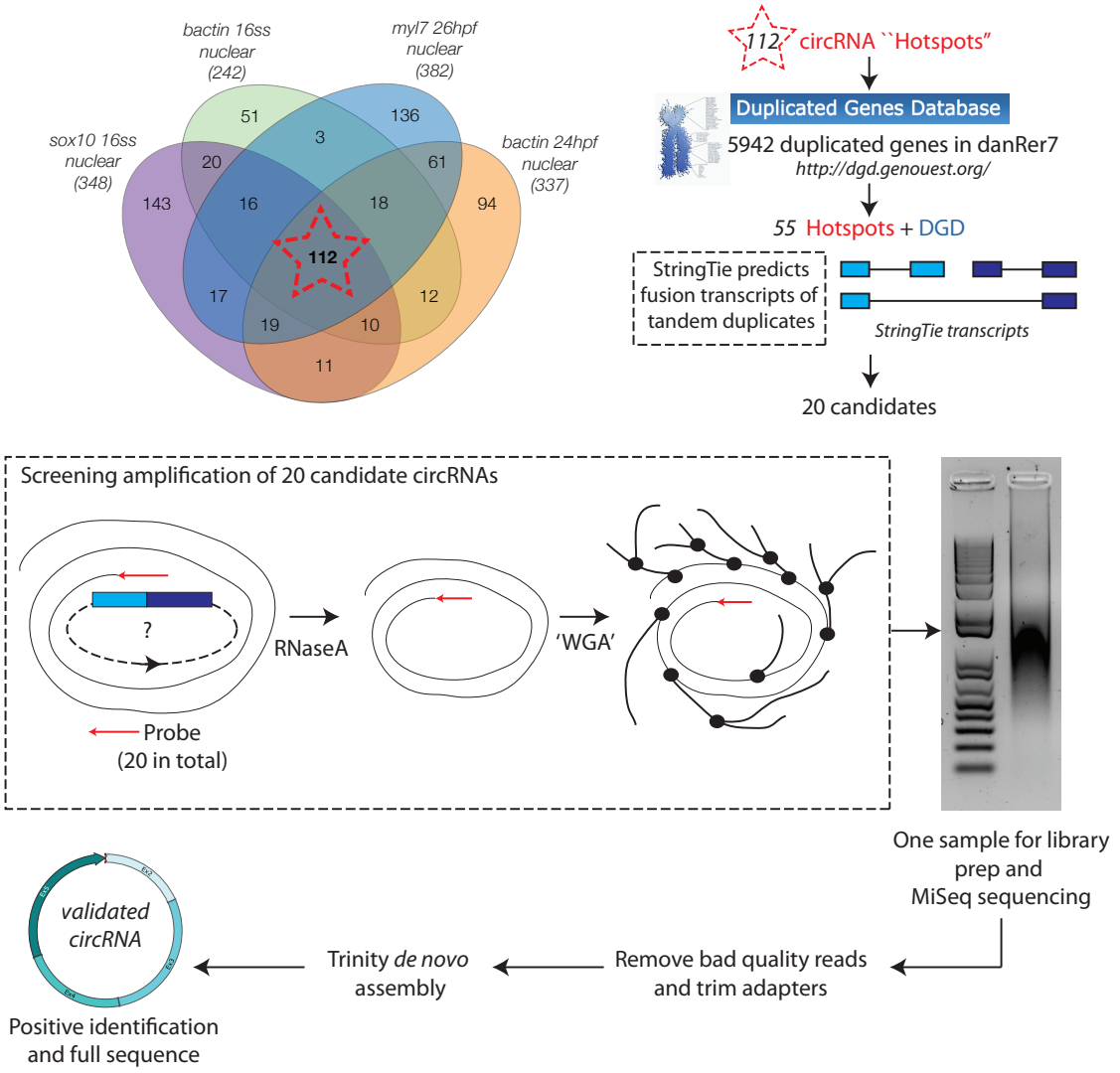


FIGURE 4.3: Validating circularisation "hotspots" enriched at tandem duplicated genes.

Venn diagram shows overlapping and mutually-exclusive putative circRNAs from nuclear compartments of each cell type with their corresponding developmental stages. 112 putative circRNAs found in all datasets ("hotspots", red star) were examined and cross-checked with 5942 zebrafish duplicated genes from the Duplicated Genes Database. 55 "hotspots" were found to overlap regions of duplicated genes. 20 high-confidence candidates were picked based on underlying splicing events between tandem duplicated genes (predicted and assembled using StringTie *de novo* assembly) that may indicate circularisation events between the linear transcripts. A novel "screening amplification" pipeline was established to simultaneously probe for circRNA expression and resolve the full sequence of circRNAs detected.

4.2.4 A new method for multiplexed screening and sequencing of putative circRNAs

Current biochemical methods to validate putative circRNAs by assaying for circularity include RT-PCR followed by inverse PCR (can be paired with qPCR for quantification), Northern Blotting with probes designed to target the circularised junction sequence, and RNaseR sensitivity tests. Another method takes advantage of the high processivity and strand displacement capabilities of reverse transcriptases which can yield cDNAs consisting of concatemers of circRNAs, a phenomenon known as rolling-circle RT. When this phenomenon is observed, characterised by a laddering appearance of PCR products on a gel, it can be strong evidence for circularisation. This was the method-of-choice used to validate neural circRNAs in the mouse (You et al., 2015).

The current methods described involved designing and testing sets of primers or probes individually, a process that is time consuming and limited to the amount of starting material available, especially for detection of rare transcripts. Here, we developed a method that allows multiplexed testing of up to 20 putative circRNAs in one reaction. Briefly, half of an RNA sample is enriched by ribo-depletion (RiboErase, to remove rRNA species but still include circRNAs) and the remaining half is poly(A)-selected (to exclude circRNAs) and focus on mRNA transcripts. Both samples are then reverse transcribed with a highly processive reverse transcriptase like SuperScript IV from ThermoScientific or SunScript™ from Expedeon, using a primer set of 20 oligos DNA probes that are reverse-complement to the exact circularised junctions-of-interest. The reverse transcription is allowed to proceed in conditions conducive to rolling-circle RT (high initial denaturation temperature, long extension time, spiking with extra dNTPs). RNaseA treatment is performed to remove residual RNA after reverse transcription is completed, and cDNA is purified

and concentrated for primer-free unbiased amplication using a whole-genome amplification approach (Picher et al., 2016). The amplified products are then cleaned-up for library preparation and sequenced using MiSeq (v2 300 cycles kit). A bioinformatics pipeline utilising *de novo* transcript assembly and BLAT is used to extract original circRNA template sequences (Figure 4.3). Full details are provided in Chapter 6 Materials and Methods.

Two novel tandem-circRNAs: *circ-cyt1* and *circ-smyhc*

Two assembled transcripts, c119_g1_i1 and c408_g1_i1, were found to overlap duplicated gene regions at chromosome locations chr19:5,975,453 - 5,987,999 (DGD ID Group_1158) and chr24:42,242,905 - 42,325,671 (DGD_ID Group_1489), respectively (Ouedraogo et al., 2012). Both transcripts were detected only in the ribo-depleted amplified sample and not in the poly(A)-selected amplified sample.

To facilitate interpretation of the splicing events that underlie transcript c119_g1_i1 at DGD ID Group_1158 (Figure 4.4A; 5th track from the top in maroon), we identified the single oligo DNA probe (that formed part of the 20-oligo probe pool) targeting a circular junction at this region. A predicted "hotspot" tandem-circRNA resided here, with a circRNA junction that spans from the 3rd exon of *cyt1* to the 2nd exon of *zgc:109868* (Figure 4.4A; blue track). Transcript c119_g1_i1 contained a back-spliced sequence, with full or partial exons from *cyt1* and/or *zgc:109868* (Figure 4.4A; pink dotted box). Due the high sequence similarity between tandem duplicated genes, the origins for certain individual exonic "parts" that form the identified template were inconclusive. Remarkably, we were able to conclude the full sequence of the circRNA (522 bp), as transcript c119_g1_i1 demonstrated a periodicity indicative of rolling-circle RT. This presented strong evidence for circularity of the identified circRNA, which we designated as *circ-cyt1* (Figure 4.4B). *Kallisto* (Bray et al., 2015) quantification estimated the abundance of *circ-cyt1* at $\sim$158 TPM, calculated based on periodicity from transcript c119_g1_i1's abundance at 330 TPM. Based on the alignment of the circRNA junction and *de novo* assembled StringTie transcripts

spanning the tandem duplicated genes (Figure 4.4A; 4th track from the top in red), we noted that most of the exons annotated within the StringTie transcripts were not present, where this "loss/excision" of exons appeared to be demarcated by the circRNA backsplicing event. Based on this "complementary splicing" observation, we hypothesise that the *de novo* assembled transcripts represent linear RNA transcripts formed as a result of circularisation events between the tandem duplicated genes, with a role to remove supernumerary exons and still generate functional protein.

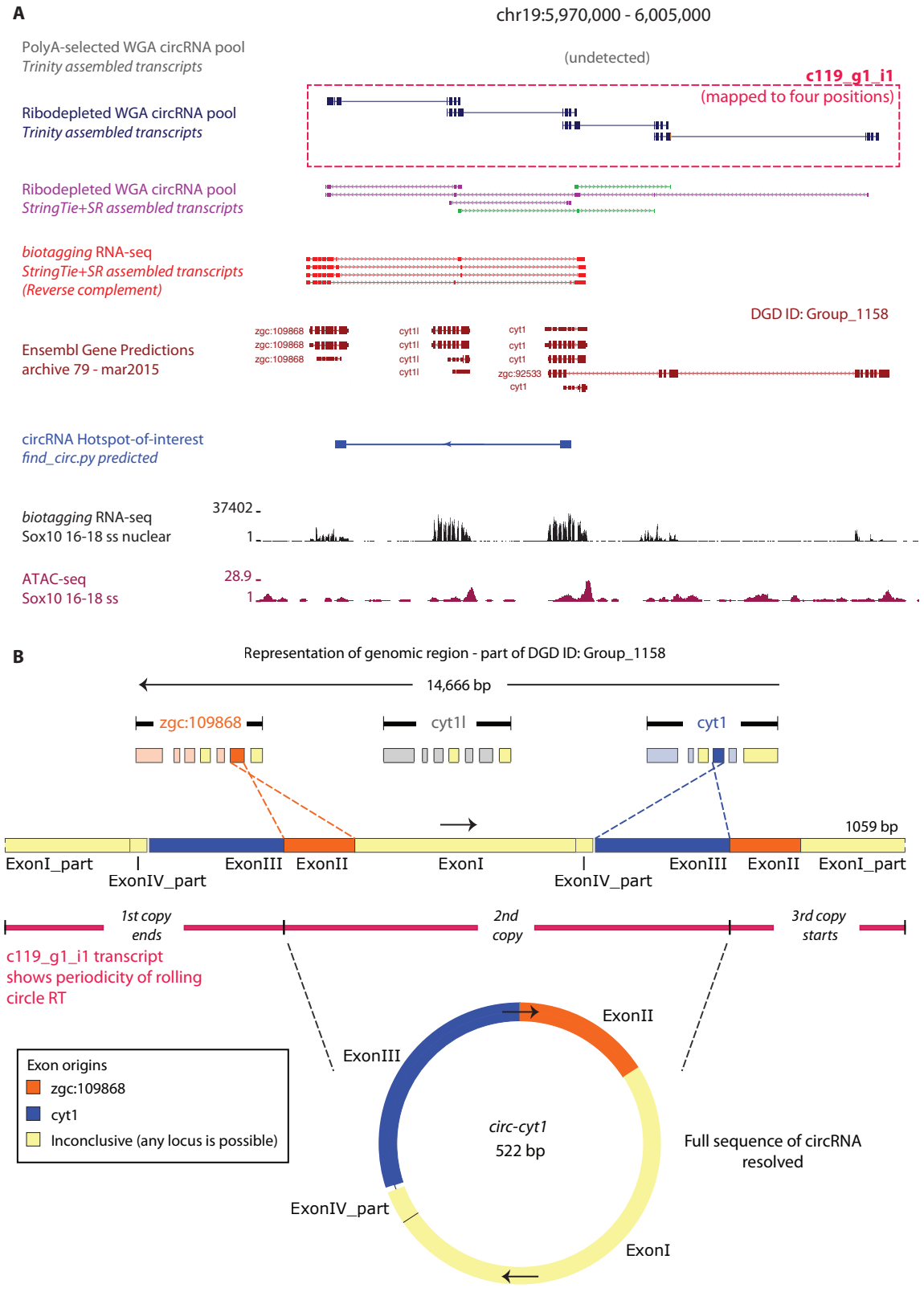


FIGURE 4.4: Identification and characterisation of *circ-cyt1*. Full legend is shown on the next page.

(Please refer to Figure 4.4 on previous page.)

A Genome Browser snapshot of genomic region where the tested circRNA "hotspot"-of-interest resides. Trinity assembled transcript c119_g1_i1 (pink box) was detected only in the ribo-depleted amplified sample but not in the poly(A)-selected sample, and therefore its sequence was extracted for further characterisation. RNA-seq and ATAC-seq tracks provide information on overall gene expression levels and position of the promoters of each tandem duplicated gene. **B** Trinity transcript c119_g1_i1 demonstrates sequence periodicity indicative of rolling-circle reverse transcription, strongly suggesting circular nature of original RNA template. Full map of identified *circ-cyt1* is shown. Exons with inconclusive gene origins (due to sequence similarity between tandem duplicated genes) are coloured pale yellow.

For transcript c408_g1_i1 at DGD ID Group_1489 (Figure 4.5A; 5th track from the top in maroon), we identified seven oligo DNA probes (that formed part of the 20-oligo probe pool) targeting circular junctions at this region. Predicted "hotspot" tandem-circRNAs resided here, with circRNA junctions that span exons between *smyhc1* to exons of *smyhc2*, *smyhc3* or *si:ch211-24n20.3* (Figure 4.5A; blue and red track). Similar to transcript c119_g1_i1, transcript c408_g1_i1 contained a back-spliced sequence, with full or partial exons within *smyhc1* or *si:ch211-24n20.3* (Figure 4.5A; pink dotted box). Again, high sequence similarity between the tandem duplicated genes hindered conclusive origins for each individual exonic "parts" that form the identified template. Transcript c408_g1_i1 also demonstrated a periodicity indicative of rolling-circle RT, thus presenting strong evidence for circularity of the identified circRNA and its full sequence (292 bp), which we designate as *circ-smyhc* (Figure 4.5B). *Kallisto* (Bray et al., 2015) quantification estimated the abundance of *circ-smyhc* at $\sim$79 TPM, calculated based on periodicity from transcript c408_g1_i1's abundance at 162 TPM. Similar to the previous locus described, "complementary splicing" events were also predicted. The identified *circ-smyhc* had contributions from *smyhc1* and possibly a combination of *smyhc2*, *smyhc3* and *si:ch211-24n20.3*.

Full sequences of both *circ-cyt1* and *circ-smyhc* are available in Appendix B.

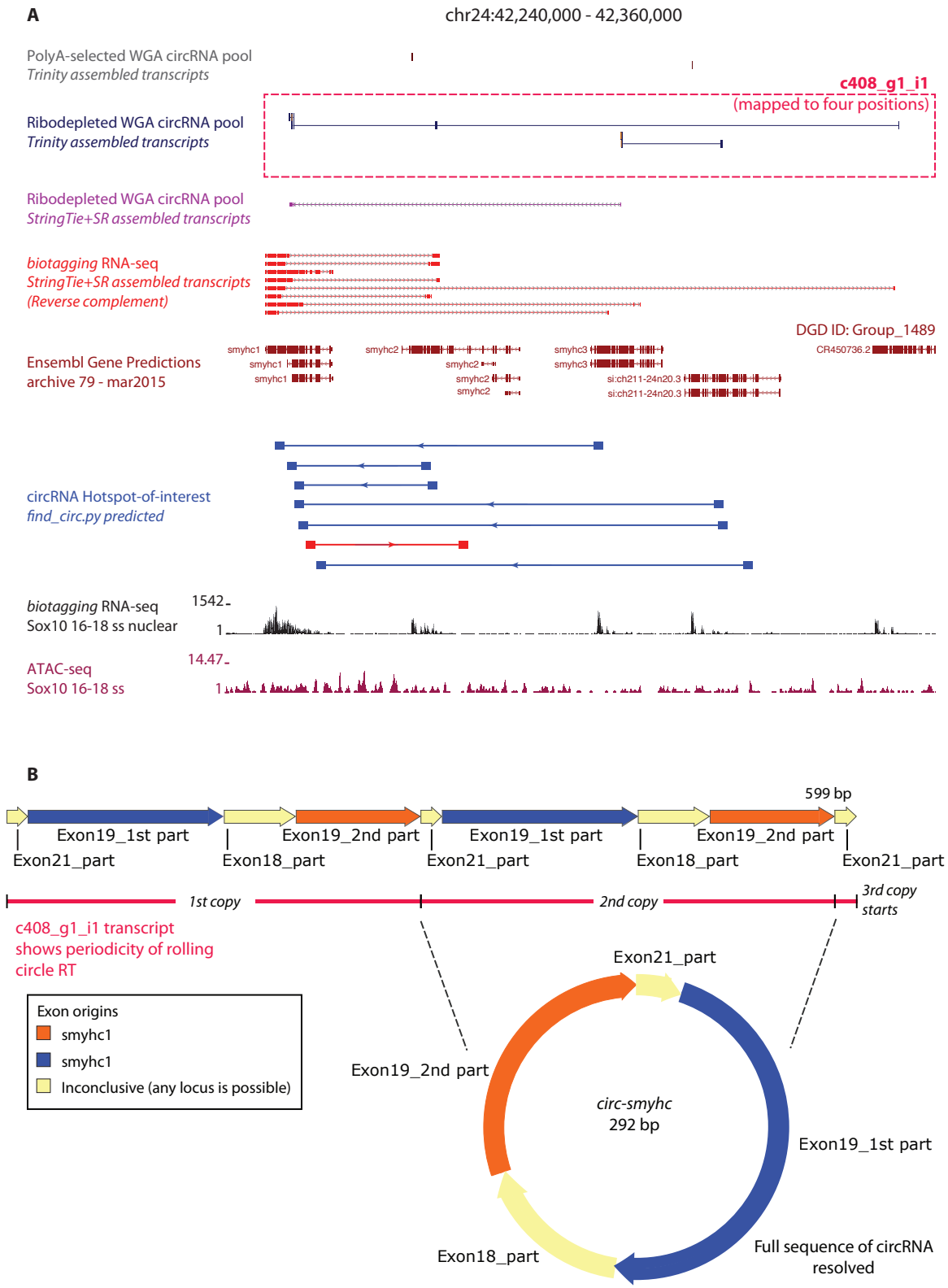


FIGURE 4.5: Identification and characterisation of *circ-smyhc*. Full legend is shown on the next page.

(Please refer to Figure 4.5 on previous page.)

A Genome Browser snapshot of genomic region where the tested circRNA "hotspots"-of-interest resides. Trinity assembled transcript c408_g1_i1 (pink box) was detected only in the ribo-depleted amplified sample but not in the poly(A)-selected sample, and therefore its sequence was extracted for further characterisation. RNA-seq and ATAC-seq tracks provide information on overall gene expression levels and position of the promoters of each tandem duplicated gene. **B** Similar to Trinity transcript c119_g1_i1, c408_g1_i1 demonstrates sequence periodicity indicative of rolling-circle reverse transcription. Full map of identified *circ-smyhc* is shown. Exons with inconclusive gene origins (due to sequence similarity between tandem duplicated genes) are coloured pale yellow.

4.3 Identification of developmentally regulated non-coding RNAs

Strand-specific, compartment-specific transcriptional profiling enables powerful insights into the ncRNA landscape, as ncRNAs often overlap protein-coding regions and their presence in different subcellular compartments indicates different molecular functions.

4.3.1 Long non-coding RNAs (lncRNAs)

Studies attempting to dissect the biology of lncRNAs have highlighted the importance of taking cellular compartmentalisation into account, a factor not considered in previous studies that utilised whole cell methods (Guttman et al., 2011; Gupta et al., 2010; Wang et al., 2011b). Whilst lncRNAs were initially described to be enriched in the nucleus (Chen and Carmichael, 2010; Derrien et al., 2012), with the use of ribosome footprinting in genome-wide studies it became evident that many lncRNAs were also associated with ribosomes (Ingolia et al., 2014; van Heesch et al., 2014).

Our *biotagging* approach allows analyses of subcellular compartments within the same defined cell type, and thus enables us to explore questions involving lncRNA function and localisation. As proof-of-principle, we quantified known zebrafish

lncRNAs (Pauli et al., 2012) in our *biotagging* datasets to identify developmentally regulated lncRNAs between different cell types and subcellular compartments. Differential expression analysis using DESeq2 revealed 51 differentially expressed lncRNAs (*padj*-value<0.05) in the *myl7 nuclear* versus *bactin2 nuclear* population at 24-26 hpf. Similarly, we found 111 differentially expressed lncRNAs in *sox10 nuclear* versus *bactin2 nuclear* at 16-18 ss. Comparing *sox10 nuclear*- and *myl7 nuclear*-specific lncRNAs revealed only 14 overlapping lncRNAs, suggesting cell type divergence in the expression of lncRNAs and their corresponding functions (Figure 4.6). Full lists of differentially expressed lncRNAs in the nuclear transcriptome are available in Appendix A.

In contrast, only 3 lncRNAs were found to be differentially expressed when comparing *sox10 ribo* versus *bactin2 ribo* at 16-18 ss. This suggested that although lncRNAs can be found on ribosomes, developmentally regulated lncRNAs are more likely to be enriched in the nucleus, a finding which is consistent with current evidence in the literature (Engreitz et al., 2016b).

(Please refer to Figure 4.6 on the next page.)

A Heatmap of differentially regulated zebrafish lncRNAs in *sox10 nuclear* versus *bactin2 nuclear* at 16-18 ss (left), and *myl7 nuclear* versus *bactin2 nuclear* at 24-26hpf (right). Venn diagram showing only 14 overlapping lncRNAs between *sox10 nuclear* and *myl7 nuclear* differentially regulated lncRNAs. **B** Transcription across annotated transposable elements (TEs) are developmentally regulated in both cell type- and compartment-specific manner. Lowly-expressed TEs (left) appear to remain in the nucleus. TEs that were transcribed at high levels but detected mostly in the polysomal compartments and TEs highly transcribed only in the developing myocardium (right).

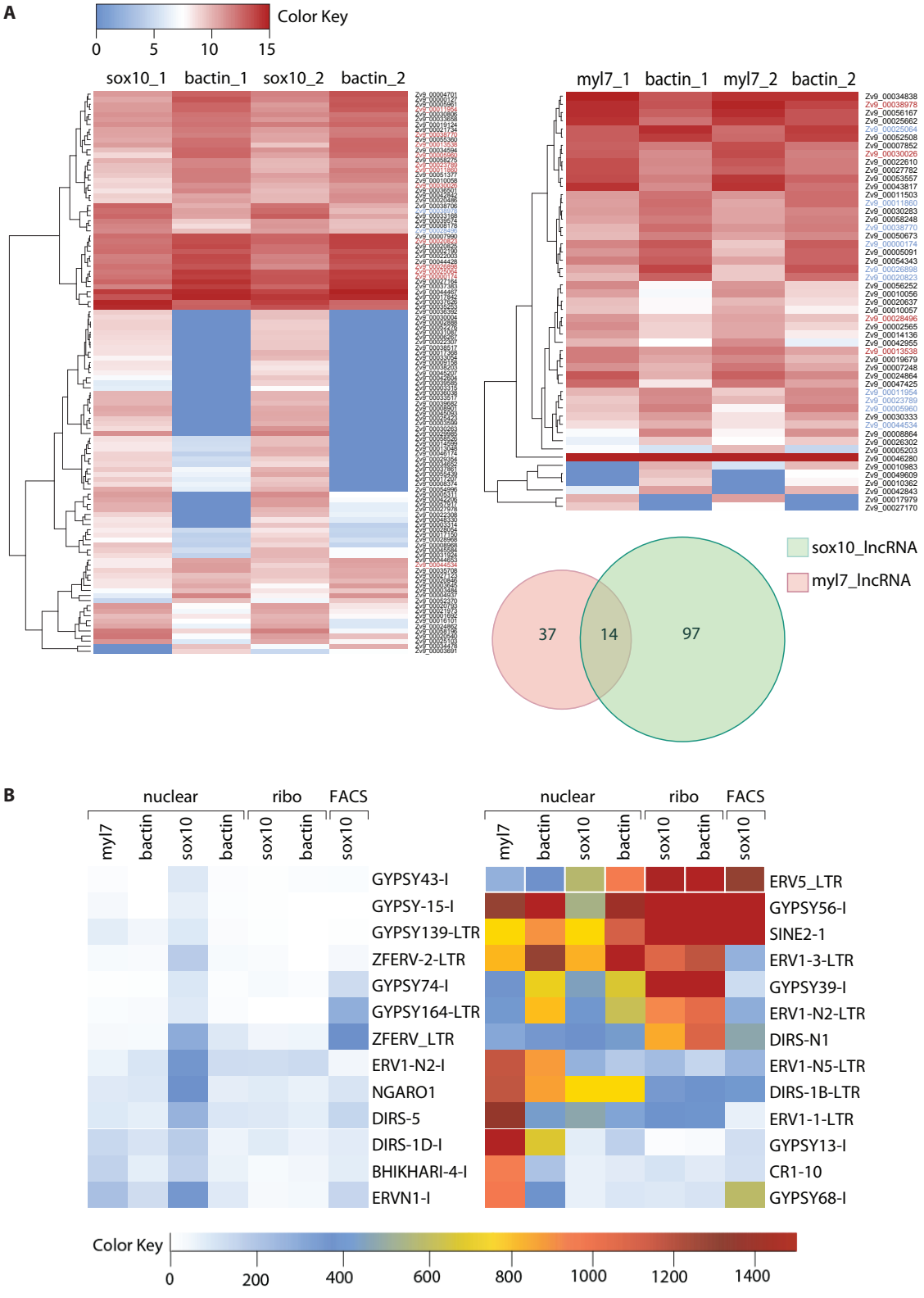


FIGURE 4.6: Identification of developmentally regulated non-coding RNAs. Full legend is shown on previous page.

4.3.2 Transposable elements

In light of studies that suggested that transposable elements (TEs) are involved in the rewiring of gene regulatory interactions during development, we next sought to quantify expression of transcribed TEs in our *biotagging* datasets. (Gifford et al., 2013; Sundaram et al., 2014). Given that >50% of the zebrafish genome consists of type I and type II TEs (Howe et al., 2013), we surveyed our cell type- and compartment-specific datasets for activity that could suggest developmental regulation. TE analysis was performed as previously described (Göke et al., 2015). Differential expression of annotated classes of zebrafish TEs performed using rank product non-parametric method revealed that a number of TEs were expressed in a tissue-specific fashion and encompassed a broad spectrum of expression levels. Heatmaps demonstrated that certain classes of lowly-expressed TEs (*ERVN1-I*, *ERV1-N2-I*, *NGARO1*, *ZFERV-2-LTR*) were enriched in *sox10 nuclear* compared to corresponding stage-matched *bactin2 nuclear*. These were not found in *sox10 ribo* or *bactin2 ribo*, suggesting they are transcribed but not exported from the nucleus and therefore may be functioning as enhancers (Pavlicev et al., 2015). We also uncovered a set of TEs that were transcribed at high levels but also detected in the polysomal compartments (*ERV1-N2-LTR*, *DIRS-N1*, *GYPSY39-I*), suggesting they could be co-regulated/co-localised with protein coding genes they associate with, possibly as parts of mRNA transcripts. A third group of TEs (*CR1-10*, *GYPSY13-I*, *GYPSY68-I*, *ERV1-1-LTR*) were highly transcribed only in the developing myocardium. Taken together, these results demonstrated that transcription of TEs is developmentally regulated in both a cell type and subcellular compartment manner.

Testing enhancer activity of transposable elements in zebrafish

We identified three putative enhancers downstream of a known NC master regulator, *tfap2*α (Rada-Iglesias et al., 2012; Wang et al., 2011c), whose transcription

was enriched in *sox10 nuclear* versus *bactin2 nuclear*. These putative regulatory elements on one hand belong to the highly bidirectionally transcribed κ-means cluster 3 highlighted in the previous chapter, but also overlap or flank annotated TEs. At the stages that we observed (14 ss - 24 hpf), Zfin-curated and published WISH data clearly highlights the expression of *tfap2*α first in migrating NC cells (both cranial and trunk) followed by expression in the hindbrain and in the branchial arches (Knight et al., 2003). *Tfap2*α *enh 2* had clear activity in the migrating cranial NC with stronger signal in the NC cells surrounding the otic vesicle at the level of rhombomere 5, which was reproducible in separate embryos (Figure 4.7; top, white arrows). *Tfap2*α *enh 3* had scattered activity in the trunk (Figure 4.7; middle), likely highlighting spinal NC (Knight et al., 2003). *Tfap2*α *enh 7* had weak expression at the earlier stages but strong activity at 24 hpf in the hindbrain as well as the branchial arches (Figure 4.7; bottom, white arrows). In short, all three TEs showed varying levels of enhancer activity with specific but different patterns of expression reminiscent of endogenous *tfap2*α expression. This finding warrants further investigation and dissection of each element in order to elucidate the role of these TEs-containing enhancers in controlling *tfap2*α expression in the embryo and in driving NC development.

Enhancer	Position (danRer7)	Forward primer	Reverse primer
Tfap2α enh 2	chr24:8486307-8488160	GGACGTTGCAACATAGGCTC	AGCCTGAGAGTGAGCTTCTG
Tfap2α enh 3	chr24:8481967-8484300	ATGATTAATAGTCTTAAATAACTTGCACTTC	GACCCACAGCCTGCATTAAA
Tfap2α enh 7	chr24:8316164-8317205	CTAAAGGCGCTGTTGACATGA	GGCAACGAGTTTACCCAATTAA

TABLE 4.2: Chromosome positions and primer sequences of *tfap2*α enhancers from **ATAC_enhancer** κ-cluster 3 in previous chapter that contain transposable elements.

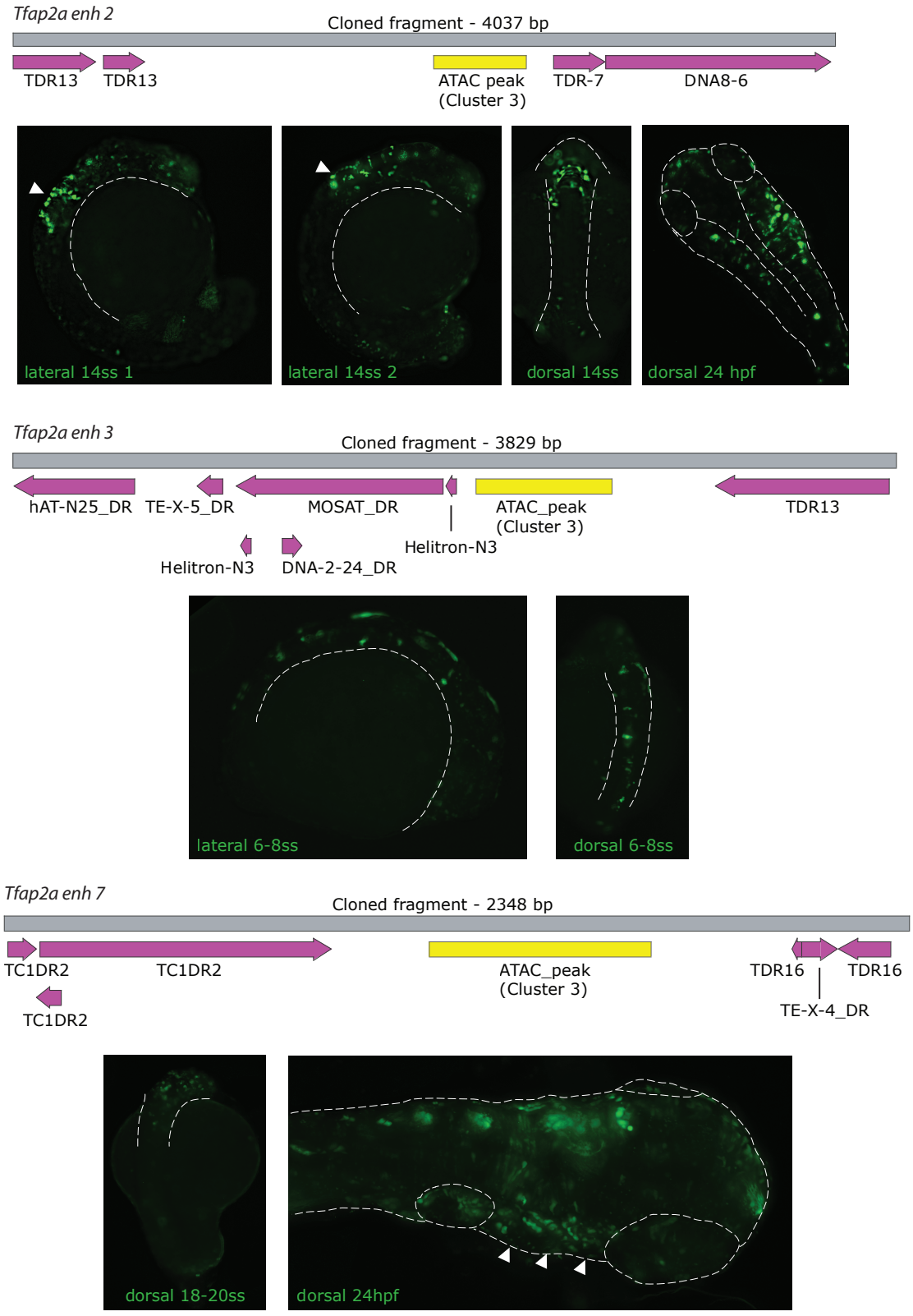


FIGURE 4.7: Enhancer activity of *cis*-regulatory transposable elements in zebrafish. Full legend is shown on next page.

(Please refer to Figure 4.7 on previous page.)

Schematic representation of three *tfap2α* putative TEs-containing enhancers from *κ*-cluster 3 in previous chapter that were tested. Microinjected embryos show *Tfap2α enh 2* activity in migrating cranial NC, with stronger eGFP signal in rhombomere 5 (top panel, white arrows). *Tfap2α enh 3* has scattered activity in the trunk NC (middle panel). *Tfap2α enh 7* has stronger activity at a later developmental stage with signal in the hindbrain regions and in the branchial arches (bottom panel, white arrows). Cloned fragments range between 2 and 4 kb, therefore requiring further characterisation to determine core region(s) of each enhancer. TEs found within the cloned enhancers include annotated repeat elements (purple arrows) consisting of Class II DNA transposons (TDR, DNA, hAT, Helitron and TC families). MOSAT and TE-X-5 are from a separate class of annotated repeat elements - satellite DNA.

4.4 *In vivo* tools for studying gene regulatory networks II

To dissect the function of non-coding RNAs identified in our *biotagging* datasets, we sought to build a reliable system to manipulate their expression *in vivo*. Several methods to interfere with gene function currently exist, such as morpholinos, TALENs, and CRISPR/Cas9. These experiments can be performed transiently via microinjection into one-cell stage embryos with the respective reagents that are available commercially or easily synthesised *in vitro* in the laboratory. The caveat to such experiments is the lack of spatiotemporal specificity as it can be exceedingly difficult to distinguish between a cell-specific phenotype and secondary effects resulting from a ubiquitious knockdown/expression of the gene-of-interest. In addition to this, targeting ncRNAs presents another set of challenges to consider as they often overlap protein-coding genes on the opposing strand. As such, care must be taken to only interfere the non-coding gene while avoiding the underlying protein-coding gene in order to interpret its functional significance in isolation from its protein-coding genomic partner.

4.4.1 CRISPR/Cas9-based *interference* or *activation* of gene expression

CRISPR/Cas9-based *interference* or *activation* (CRISPR*i* or CRISPR*a*) utilises nuclease-deficient Cas9 (dCas9) fused to transcriptional regulator domains which are targeted to specific genomic regions using "guide RNAs" (sgRNAs). To inhibit transcription, dCas9 can be targeted to transcription start sites (TSSs) of genes to inhibit RNA Polymerase II by steric hindrance (Qi et al., 2013), or fused to effector domains such as Kruppel-associated box (KRAB) (Gilbert et al., 2013) and four concatenated mSin3 repressive domains (SID4x) (Konermann et al., 2013) to induce chromatin changes inhibitive of transcription. KRAB functions by recruiting histone "writers" for H3K9me3, H3K36me3 and H4K20me3, as well as its co-repressor protein, HP1, and together they constitute a heterochromatin-forming complex that stalls transcription at the targeted region (Sripathy et al., 2006; Thakore et al., 2015). mSin3 domains were shown to function as potent transcriptional repressor complexes together with Mad proteins, by recruiting histone deacetylases such as HDAC1 and 2 (Ayer, 1999; Cowley et al., 2004). To activate transcription, dCas9 is fused to the well-characterised VP64 effector consisting of four concatenated VP16 domains (Mali et al., 2013). Crucially, as CRISPR*i* and CRISPR*a* complexes are directed to the TSS of target genes, this allows the perturbation of gene expression without modifying the underlying endogenous locus sequence, making it a well-suited and powerful tool to study the function of non-coding genes (Thakore et al., 2015; Liu et al., 2016; Konermann et al., 2015). It is important to note, however, that CRISPR*i*/CRISPR*a* success stories reported thus far were the result of experiments done in cell culture. Recent work in our lab has demonstrated promising results for adapting CRISPR*i* and CRISPR*a* in the chicken embryo (Williams et al., 2017). Here, we developed a similar approach while taking advantage of the genetic amenability of zebrafish to build a high-precision system without the caveats that often accompany transient experimental designs.

A genetic CRISPR/Cas9-based system for epigenome engineering

Two transgenic lines were built using BAC transgenesis - *TgBAC(Sox10:dCas9-SID4x-2a-Citrine)*^{ox117} and *TgBAC(Sox10:Cas9m4-VP64-2a-Citrine)*^{ox118}, expressing repressive and activating dCas9 fusions in all NC cells under the control of the Sox10 regulatory elements . Whole-mount immunohistochemistry confirmed the co-localisation of dCas9 with Sox10 proteins in the transgenic individuals (Figure 4.8). These lines ensure tissue-specific delivery of nuclease-deficient Cas9s fused to repressor or activator domains, respectively, together with a Citrine fluorescent reporter to facilitate screening. Currently, the most commonly used method to deliver sgRNAs in zebrafish is via microinjection of *in vitro*-transcribed sgRNAs. While simple and cost-effective, this approach risks decreasing the efficiency for CRISPR-mediated events in the embryo, as the nature of sgRNAs (uncapped and without poly(A) tails) renders them unstable and prone to degradation *in vivo* (A. Bassett; personal communication). In fact, when injected in the absence of Cas9 protein or Cas9 mRNA, unprotected sgRNAs degrade in a couple of hours and for our purpose would be completely absent by the time dCas9 fusions are beginning to be expressed at $\sim$10 hpf. To circumvent this problem, we cloned a cassette containing a zebrafish-specific U6a promoter (Yin et al., 2015) driving expression of sgRNAs into a custom-made *Ac/Ds* mini-vector. The 20 bp spacer region within the sgRNA (the only variable region that determines target specificity within the genome) is flanked with *Bsm*BI restriction sites to facilitate Golden Gate cloning of different sgRNA targets (Clarke et al., 2012). The small size of this specially-designed construct (<4.5 kb), coupled with small load required for activity (30 pg per embryo) permits multiplexing as well as "transient transgenesis" of sgRNAs, ensuring continued expression of sgRNAs until the tissue-specific Cas9 protein is expressed and functional. Modified versions of this cassette containing special RNA scaffolds for sgRNA-mediated KRAB transcriptional repression (Zalatan et al., 2015) and for the synergistic transcription

activation mediator (SAM) system (Konermann et al., 2015) have also been generated. As proof-of-principle, we are currently testing and optimising our system with F_1 generation transgenic fish, by targeting endogenous genes in the *sox10* GRN for ectopic and untimely activation (using VP64 line), as well as for diminished expression (using SID4x line). Once optimised, this system will be utilised for functional screens of developmentally regulated non-coding genes identified and discussed throughout the course of this thesis.

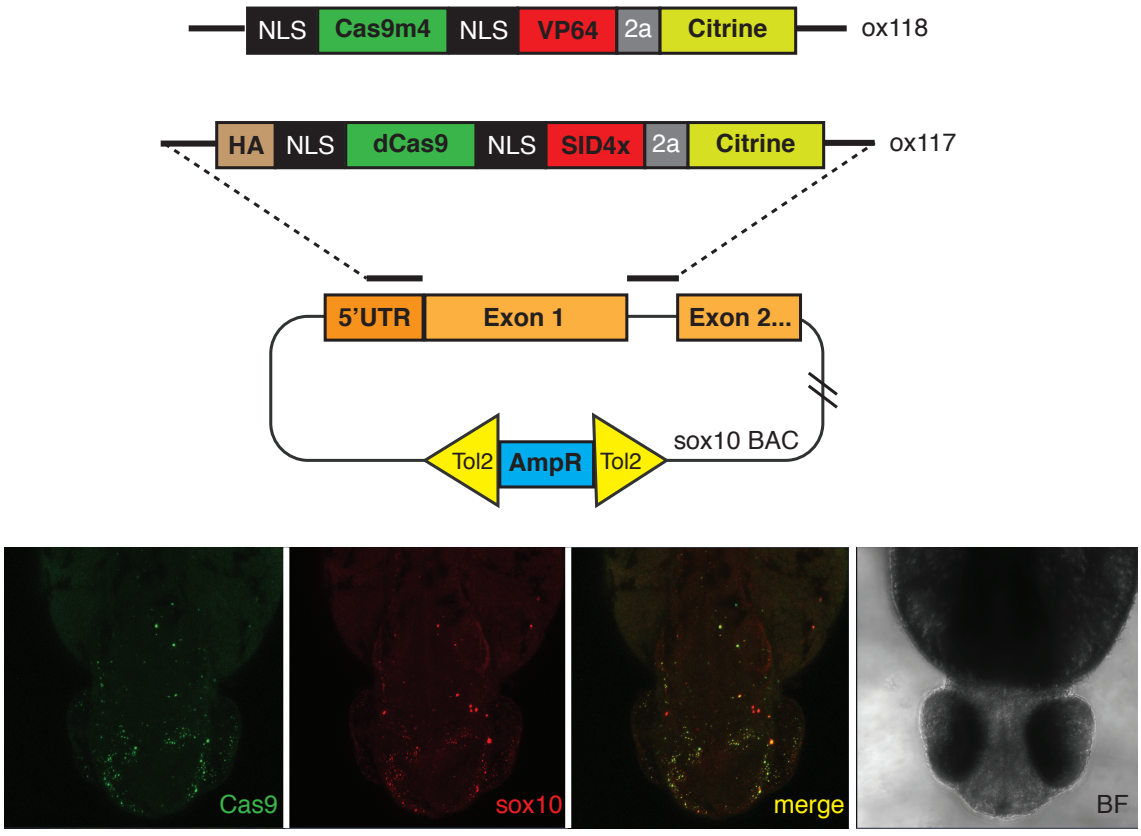


FIGURE 4.8: Tissue-specific Cas9 transgenic lines for epigenome engineering.

Schematic representation of BAC recombineering to generate *TgBAC(sox10:dCas9-SID4x-2a-Citrine)*^{ox117} and *TgBAC(sox10:Cas9m4-VP64-2a-Citrine)*^{ox118} for *sox10*-specific modulation of transcription without modifying the underlying genomic sequence. Whole-mount immunohistochemistry (IHC) in 3dpf F_1 embryos of ox117 demonstrates co-localisation of endogenous Sox10 (Alexa 568; red) and nuclease-deficient Cas9 (Alexa 488; green). These IHCs were performed by my student, Dervla Carroll.

4.5 Discussion

Mining of our *biotagging* datasets has presented a whole new world of ncRNAs whose expression and biogenesis are developmentally regulated in the embryo (lncRNAs and TEs), or provide a sneak peek into an uncharacterised mechanism during the transcription of tandem duplicated genes in the zebrafish (circRNAs).

Detecting circRNAs. Northern Blotting is disadvantageous if the circRNA-of-interest is of low abundance, as it could lead to false negatives. While circRNAs should theoretically be insensitive to RNaseR 3′ to 5′ exonuclease activity which therefore can be used to enrich for circRNAs in an RNA sample, RNaseR has been shown to be counter-productive and could decay certain RNA circles while at the same time not efficiently degrading all linear transcripts. On the other hand, inverse PCR primers are designed such that their 3′ ends face away from each other when aligned to the genome, therefore excluding non-backspliced amplicons whose exon-bound primers must face towards each other in the cDNA transcriptome. While this approach is relatively simple to perform, false positives may still occur due to template-switching during reverse transcription, trans-splicing between two pre-mRNA transcripts or genomic duplication events. In the first scenario, the reverse transcriptase "skips" between two copies of the same transcript (or between two different transcripts in the case of the second scenario), thus generating a cDNA product with abnormally-oriented exons that mimic circRNA backsplicing. In the third scenario, a genome duplication event could have yielded tandem arrangement of one or more exons within the same locus whose resulting mRNA can be templated during inverse PCR, showing exons that appear to be scrambled (Barrett and Salzman, 2016). Our rolling-RT-based method described in the results above bypasses these cons and provides a "one-step" screen-and-sequence approach to validate putative circRNAs.

Possible implications of tandem-circRNAs. Despite numerous evidence highlighting aspects of circRNA biogenesis (summarised in Chapter 1, Section 2.5.1), elucidation of their biological, functional roles have been murky at best. *CDR1as* and *circ-mbl* have been shown to regulate gene expression of their "host" genes by sponging miRNAs or autoregulating the host protein, respectively (Memczak et al., 2013; Ashwal-Fluss et al., 2014). It is currently unknown if these represent general functions for circRNAs, or simply singular examples of circRNA function (Salzman, 2016). To our knowledge, our results are the first report of a cohort of circRNAs with an enriched overlap at tandem duplicated genes. Our observation could be a species-dependent effect, as tandem duplication events have been reported to be more prevalent in the zebrafish genome when compared to three other teleost species (Lu et al., 2012). The "complementary splicing" phenomenon that we have observed between circRNAs and their underlying *de novo* assembled transcripts point towards a possible link between the mechanism of their formation and modulation of the splicing machinery during transcription of tandem duplicated genes. This is supported to a certain extent by previous studies demonstrating strong competition between circularisation and formation of mature mRNA transcripts (Ashwal-Fluss et al., 2014) as well the importance of splicing signals (whether canonical or not is still being debated) in circRNA biogenesis (Gao et al., 2016; Starke et al., 2015). However, it is also important to note the following caveat; due the high sequence similarity of tandem duplicated genes, their annotation as such in the reference genome may not entirely reflect the transcriptome of the experimental samples. They may actually be misrepresented copy number variants, leading to the inclusion of multi-mapped non-unique reads as predicted circRNA junctions.

Co-option of TEs during enhancer evolution. We identified three enhancers downstream of a known NC master regulator, *tfap2α* (Rada-Iglesias et al., 2012; Wang et al., 2011c), whose transcription was enriched in *sox10 nuclear* versus *bactin2 nuclear*, peppered with TE elements and have biological activity *in vivo*. Our results

present a tantalising insight into the possible co-option of TEs into enhancers driving NC development. In the stickleback, an insertion of a TE during evolution, acting as a tissue-specific enhancer in the freshwater species has been shown to contribute to the gain-of-function of *GDF6* gene, resulting in dramatic differences in skeletal structures, when compared to ancestral marine species (Indjeian et al., 2016). The co-opting of TEs during evolution to function as enhancers is not a new phenomenon, and has been shown to occur even in the human genome (Pavlicev et al., 2015).

Aided by the development of powerful tools such as CRISPR/Cas9, we are now in the position to actively tackle multiple avenues of uncharted territory in the study of genetic pathways that regulate embryonic development, that require scientific investigation beyond the scope of this thesis and therefore will be discussed elsewhere in the future.

Chapter 5

Concluding remarks and future work

Embryology has fascinated human civilisation since ancient times. Scientific thought and methodology through the centuries have contributed in leaps and bounds to important milestones in the understanding of how embryonic development proceeds, leading to the modern incarnation of developmental biology as we know it today. From embryological techniques to advances in molecular and cellular biology, our understanding of animal development has reached an impressive threshold albeit an incomplete picture.

In many ways, the ushering of the -omics era has revolutionised developmental biology as a field. As the lists of sequenced genomes and computational tools continue to grow, developmental and evolutionary biologists alike are presented with a tantalising "Pandora's Box", filled with secrets of life waiting to be unlocked at a pace that could only be dreamed of by our predecessors. Here, at the forefront of modern developmental genetics and genomics, we present our body of work using a combination of the old and the new to deepen our understanding of genetic programmes encoded within the genome, underlying neural crest and myocardium development.

The neural crest is a transient and multipotent population of cells invented early in vertebrate evolution. It is a versatile system for interrogation, as the genetic machinery underlying its ontogeny and subsequent differentiation into seemingly unrelated tissue types recurs throughout development and in disease. This well-oiled machinery is also sensitive to fine-tuned controls; disrupt a cog and development fails to proceed normally (e.g. neural crest defects underlying neurocristopathies

account for roughly 1/3 of all birth defects), overrun the system when it should not and disease onsets (e.g. metastasis and formation of melanoma). In order to discover ways to prevent or treat these abnormalities, we first need to fully understand what the "normal" is, at the level of genes within the context of a highly dynamic genome.

Our *biotagging* approach presents an opportunity to dissect these mechanisms at subcellular resolution within a cell-type of interest. Isolating polysomes and their associated mRNAs informed us of the specific proteins being made within a defined developmental window. At 16-18 ss, cranial neural crest cells are actively migrating and undergoing differentiation while trunk neural crest cells are still in a premigratory state. We recovered both known and novel protein markers for the neural crest at this stage. For example, the upregulation of *elavl3* in our migratory neural crest translatome dataset suggests that these cells are actively differentiating into trigeminal sensory neurons at this time. Genes without characterised roles in the neural crest or in neural crest-derived structures were also upregulated (e.g. *rbck1*, *egr1*, *npm1a*), highlighting potential lines of investigations to elucidate their biological roles in neural crest development. On the other hand, by isolating nuclei and their transcriptomes at the same developmental stage in the same cell type, we demonstrated that the molecular logic for migratory neural crest cells is encoded within the non-coding genome at the level of transcribed enhancers (eRNAs) and promoters, even if their corresponding genes were not being translated as yet. We also recovered developmentally regulated long non-coding RNAs, transcribed transposable elements and circular RNAs and presented the *biotagging* approach to tissue-specifically isolate chromatin regulators and their genomic binding regions, an approach we are currently pursuing, in order to complement our current results.

The stark contrast between the two pictures painted from within the same cells raises important questions relating not only to neural crest development as a whole, but also in our interpretation of how development proceeds at the cellular level within the framework of their molecular identity. Traditionally, cellular identity

has perhaps been defined over-simplistically via the expression of protein markers that fall within a particular gene regulatory network. We are now in the position to expand this definition by not only taking into account what *protein(s)* and where *within the organism* these players are involved, but also what *non-coding element(s)* and where *within a given cell type* these new players are exerting their functions. In the neural crest in particular, this will be exceedingly tricky given how its state transitions from being a stem cell-like population to many subpopulations committed to different fates. Further studies such as ours will be needed at a larger scale to include these different stages, and perhaps to also characterise other vertebrate systems, in order to fully tease apart this highly conserved phenomenon.

Tackling the non-coding genome has also uncovered non-coding RNA molecules that form part of the genetic regulation underlying cellular function. Previous work by many research groups has highlighted long non-coding RNAs as molecular scaffolds that shuttle proteins to their target genomic regions to regulate gene expression. Enhancer-derived RNAs (eRNAs) not only serve as "indicators" of when gene transcription onsets, but also have been proposed to facilitate chromosome-looping between *cis*-regulatory elements and the gene promoters they regulate. Transposable elements have emerged as evolutionary relics that have been re-purposed to drive the evolution of genes as species diverged from their predecessors. Circular RNAs have also benefited from the -omics revolution, revealing their omnipresence within the transcriptomes of organisms under study. The understanding of potential roles and functions of circular RNAs is very much still in its infancy, with isolated examples of circRNAs functioning as miRNA (a small non-coding RNA species) sponges or autoregulator of proteins. Here, we add to the mystery by proposing yet another class of circRNAs enriched at tandem duplicated genes in the zebrafish, with possible implications of their formation in the trans-splicing between their underlying tandem mRNA transcripts, in order to ensure production of proper protein products. Further work is needed to test our hypothesis, as we have yet to demonstrate if disrupting their formation can abrogate the trans-splicing we have observed

between these tandem transcripts.

Finally, we would like to highlight the importance of *in vivo* tools that will allow efficient interrogation of the genome to address pressing biological questions. *In vitro* methods are useful but may not recapitulate *in vivo* results obtained from genetic studies. Fully genetic studies, although rewarding, can often be time-consuming. Transient *in vivo* approaches bridge this gap, as presented by two of our toolkits for zebrafish research. By taking advantage of the highly efficient and low background features of the *Ac/Ds*-mediated integration system, we have built tools to reliably screen putative enhancers for biological activity *in vivo*, and to modulate transcription *in vivo* without altering the genomic sequence (thus avoiding spurious side effects).

In conclusion, we hope that our work not only presents new avenues of investigations into the neural crest, but will also challenge the reader to revisit how we study developmental biology as a whole. We are now ushering a new generation of scientists willing to embrace the exponential growth of both computational and wet lab methods at their disposal - the future is bright.

Chapter 6

Materials and methods

6.1 Zebrafish husbandry

All zebrafish experiments were conducted according to regulated procedures authorised by the UK Home Office within the framework of the Animals (Scientific Procedures) Act 1986. Wild type and transgenic embryos were derived from AB or AB/TL mix strains.

6.2 Plasmids

All constructs and plasmids used in this thesis are available on Addgene:
https://www.addgene.org/Tatjana_Sauka-Spengler/

6.3 Microinjection of zebrafish embryos

For maximal success, all microinjections were performed by injecting into the blastomere of one-cell stage embryos within 20 minutes of laying. Specific details and conditions for different experiments are described below.

6.3.1 Generation of *tol2* plasmid transgenic lines

Cloned plasmid constructs containing transgene cassettes flanked with *tol2* integration sequences were injected using 80 pg to 100 pg of DNA and 40 pg of *in-vitro* transcribed *tol2* mRNA per embryo. *Tol2* was transcribed from pCS2-Transposase

vector (Kwan et al., 2007) (linearised beforehand with *Asp*718) using mMESSAGE mMACHINE™ SP6 Transcription Kit (AM1340, ThermoFisher Scientific). F_0 embryos were raised and screened for founders. Positive F_1 embryos were raised for future experiments.

6.3.2 Generation of BAC transgenic lines

Donor constructs containing desired transgenes were cloned and used as PCR templates to generate linear recombination cassettes containing 50 bp homology arms for BAC recombination. Recombineering into desired BAC was performed either by replacing first coding exon with BirA cassette (for *BirA driver* lines) or by in-frame replacement of the start codon with AviTag™-containing cassette (for chromatin regulator *Avi effector*s), followed by addition of *tol2* integration arms as described previously (Abe et al., 2011). BACs were injected using 200 pg of DNA and 40 pg of *in-vitro* transcribed *tol2* mRNA per embryo. F_0 embryos were raised and screened for founders. Positive F_1 embryos were raised for future experiments.

6.3.3 *Ac*/*Ds*-mediated enhancer screening assay

Elements-of-interests were amplified from genomic DNA and cloned into the *Ac/Ds* enhancer construct, pVC-Ds-E1b:eGFP-Ds (linearised wtih *Nhe*I), using In-Fusion™ HD Cloning (638909, Takara Bio). Constructs were injected using 30 pg of DNA and 8 pg of *in-vitro* transcribed *Ac* transposase mRNA per embryo. *Ac* transposase was transcribed from pAC-SP6 vector (Emelyanov et al., 2006) (linearised beforehand with *Bam*HI) using mMESSAGE mMACHINE™ SP6 Transcription Kit (AM1340, ThermoFisher Scientific). For transient screening purposes, 50-100 embryos were injected and observed at desired developmental stages. If desired, F_0 embryos can be raised and screened for founders to generate stable transgenic lines.

6.4 *Biotagging*: Isolation of subcellular compartments using Streptavidin beads

This section is adapted from Supplemental Experimental Procedures of my co-first author paper (Trinh et al., 2017).

6.4.1 Preparation of Streptavidin beads for RNA procedures

250 µg Streptavidin Dynabeads™ were separated from the supplied supernatant using a magnetic stand (DynaMag™-2, 12321D, ThermoFisher Scientific). The supernatant was discarded and beads were washed twice with 1 mL of Solution A (DEPC-treated 0.1 M NaOH, 0.05 M NaCl) by nutating for 3 minutes each time followed by at least 3 minutes on the magnetic stand. The beads were subsequently washed once in Solution B (DEPC-treated 0.01 M and then re-suspended in Nuclear Pulldown Buffer (NPB, for nuclei isolation) or Cell Lysis Buffer (for polysomal isolation). Beads were left nutating in the cold room until nuclei/polysomal suspensions were ready. When ready to use, the beads were captured on the magnetic stand and the supernatant removed before nuclei/polysomal suspensions were added. Dynabeads™ M-280 Streptavidin (11205D, ThermoFisher Scientific) were used for nuclear isolation and Dynabeads™ MyOne™ Streptavidin T1 (65601, ThermoFisher Scientific) were used for polysomal isolations.

6.4.2 Nuclei isolation

The following protocol was adapted from the INTACT method (Deal and Henikoff, 2011) and optimised for highest yield of nuclei per zebrafish embryo. Zebrafish embryos were obtained via crosses between selected *Avi effector* and *BirA driver* lines. A maximum of 350 embryos (24 hpf or younger) were used per biological sample. Embryos were dechorionated and washed in hypotonic Buffer H (20 mM HEPES

(pH 7.4), 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT and 1X cOmplete^{TM} protease inhibitor) supplemented with 0.01% Tricaine. Embryos were resuspended in Buffer H at a **ratio of 50 embryos per 1 mL of Buffer H** and transferred to a Dounce homogenisor (available in different sizes at Sigma-Aldrich as Dounce tissue grinder sets). Embryos were dissociated using 20 strokes of Pestle A and incubated on ice for 15 minutes. This was followed by 60 strokes of Pestle B (3 x 20 strokes with 5 minutes pause in between). Nuclei were collected by centrifugation at 2,000 *g* for 10 minutes at 4 °C and resuspended by vigorous tapping (not vortexing) in NPB (10 mM HEPES (pH 7.4), 40 mM NaCl, 90 mM KCl, 0.5 mM EDTA, 0.5 mM spermidine, 0.15 mM spermine, 1 mM DTT and 1X cOmplete^{TM} protease inhibitor). The suspension was incubated with Streptavidin beads prepared for RNA procedures as described above for 30 minutes at 4 °C on a nutator set at 12 rpm. Following incubation, the nuclei-bead suspension was diluted further with NPBt (NPB + 0.1% Triton) to a final volume of 20 mL and beads were captured magnetically using a flow-based setup as described previously (Deal and Henikoff, 2011). To capture nuclei, the suspension was allowed to flow through the setup at a rate of $\sim$0.75 mL min^{-1}. The micropipette tip was then removed from the setup and attached to a p1000 pipette for release of the beads (with bound nuclei) into fresh 20 mL of NPB for another round of flow-based bead capture. After the flow is complete, beads with bound nuclei were eluted in fresh 1 mL of NPBt. For RNA isolation, the microcentrifuge tube was placed on the magnetic stand, NPBt removed, and beads resuspended with 100 µL of RNA Lysis Buffer from RNaqueous^{TM}-Micro Total RNA Isolation Kit (AM1931, ThermoFisher Scientific). The tube was removed from the magnetic rack, tapped to mix, and allowed to incubate at room temperature for 10 minutes. The tube was then replaced onto the magnetic rack for 3 minutes and the RNA Lysis Buffer removed into a fresh tube. RNA extraction was performed immediately, or snap-freezed and stored at −80 °C.

6.4.3 Polysomal isolation

The following protocol was adapted from the TRAP method (Heiman et al., 2014) and optimised for highest yield per zebrafish embryo. Zebrafish embryos were obtained via crosses between selected *Avi effector* and *BirA driver* lines. A maximum of 350 embryos (24 hpf or younger) were used per biological sample. Embryos were dechorionated and washed in Cell Lysis Buffer (20 mM HEPES (pH 7.4), 10 mM $MgCl_2$, 150 mM KCl, 0.5 mM DTT, rRNasin, RNaseOUT, SUPERaseIN and 1X cOmplete^{TM} protease inhibitor) supplemented with 0.01% Tricaine. Embryos were resuspended in Cell Lysis Buffer at a **ratio of 50 embryos per 1 mL of Cell Lysis Buffer** and transferred to a Dounce homogenisor (available in different sizes at Sigma-Aldrich as Dounce tissue grinder sets). Embryos were dissociated using 20 strokes of Pestle A, cycloheximide added to a final concentration of 100 µg mL^{-1}, and incubated at room temperature for 15 minutes. This was followed by 60 strokes of Pestle B (3 x 20 strokes with 5 minutes pause in between) on ice. Nuclei were collected by centrifugation at 2,000 *g* for 10 minutes at 4 °C and the post-nuclear supernatant removed into clean microcentrifuge tubes (1 mL supernatant per tube). IGEPAL CA-630 and 07:0 DHPC were added to each supernatant to a final concentration of 1% each and mixed by inverting for 10 times. The tubes were incubated on ice for 5 minutes before being cleared by centrifugation at 20,000 *g* for 10 minutes at 4 °C. The post-mitochondrial supernatant was then removed and incubated with Streptavidin beads prepared for RNA procedures as described above for 1 hour at 4 °C on a nutator set at 12 rpm. Following incubation, the tubes were placed on a magnetic stand for 3 minutes and the lysates (containing unbound polysomes) were removed. The captured polysomes-beads were washed 4 times in the cold room with High Salt Buffer (20 mM HEPES (pH 7.4), 10 mM $MgCl_2$, 350 mM KCl, 0.5 mM DTT, 100 µg mL^{-1} cycloheximide, 1% IGEPAL CA-630, rRNasin, RNaseOUT, SUPERaseIN and 1X cOmplete^{TM} protease inhibitor), pooling every two tubes' worth of polysomes-beads into one. After the final wash, for RNA isolation, the tubes were

placed on the magnetic stand, High Salt Buffer removed, and beads resuspended with 200 µL of RNA Lysis Buffer from RNaqueous^{TM}-Micro Total RNA Isolation Kit (AM1931, ThermoFisher Scientific). The tubes were removed from the magnetic rack, tapped to mix, and allowed to incubate at room temperature for 10 minutes. The tube were then replaced onto the magnetic rack for 3 minutes, the RNA Lysis Buffer removed and pooled into one fresh tube. RNA extraction was performed immediately, or snap-freezed and stored at −80 °C.

6.4.4 RNA extraction of isolated nuclei and polysomes

RNA samples were extracted according to manufacturer's protocol using RNaqueous^{TM}-Micro Total RNA Isolation Kit (AM1931, ThermoFisher Scientific). RNA quality was accessed on the 2100 Bioanalyzer System using RNA 6000 Pico Kit (5067-1513, Agilent Technologies).

6.4.5 ATAC-seq from FAC-sorted cells

Embryos were dissociated using 20 mg mL^{-1} collagenase (C8176, Sigma-Aldrich) in trypsin solution (0.05% Trypsin, 0.53 mM EDTA in 1X HBSS). 300 µL collagenase/trypsin solution per 100 embryos was used and dissociation performed by homogenising every 3 to 5 minutes with a p200 low-binding micropipette tip until no pieces of material was detected. The reaction was stopped by adding the dissociated material to 4 mL of Hank's solution (10 mM HEPES (pH 7.4), 2.5 mg mL^{-1} BSA in 1X HBSS) followed by gentle trituration with a glass serological pipette. Dissociated cells were spun at 500 *g* for 10 minutes at room temperature. Hank's solution was carefully removed with an aspirator and replaced with another 4 mL of fresh Hank's solution and triturated as before. The suspension was passed through a 40 µm cell strainer into a clean Falcon tube and centrifuged with strainer attached for 10 minutes at 750 *g*. The supernatant was removed until approximately 500 µL is left and the cells resuspended by gentle tapping. Positive cells were sorted using BD

FACSARIA Fusion system. Following sorting, the cells were immediately washed with PBS for the ATAC procedure as described previously (Buenrostro et al., 2015). Fragment sizes for each library were checked using TapeStation D1000 ScreenTape and Reagents (5067-5582 and 5067-5583, Agilent Technologies) and library concentrations were determined using Library Quantification Kit for Illumina (KK4835, KAPA Biosystems).

6.4.6 Protein extraction and Western Blot

Embryos were dechorionated, deyolked and lysed at 24 hpf in 500 µL of hypotonic buffer (20 mM HEPES (pH 7.4), 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT and 1X cOmplete^{TM} protease inhibitor). Lysed embryoes were centrifuged at maximum speed for 2.5 minutes at 4 °C and the supernatant (cytoplasmic fraction) collected. To obtain the nuclear fraction, embryos were further lysed in nuclear lysis buffer (20 mM HEPES (pH 7.4), 1.5 mM $MgCl_2$, 420 mM KCl, 0.2 mM EDTA, 0.4 mM PMSF, 20% glycerol, 1 mM DTT and 1X cOmplete^{TM} protease inhibitor), centrifuged at maximum speed for 2.5 minutes at 4 °C and the supernatant collected. Protein samples were snap-freezed for storage at −80 °C. Detection of BirA transgene protein was performed with mouse α-HA antibody (1:1000; 12CA5, Roche). For Avi-tagged proteins containing Cerulean or Citrine, detection was performed using rabbit α-GFP antibody (1:1000; TP401, Torrey Pines Biolabs). A Streptavin-HRP conjugate (1:20,000; NEL750001Ea, Perkin Elmer) was used to detect biotinylated protein. HA and GFP Western Blots were performed using standard methods. Biotinylated proteins were detected using a modified procedure as follows. Membranes were blocked with 5% BSA in 1X TBST (25 mM Tris, 125 mM NaCl, 0.2% Tween-20) and incubated with Streptavidin-HRP conjugate for 1 hour at RT. Membranes were washed in Strep Wash (0.5 M NaCl, 0.3% Triton X-100 in 1X TBS). For all blots, signal was detected using ECL Plus Western Blot Detection Reagent (discontinued, GE Healthcare Life Sciences).

6.5 RNA-seq

6.5.1 Preparation of strand-specific libraries

Total RNA samples (50 ng to 100 ng starting amount) were first ribo-depleted using either Ribo-Zero™ Magnetic Kit (MRZH116, Epicentre) or RiboErase kit (KK8483, KAPA Biosystems). For poly(A) selection, NEBNext™ Poly(A) mRNA Magnetic Isolation Module (E7490, New England Biolabs) was used. Libraries were prepared using Stranded RNA-Seq Library Preparation Kit (KK8483, KAPA Biosystems) and sizes determined using TapeStation D1000 ScreenTape and Reagents (5067-5582 and 5067-5583, Agilent Technologies). Library concentrations were determined using Library Quantification Kit for Illumina (KK4835, KAPA Biosystems). Sequencing experiments were performed by pooling up to 8 or 12 libraries per flow cell on HiSeq2500 or NextSeq500 Illumina platforms, respectively. All libraries were sequenced to obtain at least 20 million reads per sample.

6.5.2 Bioinformatics processing

Differential expression analysis. Raw reads were trimmed based on quality using sickle (Joshi and Fass, 2011) with the following parameters; `-t sanger`, `-l 30`, `-q 30`, to retain Q30 reads (Illumina recommended quality score) of at least 30 bp in length. Differential expression analyses were performed using either DESeq2 (Love et al., 2014) or the *sleuth* method (Pimentel et al., 2016) (both implemented as R packages). If DESeq2 was used, count tables were first obtained using HTSeq (Anders et al., 2015). If *sleuth* was used, transcript quantification was first performed using *kallisto* (Bray et al., 2015) with bootstrap parameter set as `-b 100`.

Gene set analysis. Gene set analysis of statistically more abundant genes in *myl* nuclei vs *actb2* nuclei was performed using Piano (Väremo et al., 2013), available as a Bioconductor R package. A table of PANTHER signalling pathway annotations was first pre-loaded as a gene set collection (`gsc`) using the `loadGSC` function.

The loaded `gsc` was then used as input to perform the `runGSA` function on the higher abundance gene set. To visualise statistically-significant nodes of enriched signalling pathways, the `networkPlot` function was performed on the `runGSA` output.

rRNA filtering for translatome data. As a precautionary step before transcript quantification and differential expression analysis, reads were filtered to remove those mapping to rRNA genes. Quality-trimmed reads were mapped to *D. rerio* rRNA sequences obtained from the SILVA database (Quast et al., 2013) using bowtie2 (Langmead et al., 2009). Unmapped reads were retained and used for differential expression analysis using *kallisto* and *sleuth*.

Validation of transgene alleles using *de novo* transcript assembly

Processing of reads for assembly. Illumina adapters were trimmed from raw reads using scythe (Buffalo, 2011). Adapter-trimmed reads were further trimmed based on quality using sickle (Joshi and Fass, 2011) with the following parameters; `-t sanger`, `-l 50`, `-q 30`, to retain Q30 reads (Illumina recommended quality score) of at least 50 bp in length. Using these processed reads as input, *de novo* transcript assembly was performed using Trinity (Grabherr et al., 2011) with the following parameters; `-JM 100G`, `-no_bowtie` and `-SS_lib_type RF`.

Generating a modified genome containing the transgene-of-interest. FASTA sequence of the transgene was added to the full FASTA sequence of the *D. rerio* genome as an additional chromosome using the UNIX `cat` command. The modified genome was then indexed using samtools (Li et al., 2009) faidx function.

BLAT alignment of assembled transcripts. Prior to BLAT alignment (Kent, 2002), an over-occurring 11-mer tile of the modified genome was created using `-makeOoc`. The Trinity output FASTA file containing all the assembled transcripts were aligned

using BLAT default parameters. Alignment results were converted to BAM format and viewed on Integrative Genomics Viewer (Robinson et al., 2011) to identify the transgene mRNA transcript on the additional chromosome. The identified transcript sequence was extracted and analysed on SnapGene.

6.6 Circular RNA validation

This section describes a novel method that we have developed and is currently unpublished.

6.6.1 Simultaneous screening and amplification of circRNA candidates

RNA isolation and processing. Nuclei were isolated using the *biotagging* method and RNA extracted (see above). RNA quality and quantity were verified using TapeStation RNA ScreenTape and Reagents (5067-5576 and 5067-5577, Agilent Technologies) and Qubit™ RNA HS Assay kit (Q32852, ThermoFisher Scientific), respectively. The nuclear RNA sample was split into two halves (ideally ∼100 ng each). One half was ribo-depleted and the other half was poly(A)-selected, both as previously described (see above). In both cases, the processed RNA material was eluted with 12 µL buffer to recover ∼10 µL.

Rolling reverse transcription (rolling-RT). Reverse complement[1] reads corresponding to predicted circular RNA junctions using *find_circ.py* (Memczak et al., 2013) were extracted and aligned using Clustal Omega Multiple Sequence Alignment tool (www.ebi.ac.uk/Tools/msa/clustalo/) to obtain consensus DNA sequences for each junction-of-interest. Twenty sequences were ordered as standard ssDNA oligos from Integrated DNA Technologies (IDT), and pooled to obtain a final concentration of 2 pmol μL^{-1} per probe (Liles et al., 2004). This was achieved by combining 1 µL of 40 pmol per oligo. Reverse transcription was performed in parallel for both ribo-depleted and poly(A) RNA samples, according to manufacturer's protocols using SunScript™ Reverse Transcriptase RNaseH- (421010, Expedeon) but with the following modifications. First, the following mix was prepared; 1 µL of probe pool, 1 µL of 10 mM dNTPs and 10 µL of RNA recovered in the previous section. The mix was denatured at 70 °C for 5 minutes and incubated on ice for at least 1 minute

before the remaining components were added. This was followed by incubation at 65 °C for 1 hour, pausing after 30 minutes for addition of 1 µL of 10 mM dNTPs, before allowing incubation to continue and deactivated as recommended (95 °C for 10 minutes).

[1]Due to the chemistry of stranded RNA-seq library preparation using KAPA library preparation kit described above.

Removal of RNA and cDNA purification. Following rolling-RT, 0.2 µL of 10 mg mL^{-1} RNaseA (EN0531, ThermoFisher Scientific) was added to the reactions and incubated at 37 °C for 1 hour. The resulting cDNA pools were purified using Monarch™ PCR & DNA Cleanup kit (T1030, New England Biolabs), following manufacturer's protocol for cDNA cleanup, and final elution performed with 10 µL of buffer (typical yield was ∼5 ng μL^{-1} per sample).

Unbiased amplification of cDNA pool. Amplification was performed according to manufacturer's protocol using TruePrime Whole Genome Amplification kit (370025, Expedeon) (Picher et al., 2016). If desired, 5 µL of the final reaction can be assessed by gel electrophoresis using Gel Loading Buffer II provided with mMESSAGE mMACHINE kits (ThermoFisher Scientific).

Phenol/chloroform purification of WGA pool and library preparation. To each WGA sample, volume was made up to 100 µL with nuclease-free water, followed by addition of 11 µL of 3 M NaOAc and 111 µL of phenol-chloroform (15593031, ThermoFisher Scientific). The mixes were vortexed for 10 seconds, and centrifuged at maximum speed for 5 minutes at RT. The aqueous phases were carefully removed and amplified DNA purified using conventional ethanol-based purification method. The pellets were resuspended in 20 µL of nuclease-free water and the final products quantified using Qubit™ dsDNA HS Assay Kit (Q32851, ThermoFisher Scientific).

Starting with 100 ng of purified amplified DNA, libraries were prepared using Nextera DNA Library Prep kit (FC-121-1030, Illumina) with 5 µL each of 10 µM Nextera index primers, and sequenced on MiSeq platform using MiSeq v2 300 cycles kit (MS-102-2002, Illumina).

6.6.2 Bioinformatics processing

Reconstruction of circRNA sequences. Illumina adapters were trimmed from raw reads using scythe (Buffalo, 2011). Adapter-trimmed reads were further trimmed based on quality using sickle (Joshi and Fass, 2011) and the following parameters; `-t sanger`, `-l 50`, `-q 30`, to retain Q30 reads (Illumina recommended quality score) of at least 50 bp in length. Using these processed reads as input, *de novo* transcript assembly was performed using Trinity (Grabherr et al., 2011) with the following parameters; `-JM 100G` and `-no_bowtie`. Prior to BLAT alignment (Kent, 2002), an over-occurring 11-mer tile of the *D. rerio* genome (danRer10) was created using `-makeOoc`. The Trinity output FASTA file containing all the assembled transcripts were aligned using BLAT default parameters (alternative approach can be used using StringTie+SR, see next paragraph). Alignment results were converted to BAM format and viewed on UCSC Genome Browser to identify Trinity-assembled transcripts that overlapped loci containing junctions-of-interest, in the ribo-depleted sample only. Full sequences of promising transcripts were imported into SnapGene software for further sequence analysis (e.g. sequence periodicity, exon/intron structure).

***De novo* assembly of mRNA transcripts using StringTie.** To obtain possible mRNA splicing events between tandem duplicated genes at loci with putative circRNAs, an alternative *de novo* transcript assembly approach was used on *biotagging* RNA-seq datasets. Raw reads were processed as described above for Trinity, but assembled using the StringTie+SR approach (Pertea et al., 2015) with the following superreads.pl paramaters; `-s 1` and `-j 14100000000`. "LongReads" were mapped

to the genome using HISAT (Kim et al., 2015). Whilst much faster than the Trinity + BLAT method described above, with our parameters, this method appeared to provide slightly varied results when used for circRNA sequence validation. In any case, putative transcripts should always be extracted and manually inspected on a sequence analysis software (like SnapGene) and not relied upon genomic alignments alone.

6.7 Whole-mount Immunohistochemistry (Sox10 and Cas9)

Dechorionated 3 dpf CRISPR*i* transgenic embryos were collected and fixed with 4% PFA in 1X PBS for 1 hour at room temperature, in small glass beakers. Embryos were rinsed 2 X 5 minutes and then washed 3 X 15 minutes in PBT (2% DMSO and 0.5% Triton X-100 in 1X PBS, freshly prepared). To block, embryos were incubated with 10% Donkey serum in PBT for 1 to 2 hours at room temperature. Embryos were incubated with the following primary antibodies in 10% Donkey serum in PBT - rabbit α-Sox10 (1:500, GTX128374, GeneTex) and mouse α-Cas9 (1:100, 7A9-3A3, Cell Signaling Technology), overnight at 4 °C on the nutator. Excess primary antibodies were washed 5 X 1 hour followed by 1 X overnight with PBT. Embryos were incubated with the following secondary antibodies in PBT - α-rabbit Alexa 568 (1:200) to detect Sox10 and α-mouse Alexa 488 (1:200) to detect Cas9. Prior to secondary antibody dilutions, aliquots were centrifuged at maximum speed for 1 minute to remove fluorescent precipitates. Embryos were washed 8 X 1 hour followed by 1 X overnight in PBT before imaging.

Appendix A

Gene Lists

A.1 Chapter 3

The following tables contain these column headings:

- ensTranscript .. Ensembl transcript ID
- ensGene .. Ensembl gene ID
- pval .. *p*-value
- qval .. FDR adjusted *p*-value using Benjamini-Hochberg
- b .. the β value (analogous to fold change, though technically a bias estimator)
- se_b .. the standard error of β
- mean_obs .. the mean of the observations
- var_obs .. the variance of the observations
- tech_var .. the technical variance from the bootstraps

Adapted from a response by the first author of the *sleuth* paper, Harold Pimentel, to a question posed by a user to the Google group of *kallisto* and *sleuth* users.
(Source: https://groups.google.com/d/msg/kallisto-sleuth-users/kWodd7CQejE/g8Tdqcq3BgAJ)

A.1.1 Genes more abundant in the *sox10* translatome

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000166505	elavl3	1.7700E-05	0.00749	6.12206	1.42616	2.36894	12.92942	1.66588
ENSDART00000171062	srsf1a	2.0500E-05	0.00817	5.87861	1.38004	2.24616	12.78902	1.43236
ENSDART00000140051	cica	3.3300E-08	0.00005	5.71327	1.03444	2.16349	10.91109	0.58168
ENSDART00000140467	ugp2b	6.1900E-06	0.00338	5.56978	1.23235	2.09174	10.58482	0.97968
ENSDART00000035538	ptp4a3	6.8622E-04	0.09874	5.33919	1.57265	2.77312	9.91926	2.26126
ENSDART00000168304	rbck1	2.2019E-04	0.04457	5.32114	1.44023	1.99788	9.53113	1.46121
ENSDART00000129634	npm1a	2.6000E-05	0.00968	4.87246	1.15844	1.74308	7.98994	0.47512
ENSDART00000054460	egr1	7.0900E-06	0.00366	1.95679	0.43572	3.69227	1.40291	0.04555
ENSDART00000043298	fosab	2.2200E-10	0.00000	1.93933	0.30562	3.94767	1.31593	0.03413
ENSDART00000142420	kat5a	1.5695E-04	0.03678	1.86852	0.49434	3.89746	1.18174	0.18824
ENSDART00000136344	CU570881.3	2.8963E-04	0.05395	1.72696	0.47648	3.24544	1.03036	0.11249
ENSDART00000169420	nme2b.2	1.0587E-04	0.02730	1.44677	0.37319	3.49406	0.77779	0.05440
ENSDART00000121226	Telomerase-vert	1.6900E-13	0.00000	1.38910	0.18845	5.85799	0.66687	0.00330
ENSDART00000010477	hsp70.3	6.9400E-05	0.02005	1.37656	0.34601	6.73953	0.71145	0.00250
ENSDART00000161982	SCARNA6	4.0500E-08	0.00005	1.31071	0.23881	4.85630	0.61068	0.00832
ENSDART00000161602	FP236632.1	6.1110E-04	0.09106	1.28786	0.37584	3.43921	0.58074	0.05075

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000130025	Metazoa_SRP	2.2600E-05	0.00880	1.27725	0.30139	3.74078	0.56840	0.02617
ENSDART00000145909	sepp1b	1.5972E-04	0.03693	1.25498	0.33240	4.24618	0.59865	0.06385
ENSDART00000164935	CU457819.3	5.5539E-04	0.08587	1.24964	0.36195	4.44155	0.60787	0.02139
ENSDART00000129710	crtac1a	1.6427E-04	0.03772	1.23319	0.32724	3.74514	0.57831	0.03167
ENSDART00000169958	BX908386.1	4.6600E-05	0.01515	1.21959	0.29949	3.84944	0.50272	0.03134
ENSDART00000173347	zp2.2	6.7457E-04	0.09790	1.21453	0.35725	4.25352	0.57678	0.06083
ENSDART00000145709	BX569801.1	4.3768E-04	0.07248	1.09902	0.31255	3.82574	0.40977	0.03812
ENSDART00000144759	zgc:136791	1.4319E-04	0.03488	1.06934	0.28121	4.36536	0.40969	0.03600
ENSDART00000135029	BX927398.3	4.1400E-06	0.00234	1.06632	0.23160	4.54933	0.40423	0.01502
ENSDART00000030103	krt94	1.7600E-06	0.00131	1.03694	0.21697	4.64407	0.35986	0.01066
ENSDART00000128108	urahb	1.0800E-10	0.00000	1.02069	0.15812	5.64045	0.35185	0.00319
ENSDART00000118923	snoU85	1.7900E-14	0.00000	1.01374	0.13225	6.45211	0.34377	0.00183
ENSDART00000161596	si:ch211-193l2.4	5.5165E-04	0.08587	1.00893	0.29208	5.26385	0.39618	0.00931
ENSDART00000043743	HSPA8	2.1257E-04	0.04372	0.99196	0.26784	4.06154	0.34089	0.02140
ENSDART00000138513	CR936321.1	9.0100E-05	0.02437	0.97500	0.24900	4.33959	0.34751	0.01834
ENSDART00000141092	BX469930.1	4.8690E-04	0.07811	0.95909	0.27498	4.86927	0.35703	0.01184
ENSDART00000119160	SCARNA6	1.2400E-05	0.00580	0.94254	0.21563	4.68412	0.31433	0.01092

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000142488	taco1	1.5554E-04	0.03672	0.93633	0.24757	4.41308	0.29245	0.01930
ENSDART00000161819	U3	1.5900E-06	0.00121	0.93433	0.19465	5.94879	0.31625	0.00306
ENSDART00000154809	CU467861.1	3.3788E-04	0.06090	0.92710	0.25865	4.86050	0.33111	0.01373
ENSDART00000170761	FP325123.2	1.1695E-04	0.02948	0.91855	0.23843	4.41623	0.29726	0.01493
ENSDART00000144030	si:ch211-214b16.4	1.5000E-06	0.00119	0.88632	0.18424	5.38066	0.26863	0.00921
ENSDART00000169891	snoU85	1.6600E-05	0.00725	0.84636	0.19654	4.96271	0.25146	0.00792
ENSDART00000032996	ppdpfa	1.5400E-07	0.00017	0.76255	0.14531	5.99498	0.19871	0.00248
ENSDART00000023550	hsp90aa1.2	1.9600E-09	0.00000	0.69743	0.11622	7.18719	0.16608	0.00107
ENSDART00000025852	tnni2b.1	6.1600E-05	0.01878	0.66389	0.16570	6.00280	0.15675	0.00888
ENSDART00000134206	hspb8	3.9200E-06	0.00230	0.63147	0.13681	6.34869	0.13440	0.00247
ENSDART00000015224	gadd45gb.1	5.9500E-05	0.01829	0.62372	0.15535	5.76613	0.14105	0.00354
ENSDART00000144750	tnni2b.1	3.2566E-04	0.05902	0.60341	0.16789	7.16699	0.14016	0.00140
ENSDART00000145377	rpl38	1.3500E-07	0.00015	0.59460	0.11278	7.58087	0.12371	0.00142
ENSDART00000171678	ubb	6.0300E-07	0.00058	0.58185	0.11660	7.44884	0.11370	0.00195
ENSDART00000103330	chac1	5.4318E-04	0.08550	0.57806	0.16714	5.61544	0.12746	0.00587
ENSDART00000123878	ppiaa	3.4537E-04	0.06192	0.56987	0.15924	7.85892	0.12516	0.00062
ENSDART00000048968	EIF1B	5.8200E-05	0.01822	0.56837	0.14138	6.85176	0.11532	0.00628

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000124935	brk1	1.4583E-04	0.03512	0.55662	0.14655	6.17226	0.10478	0.00412
ENSDART00000165314	AL954715.2	1.8423E-04	0.03993	0.55317	0.14792	6.04582	0.11614	0.00363
ENSDART00000132657	si:ch211-163l21.7	1.2000E-05	0.00576	0.55296	0.12633	6.68866	0.10790	0.00151
ENSDART00000159326	gstp1	2.0995E-04	0.04372	0.55088	0.14862	6.12543	0.11588	0.00227
ENSDART00000145370	zgc:112234	1.8247E-04	0.03993	0.55042	0.14709	6.50579	0.11541	0.00152
ENSDART00000152541	BX548015.2	9.0200E-06	0.00455	0.54491	0.12274	6.81597	0.10304	0.00120
ENSDART00000159666	myl10	1.9900E-07	0.00021	0.54469	0.10474	7.91790	0.10377	0.00039
ENSDART00000146994	txlnba	7.1400E-05	0.02043	0.54049	0.13609	6.66795	0.09824	0.00397
ENSDART00000156873	BX324206.2	1.4665E-04	0.03512	0.53966	0.14214	6.22234	0.11051	0.00318
ENSDART00000172764	CR769769.2	1.7169E-04	0.03845	0.52291	0.13917	6.35570	0.09955	0.00316
ENSDART00000100941	chchd1	6.5025E-04	0.09561	0.51737	0.15173	5.86070	0.09084	0.00328
ENSDART00000162924	SCARNA2	2.0100E-05	0.00812	0.51521	0.12083	6.91385	0.09293	0.00115
ENSDART00000018523	ahcy	6.5900E-06	0.00345	0.50895	0.11293	7.64567	0.09232	0.00161
ENSDART00000173104	CABZ01044764.1	6.8988E-04	0.09884	0.50593	0.14908	5.92050	0.09093	0.00299
ENSDART00000104687	pfn2	2.1200E-06	0.00155	0.50125	0.10572	7.83065	0.08842	0.00043
ENSDART00000075926	scel	4.0685E-04	0.06991	0.49757	0.14073	6.25427	0.08277	0.00298
ENSDART00000040926	eif4ebp2	3.5100E-05	0.01239	0.49228	0.11898	7.00792	0.08968	0.00107

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000075028	rps11	1.0400E-05	0.00507	0.47256	0.10721	8.74953	0.08210	0.00018
ENSDART00000127033	ppp1r3ab	7.0506E-04	0.09974	0.45319	0.13378	6.53141	0.08039	0.00166
ENSDART00000168840	rpl35a	9.8500E-06	0.00486	0.45080	0.10198	8.29463	0.07064	0.00038
ENSDART00000166238	rplp2	2.4100E-05	0.00922	0.44991	0.10654	7.76302	0.06967	0.00047
ENSDART00000123908	si:ch211-202a12.4	5.8322E-04	0.08768	0.42304	0.12300	6.98060	0.06215	0.00194
ENSDART00000167035	mibp2	2.4174E-04	0.04796	0.39792	0.10840	7.60693	0.05334	0.00052
ENSDART00000020311	rpl27	8.2100E-05	0.02276	0.39194	0.09953	8.62270	0.05137	0.00018
ENSDART00000128025	rps26l	1.8350E-04	0.03993	0.39083	0.10448	7.95355	0.05405	0.00040
ENSDART00000129530	rps27a	3.9138E-04	0.06764	0.39014	0.11003	9.12670	0.05881	0.00013
ENSDART00000031173	rps27.1	1.0947E-04	0.02802	0.38584	0.09974	8.60383	0.05405	0.00021
ENSDART00000033663	rps21	1.9446E-04	0.04153	0.37800	0.10145	8.27302	0.04883	0.00024
ENSDART00000052761	rpl39	2.2068E-04	0.04457	0.36640	0.09919	8.76768	0.04990	0.00019
ENSDART00000029646	rplp1	3.5043E-04	0.06225	0.36122	0.10105	8.57051	0.04358	0.00045
ENSDART00000060444	rps29	5.6309E-04	0.08620	0.34240	0.09928	8.70169	0.04181	0.00017

TABLE A.1: Genes more abundant in the *sox10* translatome

A.1.2 Genes less abundant in the *sox10* translatome

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000061117	rrbp1b	0.00070	0.09974	-0.34392	0.10151	8.28624	0.04067	0.00027
ENSDART00000143005	mdn1	0.00050	0.07957	-0.34406	0.09889	8.83742	0.04380	0.00015
ENSDART00000132536	hcfc1b	0.00055	0.08556	-0.35731	0.10336	8.05077	0.04519	0.00033
ENSDART00000171015	CABZ01080518.1	0.00026	0.04983	-0.38544	0.10548	7.85323	0.05155	0.00043
ENSDART00000142056	megf8	0.00041	0.07003	-0.39361	0.11147	7.40950	0.05453	0.00067
ENSDART00000080289	prrc2c	0.00069	0.09874	-0.39497	0.11633	7.83186	0.06102	0.00174
ENSDART00000111806	cenpe	0.00006	0.01903	-0.39528	0.09881	8.84259	0.05798	0.00014
ENSDART00000093596	mt-nd1	0.00027	0.05089	-0.39931	0.10950	7.54425	0.05779	0.00060
ENSDART00000077125	aplp2	0.00014	0.03488	-0.40085	0.10543	7.86373	0.05818	0.00044
ENSDART00000110770	herc2	0.00017	0.03845	-0.40621	0.10808	7.66325	0.05586	0.00058
ENSDART00000019145	lbr	0.00056	0.08587	-0.41178	0.11928	7.15549	0.05659	0.00168
ENSDART00000115011	ddr1	0.00061	0.09106	-0.41482	0.12105	6.93860	0.05877	0.00130
ENSDART00000008043	adgrv1	0.00011	0.02917	-0.41600	0.10786	7.86895	0.06384	0.00096
ENSDART00000020456	MMP15	0.00066	0.09693	-0.41699	0.12248	6.80740	0.05831	0.00110
ENSDART00000167331	appa	0.00035	0.06225	-0.41879	0.11716	7.06753	0.06361	0.00087
ENSDART00000108851	dsg2.1	0.00041	0.06991	-0.43016	0.12171	6.87065	0.06610	0.00118

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000009790	cx43.4	0.00003	0.00968	-0.43140	0.10254	8.15243	0.06554	0.00031
ENSDART00000138988	akap1b	0.00024	0.04796	-0.43289	0.11790	7.02319	0.06396	0.00088
ENSDART00000013676	chd8	0.00026	0.04983	-0.43464	0.11896	7.57355	0.06928	0.00284
ENSDART00000082453	celsr2	0.00020	0.04240	-0.43550	0.11714	7.08637	0.07071	0.00093
ENSDART00000160236	fras1	0.00010	0.02557	-0.43600	0.11173	7.81757	0.06432	0.00171
ENSDART00000123104	notch2	0.00003	0.01168	-0.44017	0.10592	7.79769	0.06571	0.00041
ENSDART00000044000	plxna3	0.00003	0.01168	-0.44077	0.10612	7.77603	0.06655	0.00041
ENSDART00000003932	erbb2	0.00066	0.09693	-0.44109	0.12959	6.56877	0.06516	0.00176
ENSDART00000145562	robo4	0.00057	0.08667	-0.44878	0.13023	6.49118	0.06755	0.00152
ENSDART00000165374	tpra	0.00004	0.01387	-0.45810	0.11168	7.84956	0.07231	0.00176
ENSDART00000169242	EPHB3	0.00044	0.07248	-0.46408	0.13201	7.32618	0.08341	0.00071
ENSDART00000126314	pcdh8	0.00050	0.07910	-0.46745	0.13420	6.92802	0.08484	0.00082
ENSDART00000167903	CABZ01080702.1	0.00031	0.05672	-0.46861	0.12986	6.51688	0.07545	0.00155
ENSDART00000093609	mt-co2	0.00000	0.00185	-0.47002	0.10026	8.47088	0.07410	0.00021
ENSDART00000075259	fgfr1b	0.00031	0.05732	-0.47366	0.13141	6.52368	0.07632	0.00200
ENSDART00000039280	fzd7b	0.00019	0.04034	-0.47533	0.12727	6.60044	0.07554	0.00132
ENSDART00000007058	col18a1	0.00021	0.04372	-0.47714	0.12881	7.71430	0.08244	0.00560

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000037025	slc39a10	0.00005	0.01730	-0.47932	0.11867	7.00504	0.08021	0.00099
ENSDART00000098461	ptk7a	0.00008	0.02276	-0.48002	0.12188	7.30510	0.08671	0.00061
ENSDART00000100032	hyou1	0.00002	0.00725	-0.48320	0.11230	7.35658	0.07833	0.00071
ENSDART00000093625	mt-cyb	0.00000	0.00227	-0.48372	0.10455	7.94852	0.08149	0.00041
ENSDART00000034216	dync1h1	0.00000	0.00119	-0.48511	0.10069	8.50337	0.08144	0.00032
ENSDART00000004392	fkbp9	0.00000	0.00228	-0.48519	0.10499	7.88351	0.07934	0.00038
ENSDART00000061973	tgfbr1b	0.00015	0.03569	-0.48801	0.12873	6.56615	0.08094	0.00152
ENSDART00000013604	lemd3	0.00036	0.06349	-0.49196	0.13788	6.37606	0.08734	0.00293
ENSDART00000003503	col11a1a	0.00006	0.01903	-0.49537	0.12399	7.49799	0.09130	0.00386
ENSDART00000123163	zgc:85936	0.00009	0.02334	-0.50394	0.12829	6.58116	0.08553	0.00148
ENSDART00000098267	nrp1b	0.00027	0.05168	-0.50460	0.13863	6.21952	0.08559	0.00218
ENSDART00000081122	COLGALT1	0.00010	0.02725	-0.50688	0.13067	6.86335	0.09703	0.00127
ENSDART00000092389	nup210	0.00000	0.00072	-0.51173	0.10388	8.00259	0.08842	0.00036
ENSDART00000079046	cpamd8	0.00064	0.09399	-0.51497	0.15077	6.13281	0.09920	0.00511
ENSDART00000059321	epcam	0.00006	0.01903	-0.52051	0.13021	7.46017	0.10161	0.00099
ENSDART00000067599	angptl2a	0.00007	0.02043	-0.52179	0.13144	6.47385	0.09348	0.00174
ENSDART00000144757	mxra5b	0.00038	0.06657	-0.52399	0.14743	5.95156	0.09515	0.00276

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000093606	mt-co1	0.00000	0.00015	-0.52997	0.10032	9.77730	0.09608	0.00006
ENSDART00000170076	frem1b	0.00002	0.00779	-0.53179	0.12427	6.80281	0.10123	0.00152
ENSDART00000140961	RANBP2	0.00024	0.04746	-0.53265	0.14488	7.03955	0.09805	0.00803
ENSDART00000155857	nid2a	0.00000	0.00069	-0.53363	0.10791	8.11654	0.09876	0.00139
ENSDART00000047201	atp2a2b	0.00001	0.00653	-0.53453	0.12324	7.25272	0.10174	0.00296
ENSDART00000085764	PLOD3	0.00000	0.00075	-0.53574	0.10901	7.57148	0.09685	0.00056
ENSDART00000145114	ptprsa	0.00000	0.00191	-0.54167	0.11596	7.29551	0.09953	0.00136
ENSDART00000154728	col12a1a	0.00000	0.00191	-0.54394	0.11638	7.09858	0.10095	0.00080
ENSDART00000158574	adam10b	0.00022	0.04445	-0.54599	0.14765	6.29367	0.09951	0.00522
ENSDART00000162714	pcdh10b	0.00018	0.03993	-0.54792	0.14653	6.33487	0.10454	0.00514
ENSDART00000093600	mt-nd2	0.00001	0.00338	-0.54876	0.12148	6.86868	0.10197	0.00112
ENSDART00000158523	CABZ01065684.1	0.00006	0.01903	-0.54965	0.13759	6.31143	0.11184	0.00246
ENSDART00000146299	slit3	0.00002	0.00847	-0.54999	0.12944	6.55214	0.10760	0.00163
ENSDART00000111613	adgrl2b.1	0.00004	0.01353	-0.55050	0.13394	6.49520	0.10423	0.00252
ENSDART00000041577	kif7	0.00047	0.07709	-0.55088	0.15756	6.89999	0.11123	0.01132
ENSDART00000145068	fat1a	0.00003	0.01031	-0.55815	0.13323	8.84315	0.10546	0.00813
ENSDART00000093618	mt-nd4	0.00000	0.00006	-0.55866	0.10216	8.19455	0.10831	0.00029

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000169404	hspa5	0.00000	0.00004	-0.55932	0.10037	8.45887	0.10852	0.00022
ENSDART00000022685	nradd	0.00000	0.00234	-0.55947	0.12141	7.36654	0.11416	0.00078
ENSDART00000087247	glg1a	0.00000	0.00053	-0.56506	0.11277	7.49989	0.10714	0.00121
ENSDART00000158147	calr3b	0.00000	0.00005	-0.56553	0.10216	8.19387	0.11248	0.00029
ENSDART00000113418	igsf3	0.00043	0.07190	-0.56671	0.16098	5.59250	0.11048	0.00361
ENSDART00000148741	met	0.00050	0.07911	-0.57042	0.16382	5.56986	0.10991	0.00429
ENSDART00000127476	erlec1	0.00041	0.07003	-0.57046	0.16154	6.60707	0.12587	0.00815
ENSDART00000124072	atp7a	0.00007	0.01953	-0.57070	0.14315	6.43412	0.12153	0.00473
ENSDART00000127875	bmp1a	0.00018	0.03993	-0.57457	0.15358	6.50205	0.12577	0.00147
ENSDART00000092627	CABZ01052815.1	0.00008	0.02297	-0.58044	0.14755	5.94903	0.11917	0.00277
ENSDART00000013638	cdh6	0.00002	0.00725	-0.58174	0.13520	6.84858	0.12499	0.00314
ENSDART00000073930	notch3	0.00000	0.00001	-0.58182	0.10069	8.41592	0.11701	0.00024
ENSDART00000168795	col9a1b	0.00013	0.03234	-0.58414	0.15267	6.10218	0.11408	0.00547
ENSDART00000093612	mt-atp6	0.00039	0.06711	-0.58517	0.16487	7.03704	0.13226	0.00086
ENSDART00000149264	pabpc1a	0.00025	0.04944	-0.58522	0.15989	6.24053	0.11752	0.00866
ENSDART00000114067	agrn	0.00021	0.04310	-0.58929	0.15876	6.83837	0.13001	0.01143
ENSDART00000143442	p3h3	0.00025	0.04869	-0.59722	0.16293	5.55939	0.13023	0.00389

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000159555	FBLN1	0.00047	0.07709	-0.59761	0.17098	6.52116	0.13854	0.00175
ENSDART00000159077	col4a2	0.00002	0.00922	-0.59855	0.14177	6.78625	0.13282	0.00220
ENSDART00000019975	acvr2b	0.00002	0.00700	-0.61230	0.14176	6.85493	0.13837	0.00142
ENSDART00000170766	glg1b	0.00046	0.07554	-0.62099	0.17726	5.29563	0.13558	0.00559
ENSDART00000073827	col4a5	0.00005	0.01739	-0.62804	0.15563	7.33414	0.14762	0.00720
ENSDART00000077296	col4a1	0.00004	0.01446	-0.62952	0.15408	6.69941	0.14793	0.00236
ENSDART00000088208	lrp1ab	0.00000	0.00000	-0.63398	0.09947	8.64765	0.13417	0.00018
ENSDART00000034683	robo1	0.00008	0.02117	-0.63865	0.16130	5.88876	0.15330	0.00503
ENSDART00000159903	lamb1a	0.00000	0.00000	-0.64024	0.09964	8.86978	0.13878	0.00031
ENSDART00000159596	zgc:91849	0.00026	0.04983	-0.64098	0.17544	5.34467	0.13906	0.00559
ENSDART00000156460	galnt16	0.00010	0.02631	-0.64852	0.16668	5.53895	0.14101	0.00490
ENSDART00000126602	bcam	0.00000	0.00000	-0.65628	0.10981	7.57030	0.14481	0.00073
ENSDART00000088643	col15a1b	0.00000	0.00009	-0.65928	0.12268	6.99993	0.14576	0.00194
ENSDART00000076850	cldni	0.00000	0.00119	-0.66261	0.13783	6.36232	0.15046	0.00283
ENSDART00000004277	lamc1	0.00000	0.00000	-0.66552	0.09863	8.93988	0.14768	0.00013
ENSDART00000130565	tmem132e	0.00001	0.00345	-0.66679	0.14784	6.05437	0.15648	0.00367
ENSDART00000165418	lrp6	0.00000	0.00191	-0.66740	0.14292	6.37781	0.15530	0.00435

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000155907	nid2a	0.00000	0.00047	-0.67965	0.13483	6.61779	0.16089	0.00339
ENSDART00000160069	lama1	0.00000	0.00000	-0.68274	0.09868	9.30462	0.15895	0.00009
ENSDART00000039345	myoc	0.00001	0.00598	-0.68532	0.15720	5.70566	0.16452	0.00354
ENSDART00000112067	hmcn2	0.00000	0.00000	-0.68548	0.10148	8.31257	0.15719	0.00030
ENSDART00000093617	mt-nd4l	0.00001	0.00580	-0.68767	0.15738	5.70190	0.16426	0.00356
ENSDART00000030138	alcama	0.00027	0.05157	-0.68830	0.18900	5.58798	0.16509	0.01337
ENSDART00000171530	FQ312024.2	0.00012	0.02970	-0.68836	0.17885	5.41069	0.16471	0.00763
ENSDART00000166842	cyt1l	0.00002	0.00779	-0.69128	0.16157	7.12790	0.17248	0.01346
ENSDART00000172387	col7a1l	0.00063	0.09370	-0.71308	0.20865	5.03623	0.16989	0.01394
ENSDART00000148481	thbs2a	0.00006	0.01802	-0.72391	0.17985	5.22338	0.17883	0.00551
ENSDART00000168201	golim4a	0.00004	0.01300	-0.72654	0.17617	5.28329	0.18766	0.00504
ENSDART00000103753	fn1a	0.00004	0.01446	-0.73316	0.17945	6.17216	0.20064	0.01208
ENSDART00000016422	mknk2b	0.00020	0.04178	-0.73548	0.19755	5.88106	0.18913	0.01946
ENSDART00000145570	ctslb	0.00003	0.01168	-0.74788	0.17990	8.00129	0.20802	0.00077
ENSDART00000102981	col8a1a	0.00000	0.00000	-0.75488	0.10387	8.17338	0.19714	0.00028
ENSDART00000125302	fbn2b	0.00000	0.00169	-0.75776	0.16086	6.89521	0.20865	0.00124
ENSDART00000077908	appb	0.00016	0.03678	-0.76165	0.20159	5.28352	0.19982	0.01464

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000170951	PXDN	0.00048	0.07770	-0.76199	0.21830	5.85502	0.22531	0.00303
ENSDART00000133365	si:ch211-246m6.5	0.00043	0.07152	-0.77332	0.21951	4.57642	0.20372	0.01023
ENSDART00000093623	mt-nd6	0.00000	0.00000	-0.77497	0.10469	7.91097	0.20724	0.00037
ENSDART00000167748	fbn2b	0.00000	0.00000	-0.77502	0.09932	8.69220	0.20245	0.00018
ENSDART00000168209	nid1b	0.00000	0.00072	-0.78394	0.15912	7.28711	0.22173	0.00149
ENSDART00000155534	CR774184.1	0.00042	0.07086	-0.78641	0.22298	4.55033	0.21275	0.01113
ENSDART00000108738	col4a5	0.00000	0.00161	-0.79028	0.16713	7.06188	0.21119	0.01505
ENSDART00000010521	fn1a	0.00029	0.05386	-0.79267	0.21859	7.07343	0.21415	0.03494
ENSDART00000091158	irg1l	0.00010	0.02631	-0.80232	0.20628	4.79205	0.24140	0.00903
ENSDART00000045126	lama5	0.00000	0.00000	-0.80262	0.10191	8.21204	0.21649	0.00026
ENSDART00000167645	zgc:174855	0.00002	0.00812	-0.81362	0.19073	6.35457	0.22875	0.02017
ENSDART00000077840	cmn	0.00039	0.06711	-0.82120	0.23137	6.86288	0.26048	0.00323
ENSDART00000100234	col2a1a	0.00000	0.00000	-0.82536	0.13169	9.19520	0.23864	0.00016
ENSDART00000097295	col18a1	0.00000	0.00221	-0.83391	0.17990	7.92412	0.25338	0.00392
ENSDART00000171013	tenm4	0.00001	0.00455	-0.84317	0.19000	6.40151	0.25068	0.02016
ENSDART00000164152	egfra	0.00017	0.03845	-0.84887	0.22595	5.22902	0.24711	0.02429
ENSDART00000047844	col2a1b	0.00004	0.01426	-0.86819	0.21210	7.52370	0.28124	0.00563

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000165181	wnt5b	0.00029	0.05447	-0.87885	0.24275	5.67931	0.29675	0.02043
ENSDART00000166808	FP325123.3	0.00032	0.05832	-0.89217	0.24793	4.33674	0.27286	0.01775
ENSDART00000109807	EMILIN3	0.00000	0.00000	-0.90610	0.13501	7.44429	0.28583	0.00061
ENSDART00000087597	col6a2	0.00000	0.00234	-0.91183	0.19813	4.95795	0.27783	0.00847
ENSDART00000174157	LTBP4	0.00000	0.00069	-0.96432	0.19494	5.05466	0.31369	0.00868
ENSDART00000097176	CABZ01090749.1	0.00000	0.00000	-0.97437	0.14213	6.10403	0.31941	0.00238
ENSDART00000128937	ipo8	0.00010	0.02631	-0.97795	0.25140	4.97224	0.34347	0.03264
ENSDART00000128837	grin2da	0.00014	0.03488	-1.03140	0.27106	4.01949	0.35853	0.02187
ENSDART00000156287	nid2a	0.00000	0.00169	-1.03402	0.21930	5.40991	0.37288	0.02372
ENSDART00000162331	si:dkey-22o22.2	0.00000	0.00005	-1.05657	0.19235	5.14743	0.37605	0.00904
ENSDART00000102629	nit1	0.00070	0.09944	-1.12979	0.33320	3.54483	0.44739	0.03095
ENSDART00000159570	CABZ01072043.1	0.00000	0.00005	-1.15832	0.20988	4.75260	0.45214	0.00981
ENSDART00000006983	kat5a	0.00058	0.08768	-1.23274	0.35843	3.90973	0.57640	0.07286
ENSDART00000166889	nptna	0.00056	0.08587	-1.44619	0.41906	3.19632	0.70311	0.05379
ENSDART00000152863	FP565451.1	0.00017	0.03785	-1.70942	0.45392	3.23580	1.00599	0.09011
ENSDART00000162328	snx7	0.00028	0.05287	-4.83799	1.33187	1.72586	7.82231	0.88664
ENSDART00000157697	sec22bb	0.00057	0.08667	-5.01406	1.45551	1.81388	8.40584	1.33080

ensTranscript	ensGene	pval	qval	b	se_b	mean_obs	var_obs	tech_var
ENSDART00000088833	si:ch73-233f7.1	0.00006	0.01829	-5.39221	1.34254	2.33527	9.77415	1.41690
ENSDART00000150045	adamts18	0.00048	0.07762	-5.46787	1.56579	2.04092	10.04626	1.87375
ENSDART00000125431	prdm2b	0.00000	0.00000	-5.57053	0.84143	2.09212	10.43468	0.16926
ENSDART00000025509	tdh	0.00004	0.01320	-5.78051	1.40367	2.19716	11.17061	1.50400

TABLE A.2: Genes less abundant in the *sox10* translatome

A.2 Chapter 4

The following tables contain these column headings:

- Pauli_lncRNA .. LncRNA ID from Pauli *et al.* 2012
- Position (danRer7) .. Chromosomal position of annotated lncRNA in danRer7
- baseMean .. base mean over all rows (genes)
- log2FoldChange .. log2 fold change (over control)
- lfcSE .. log2FoldChange standard error
- stat .. Wald statistic (log2FoldChange divided by its standard error)
- pvalue .. Wald test, *p*-value
- padj .. Wald test, Benjamini-Hochberg adjusted *p*-value

Adapted from DESeq2 vignette.

A.2.1 Differentially expressed lncRNAs in *sox10 nuclear* versus *bactin2 nuclear* at 16-18 ss

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00038706	chr16:49461761-49495118	616.79385	2.05771	0.33363	6.16769	6.9300E-10	3.9000E-07
Zv9_00022003	chr9:20207660-20209274	300.65899	-2.01907	0.33660	-5.99842	1.9900E-09	5.6100E-07
Zv9_00000127	chr1:2804472-2807213	274.80375	-2.20022	0.38384	-5.73214	9.9200E-09	1.4000E-06
Zv9_00011860	chr5:67609742-67611640	153.63510	-2.25816	0.39285	-5.74813	9.0200E-09	1.4000E-06
Zv9_00044534	chr19:20674252-20681464	206.50855	-2.17051	0.38553	-5.62995	1.8000E-08	2.0300E-06
Zv9_00023789	chr9:55417516-55422084	244.25615	-1.92790	0.34925	-5.52018	3.3900E-08	3.1800E-06
Zv9_00044428	chr19:15458926-15462802	328.99362	-1.77337	0.32924	-5.38621	7.2000E-08	5.7900E-06
Zv9_00058526	Zv9_NA545:48300-65881	39.74709	4.35536	0.81618	5.33628	9.4900E-08	6.6800E-06
Zv9_00005960	chr3:30845519-30846926	112.73938	-2.66249	0.51080	-5.21239	1.8600E-07	1.1700E-05
Zv9_00011954	chr5:72043356-72044780	184.40367	-1.92951	0.37561	-5.13698	2.7900E-07	1.5700E-05
Zv9_00036392	chr16:459487-461852	24.99772	5.56765	1.09645	5.07791	3.8200E-07	1.9500E-05
Zv9_00004937	chr2:47018360-47021627	28.71794	-3.89828	0.78446	-4.96940	6.7200E-07	3.1500E-05
Zv9_00037383	chr16:45735633-45738732	173.02005	-2.02175	0.41293	-4.89612	9.7700E-07	4.2300E-05
Zv9_00013538	chr5:64464453-64465996	180.40061	-2.25840	0.46665	-4.83959	1.3000E-06	5.2300E-05
Zv9_00005961	chr3:30845520-30846328	96.92356	-2.12807	0.45816	-4.64478	3.4000E-06	1.2778E-04

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00020825	chr8:17129336-17130464	227.57878	-1.58963	0.35400	-4.49046	7.1100E-06	2.5008E-04
Zv9_00046174	chr20:52274652-52355796	20.53828	4.54095	1.02364	4.43609	9.1600E-06	3.0338E-04
Zv9_00025064	chr10:7009618-7012244	611.16761	-1.55599	0.35498	-4.38325	1.1700E-05	3.5466E-04
Zv9_00028968	chr12:44135841-44147664	40.82641	3.21054	0.73331	4.37815	1.2000E-05	3.5466E-04
Zv9_00025103	chr10:8709102-8711727	35.35075	3.11828	0.74949	4.16052	3.1800E-05	8.5126E-04
Zv9_00003314	chr2:32572090-32579874	18.00625	4.31533	1.03549	4.16744	3.0800E-05	8.5126E-04
Zv9_00055360	chr25:29679-30465	67.63899	-2.05896	0.50282	-4.09478	4.2300E-05	1.0527E-03
Zv9_00036038	chr15:31721484-31772034	13.82866	4.73885	1.15844	4.09071	4.3000E-05	1.0527E-03
Zv9_00034594	chr15:10541352-10543321	77.09088	-2.14125	0.53004	-4.03983	5.3500E-05	1.2548E-03
Zv9_00020823	chr8:17129153-17132758	358.17729	-1.36240	0.34247	-3.97821	6.9400E-05	1.5128E-03
Zv9_00004701	chr2:37535504-37536100	104.60776	-1.81897	0.45740	-3.97676	6.9900E-05	1.5128E-03
Zv9_00054996	chr24:29475091-29483696	18.33578	3.88893	0.98024	3.96731	7.2700E-05	1.5157E-03
Zv9_00027164	chr11:7384484-7385648	174.29202	-1.50115	0.38237	-3.92590	8.6400E-05	1.6774E-03
Zv9_00003645	chr2:47086248-47185256	38.55853	-2.51037	0.63911	-3.92793	8.5700E-05	1.6774E-03
Zv9_00008968	chr4:45204254-45206439	17.45732	3.78980	0.98128	3.86209	1.1242E-04	2.0594E-03
Zv9_00025423	chr10:27528059-27530569	12.22254	4.53671	1.17532	3.85998	1.1339E-04	2.0594E-03
Zv9_00024862	chr10:46574126-46577344	18.93154	3.55280	0.92441	3.84331	1.2138E-04	2.1356E-03

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00021734	chr9:2585197-2588041	213.95490	-1.33389	0.35373	-3.77098	1.6261E-04	2.7742E-03
Zv9_00037626	chr16:58676351-58679193	190.77608	1.57817	0.42211	3.73875	1.8494E-04	3.0623E-03
Zv9_00038770	chr16:54498983-54501006	138.22985	-1.45046	0.39849	-3.63994	2.7270E-04	4.3866E-03
Zv9_00057917	Zv9_NA250:14182-16287	13.50002	3.88721	1.07365	3.62055	2.9397E-04	4.5974E-03
Zv9_00058196	Zv9_scaffold3492:1682-106214	22.38453	3.03044	0.84489	3.58677	3.3480E-04	5.0944E-03
Zv9_00033658	chr14:17228181-17270812	149.94110	-1.39285	0.39090	-3.56321	3.6635E-04	5.4277E-03
Zv9_00007990	chr3:56902276-56962430	219.88684	-1.29659	0.36741	-3.52898	4.1717E-04	6.0222E-03
Zv9_00044653	chr19:27949598-27956054	114.84363	-1.68785	0.48581	-3.47432	5.1216E-04	7.0330E-03
Zv9_00030026	chr12:44254847-44257953	111.08646	-1.59021	0.45790	-3.47281	5.1504E-04	7.0330E-03
Zv9_00051377	chr22:41797058-41840513	95.85645	-1.70613	0.49367	-3.45600	5.4826E-04	7.0330E-03
Zv9_00020540	chr8:4635953-4641256	51.02654	2.42858	0.70285	3.45531	5.4965E-04	7.0330E-03
Zv9_00029985	chr12:37675757-37678223	9.32133	4.16323	1.20283	3.46119	5.3779E-04	7.0330E-03
Zv9_00039574	chr17:29081196-29096892	71.11734	1.83437	0.54804	3.34715	8.1647E-04	9.9928E-03
Zv9_00045584	chr20:26089316-26095551	26.41444	2.65193	0.79178	3.34933	8.1007E-04	9.9928E-03
Zv9_00000174	chr1:6993817-6995193	253.78548	-1.28166	0.38544	-3.32523	8.8345E-04	1.0583E-02
Zv9_00010058	chr4:28564924-28567167	52.42975	-1.98751	0.60121	-3.30587	9.4682E-04	1.1105E-02
Zv9_00035708	chr15:16770170-16773603	90.13867	-1.51504	0.46102	-3.28628	1.0152E-03	1.1665E-02

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00017842	chr7:75486474-75489532	2015.83504	-1.03227	0.31689	-3.25753	1.1239E-03	1.2655E-02
Zv9_00033168	chr14:40177321-40283404	44.27677	1.93336	0.59658	3.24072	1.1923E-03	1.3162E-02
Zv9_00036501	chr16:6405163-6528449	35.95612	-2.13204	0.66077	-3.22659	1.2527E-03	1.3415E-02
Zv9_00055439	chr25:2479072-2487598	16.07928	3.23920	1.00462	3.22429	1.2628E-03	1.3415E-02
Zv9_00028054	chr12:1939872-2212017	17.37116	2.84018	0.89027	3.19026	1.4215E-03	1.4820E-02
Zv9_00017368	chr7:45370609-45371390	7.95899	3.89535	1.22424	3.18184	1.4634E-03	1.4980E-02
Zv9_00033517	chr14:6428301-6434768	7.68431	3.88395	1.22354	3.17437	1.5016E-03	1.5097E-02
Zv9_00045093	chr19:50220711-50233538	7.52215	3.85235	1.22583	3.14265	1.6742E-03	1.6537E-02
Zv9_00052276	chr23:37211396-37216224	7.47361	3.84288	1.22654	3.13309	1.7297E-03	1.6790E-02
Zv9_00031087	chr13:40244710-40287683	7.45707	3.83550	1.22714	3.12555	1.7747E-03	1.6935E-02
Zv9_00026898	chr11:46193880-46194580	126.79300	-1.38451	0.44878	-3.08503	2.0353E-03	1.8785E-02
Zv9_00042842	chr18:44662300-44665013	82.77183	-1.54008	0.49900	-3.08635	2.0263E-03	1.8785E-02
Zv9_00027978	chr12:422221-468111	10.08035	3.44194	1.11753	3.07994	2.0704E-03	1.8801E-02
Zv9_00017150	chr7:35832416-35854864	11.73327	3.18966	1.04077	3.06472	2.1787E-03	1.9470E-02
Zv9_00013048	chr5:38534730-38544138	10.14692	3.44783	1.13473	3.03846	2.3779E-03	2.0918E-02
Zv9_00002888	chr2:13973773-13975275	6.93860	3.73325	1.23437	3.02441	2.4912E-03	2.1577E-02
Zv9_00030263	chr13:1789778-1808949	7.40744	3.73389	1.23778	3.01660	2.5562E-03	2.1806E-02

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00033054	chr14:34453646-34456452	6.88897	3.66403	1.24115	2.95213	3.1559E-03	2.6129E-02
Zv9_00003315	chr2:32740771-32745731	6.79189	3.66448	1.24036	2.95438	3.1330E-03	2.6129E-02
Zv9_00052370	chr23:41516685-41526407	10.92591	-2.98516	1.01310	-2.94656	3.2133E-03	2.6219E-02
Zv9_00044467	chr19:18416479-18423610	348.99231	-0.99271	0.34144	-2.90747	3.6437E-03	2.8893E-02
Zv9_00003599	chr2:44434721-44448194	6.48412	3.61491	1.24308	2.90802	3.6373E-03	2.8893E-02
Zv9_00038978	chr17:2330424-2333257	39.55757	1.82710	0.63108	2.89519	3.7892E-03	2.9224E-02
Zv9_00038203	chr16:29583679-29602054	6.35505	3.60293	1.24345	2.89754	3.7611E-03	2.9224E-02
Zv9_00027123	chr11:5991570-5998933	200.84258	-1.17080	0.40823	-2.86797	4.1311E-03	3.1430E-02
Zv9_00030004	chr12:42426128-42551988	6.22489	3.56297	1.24632	2.85880	4.2525E-03	3.1922E-02
Zv9_00045207	chr20:2535552-2536748	6.17635	3.55562	1.24669	2.85205	4.3438E-03	3.2179E-02
Zv9_00037861	chr16:14275174-14282470	12.53054	2.90532	1.02432	2.83634	4.5634E-03	3.2659E-02
Zv9_00014599	chr6:32013007-32015418	8.78567	3.26048	1.14845	2.83902	4.5252E-03	3.2659E-02
Zv9_00039585	chr17:29782460-29787852	6.71027	3.54770	1.25140	2.83499	4.5827E-03	3.2659E-02
Zv9_00016101	chr6:47987343-48002248	15.02410	2.62483	0.93137	2.81826	4.8285E-03	3.3787E-02
Zv9_00008374	chr4:7520822-7524897	14.20000	2.78236	0.98802	2.81610	4.8610E-03	3.3787E-02
Zv9_00017207	chr7:37998433-38000115	10.45623	3.02738	1.07671	2.81169	4.9283E-03	3.3837E-02
Zv9_00031924	chr13:33017465-33031243	19.67388	2.47818	0.88627	2.79621	5.1707E-03	3.5073E-02

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00022308	chr9:32380345-32461821	10.87026	2.97925	1.07054	2.78294	5.3869E-03	3.6105E-02
Zv9_00002190	chr1:47223133-47235705	255.41134	-0.95462	0.34541	-2.76374	5.7143E-03	3.7165E-02
Zv9_00035253	chr15:44821274-44824535	220.49064	1.24951	0.45402	2.75212	5.9212E-03	3.7165E-02
Zv9_00030806	chr13:28009225-28027481	192.31253	-0.99979	0.36252	-2.75792	5.8170E-03	3.7165E-02
Zv9_00058275	Zv9_NA303:15459-35646	50.16829	-1.52037	0.55266	-2.75101	5.9411E-03	3.7165E-02
Zv9_00008178	chr4:388071-389997	33.49947	1.89452	0.68600	2.76169	5.7503E-03	3.7165E-02
Zv9_00029354	chr12:6437865-6655843	8.13813	3.17353	1.15328	2.75174	5.9279E-03	3.7165E-02
Zv9_00034652	chr15:15054883-15064519	8.08960	3.16397	1.15447	2.74062	6.1323E-03	3.7940E-02
Zv9_00021973	chr9:17853368-17859762	16.10007	2.43929	0.89718	2.71883	6.5513E-03	3.9237E-02
Zv9_00022307	chr9:32380345-32428997	5.64134	3.41956	1.25560	2.72345	6.4605E-03	3.9237E-02
Zv9_00008901	chr4:37628504-37628862	5.60934	3.41782	1.25565	2.72196	6.4897E-03	3.9237E-02
Zv9_00006267	chr3:39622242-39626858	5.59280	3.41066	1.25607	2.71534	6.6208E-03	3.9237E-02
Zv9_00003484	chr2:38763830-38774500	37.11672	-1.73922	0.64280	-2.70571	6.8159E-03	3.9624E-02
Zv9_00034478	chr15:2836464-2847282	7.92803	-3.05565	1.12956	-2.70518	6.8268E-03	3.9624E-02
Zv9_00009158	chr4:61139037-61145088	5.49572	3.38841	1.25743	2.69472	7.0447E-03	4.0471E-02
Zv9_00038517	chr16:43005101-43007036	5.47918	3.37371	1.25854	2.68065	7.3480E-03	4.1787E-02
Zv9_00001692	chr1:18431669-18436630	44.84359	1.58211	0.59108	2.67667	7.4359E-03	4.1864E-02

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00020846	chr8:18297259-18302323	68.95060	-1.31867	0.49603	-2.65845	7.8500E-03	4.3750E-02
Zv9_00042206	chr18:4708883-4729102	7.92376	3.06729	1.15520	2.65519	7.9263E-03	4.3750E-02
Zv9_00048330	chr21:41373273-41390573	12.00213	2.71471	1.03103	2.63302	8.4630E-03	4.6243E-02
Zv9_00005311	chr3:4105305-4106310	7.81014	3.04184	1.15809	2.62660	8.6243E-03	4.6243E-02
Zv9_00039682	chr17:35445917-35454439	5.23649	3.31583	1.26198	2.62749	8.6017E-03	4.6243E-02
Zv9_00019124	chr7:68245262-68247020	202.27424	-0.99168	0.37887	-2.61749	8.8578E-03	4.6915E-02
Zv9_00028496	chr12:21845376-21848363	114.99519	1.19290	0.45613	2.61525	8.9163E-03	4.6915E-02
Zv9_00020486	chr8:1538577-1539419	38.42120	-1.62325	0.62501	-2.59715	9.4000E-03	4.9002E-02
Zv9_00042604	chr18:23467379-23617603	5.39756	3.28200	1.26582	2.59280	9.5199E-03	4.9172E-02
Zv9_00003691	chr2:48553631-48556829	5.99803	-3.11033	1.20353	-2.58434	9.7566E-03	4.9936E-02
Zv9_00020793	chr8:16581571-16583810	16.29466	2.27201	0.88022	2.58117	9.8467E-03	4.9943E-02

TABLE A.3: Differentially expressed lncRNAs in *sox10 nuclear*

A.2.2 Differentially expressed lncRNAs in *myl7 nuclear* versus *bactin2 nuclear* at 24-26 hpf

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00046280	chr20:55785164-55797386	4294.37337	2.20957	0.28093	7.86528	3.6828E-15	1.0754E-12
Zv9_00043817	chr19:44163750-44206084	339.84020	2.52472	0.35041	7.20509	5.8005E-13	8.4687E-11
Zv9_00026898	chr11:46193880-46194580	99.40541	-3.15636	0.49360	-6.39453	1.6105E-10	1.5675E-08
Zv9_00020823	chr8:17129153-17132758	109.91255	-2.75570	0.45777	-6.01989	1.7454E-09	1.2741E-07
Zv9_00030026	chr12:44254847-44257953	417.29310	1.81054	0.30594	5.91803	3.2582E-09	1.9028E-07
Zv9_00047425	chr21:286440-290860	153.66162	2.49923	0.49579	5.04087	4.6341E-07	2.2553E-05
Zv9_00005960	chr3:30845519-30846926	54.34987	-2.56852	0.58438	-4.39529	1.1063E-05	4.6147E-04
Zv9_00011860	chr5:67609742-67611640	175.83500	-1.62413	0.37421	-4.34021	1.4234E-05	5.1956E-04
Zv9_00023789	chr9:55417516-55422084	168.31309	-1.52292	0.36142	-4.21368	2.5124E-05	7.3363E-04
Zv9_00017979	chr7:4511174-4529094	11.77928	3.90548	0.92419	4.22585	2.3804E-05	7.3363E-04
Zv9_00024864	chr10:66084-72595	75.71218	2.03730	0.51457	3.95925	7.5185E-05	1.9958E-03
Zv9_00010056	chr4:28447351-28450190	35.75891	2.53583	0.64746	3.91657	8.9818E-05	2.1856E-03
Zv9_00025064	chr10:7009618-7012244	556.79460	-1.30979	0.34530	-3.79321	1.4871E-04	3.3403E-03
Zv9_00010057	chr4:28450578-28455031	58.37068	1.93905	0.51585	3.75896	1.7062E-04	3.5586E-03
Zv9_00011503	chr5:51536461-51556642	220.97306	-1.17536	0.33128	-3.54796	3.8823E-04	6.5536E-03

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00019679	chr8:20277144-20341397	93.60573	1.65470	0.46776	3.53746	4.0399E-04	6.5536E-03
Zv9_00005091	chr2:56215978-56216907	74.68109	-1.85227	0.52201	-3.54837	3.8762E-04	6.5536E-03
Zv9_00005203	chr2:59926871-59929843	18.64968	2.95738	0.82828	3.57050	3.5631E-04	6.5536E-03
Zv9_00010362	chr4:59767798-59769626	9.17462	-3.30667	0.94832	-3.48685	4.8874E-04	7.5112E-03
Zv9_00056167	chr25:31968656-32018548	95.12790	1.48437	0.43005	3.45162	5.5724E-04	8.1357E-03
Zv9_00007248	chr3:23511510-23520283	101.42136	1.49608	0.43687	3.42458	6.1576E-04	8.3639E-03
Zv9_00058248	Zv9_scaffold3546:141026-145912	205.49428	-1.17990	0.34517	-3.41829	6.3015E-04	8.3639E-03
Zv9_00026302	chr11:18456416-18462949	41.42663	-1.94542	0.59159	-3.28843	1.0075E-03	1.1192E-02
Zv9_00025662	chr10:38374481-38375433	125.18318	1.40778	0.42908	3.28088	1.0348E-03	1.1192E-02
Zv9_00027170	chr11:7929707-7930978	8.09461	3.17891	0.95685	3.32226	8.9293E-04	1.1192E-02
Zv9_00000174	chr1:6993817-6995193	93.47845	-1.83541	0.55422	-3.31171	9.2727E-04	1.1192E-02
Zv9_00042843	chr18:44662367-44667102	10.68308	-3.02155	0.91603	-3.29853	9.7191E-04	1.1192E-02
Zv9_00010983	chr5:28540595-28544534	9.73943	-3.00058	0.92438	-3.24605	1.1702E-03	1.1390E-02
Zv9_00038770	chr16:54498983-54501006	127.16395	-1.39056	0.42810	-3.24826	1.1611E-03	1.1390E-02
Zv9_00013538	chr5:64464453-64465996	120.59049	1.31241	0.40391	3.24928	1.1570E-03	1.1390E-02
Zv9_00007852	chr3:48382957-48385180	153.10826	1.20283	0.37275	3.22695	1.2512E-03	1.1785E-02
Zv9_00028496	chr12:21845376-21848363	79.10082	1.52423	0.47474	3.21063	1.3244E-03	1.2085E-02

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00038978	chr17:2330424-2333257	186.05717	1.11555	0.34924	3.19425	1.4019E-03	1.2405E-02
Zv9_00022610	chr9:48988601-48989328	66.45114	1.53591	0.48998	3.13466	1.7206E-03	1.4777E-02
Zv9_00030333	chr13:5247352-5250749	76.20030	-1.53690	0.49361	-3.11362	1.8481E-03	1.5418E-02
Zv9_00030283	chr13:2376291-2381118	341.28638	-1.04229	0.34681	-3.00537	2.6525E-03	2.0934E-02
Zv9_00049609	chr22:1845320-1864245	7.25944	-2.89691	0.96265	-3.00931	2.6184E-03	2.0934E-02
Zv9_00053557	chr23:46386320-46386611	55.44766	1.57355	0.53209	2.95731	3.1034E-03	2.3847E-02
Zv9_00008864	chr4:33831729-34083045	11.46258	-2.58397	0.87842	-2.94162	3.2650E-03	2.4446E-02
Zv9_00052508	chr23:266905-269445	783.47597	-0.88650	0.30548	-2.90202	3.7077E-03	2.7066E-02
Zv9_00014136	chr6:8463418-8464777	37.08078	1.80266	0.62300	2.89350	3.8098E-03	2.7133E-02
Zv9_00050673	chr22:6451500-6453464	116.58394	-1.22280	0.42536	-2.87471	4.0439E-03	2.8115E-02
Zv9_00042955	chr19:58751-62063	10.38295	2.55359	0.89590	2.85031	4.3677E-03	2.9660E-02
Zv9_00011954	chr5:72043356-72044780	56.22962	-1.50941	0.53443	-2.82432	4.7381E-03	3.1444E-02
Zv9_00044534	chr19:20674252-20681464	104.56383	-1.19315	0.42889	-2.78195	5.4033E-03	3.4540E-02
Zv9_00027782	chr11:38779322-38780314	52.97769	1.51448	0.54484	2.77968	5.4413E-03	3.4540E-02
Zv9_00020637	chr8:8278868-8300470	49.14177	1.49429	0.54113	2.76145	5.7545E-03	3.5006E-02
Zv9_00054343	chr24:40848449-40849753	110.39713	-1.29215	0.46744	-2.76432	5.7042E-03	3.5006E-02
Zv9_00034838	chr15:23353116-23356845	249.80502	1.14182	0.42689	2.67475	7.4784E-03	4.4565E-02

Pauli_lncRNA	Position (danRer7)	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Zv9_00056252	chr25:35465287-35466284	14.33969	2.19137	0.82151	2.66749	7.6420E-03	4.4629E-02
Zv9_00002565	chr2:1496-2676	27.01832	1.80572	0.67950	2.65740	7.8746E-03	4.5086E-02

TABLE A.4: Differentially expressed lncRNAs in *myl7 nuclear*

Appendix B

circRNA sequences

B.1 Chapter 4

B.1.1 Transcript c119_g1_i1 containing *circ-cyt1*

TCTCCAGGCTGCGCACCTTCTCCAGGTAGGAGGCCAGACGGTCATTGAGGTTTTGCATGGT
GGCTTTCTCATTTGTTGGCACGTCAATGGCATCAGACAGGTTGAAGCCAGCACCAGCACCA
CCACCAAATCCTGCGCCCCCACCATGGGTGCGTGCTGCCAAGAGTTCCTCCTCGTGGTTCT
TCTTGAGGAAGATCAGTTCTTCTTTCAGACCCTCGATCTGCAACTCCAGGTCAGATCTGGT
CATTGTGAGCTCATCCAGCACCCTCCTCAAGCCAACAATGTCTGCTTCCACAGACTGCCTC
ATGCTCAGCTCATTCTCATAC[TTgACTCTGAAGTCATCtGCAGCCAGCTTGGCATTGTCA
ATGCTGAGGTAGATGCCTCCATTTACACCAGTAGCATGGTGGATCTTTGCCTGGAGGTCAC
TGATTGTTGCATAGTAGGCACTGTAGTCACGAGCAGTAGGCGTGGCCTTGCTGTCCAGGAA
CTGGCGGATCTTCAGCTCAAGATCTGCATTGGCCTTCTCCAGGCTGCGCACCTTCTCCAGG
TAGGAGGCCAGACGGTCATTGAGGTTTTGCATGGTGGCTTTCTCATTTGTTGGCACGTCAA
TGGCATCAGACAGGTTGAAGCCAGCACCAGCACCACCACCAAATCCTGCGCCCCCACCATG
GGTGCGTGCTGCCAAGAGTTCCTCCTCGTGGTTCTTCTTGAGGAAGATCAGTTCTTCTTTC
AGACCCTCGATCTGCAACTCCAGGTCAGATCTGGTCATTGTGAGCTCATCCAGCACCCTCC
TCAAGCCAACAATGTCTGCTTCCACAGACTGCCTCATGCTCAGCTCATTCTCATAC]TTGA
CTCTGAAGTCATCTGCAGCCAGCTTGGCATTGTCAATGCTGAGGTAGATGCCTCCATTTAC
ACCAGTAGCATGGTGGATCTTTGCCTGGAGGTCACTGATTGTTGCATAGTAGGCACTGTAG
TCACGAGCAGTAGGCGTGGCCTTGCTGTCCAGGAACTGGCGGATCTTCAGCTCAAGATCTG
CATTGGCCTTCTCCAGGC

Full sequence of *circ-cyt1* is demarcated within square brackets []. Parts of the full sequence are repeated in the flanking regions outside the brackets.

B.1.2 Transcript c408_g1_i1 containing *circ-smyhc*

[CTCCAGCTTTCTCTTCTTTGCCACTAATTCAGCATTCATTTCCTCTTCATCTTCCAGCCT
CTCAGTCAGCTCTTTGGCTTTGGCCTCAAGCTGGATCTTACTTTTGATTAGACCCTCACAT
CTCTCCTCAGCATCGACAAGATTATCTTGCTCAGACTGCACTGCAAGCTGCAGGTCATTCT
TCTCTTGGAGAAGGGAAACCATTGTTCTCAGTGGCATGTTTCTCCTTCTCCACTTTGGCCA
GAGTGAGCTCCAGATCATCAATGTCTTTCTTGAGTTCAGAACATTCATC][CTCCAGCTTT
CTCTTCTTTGCCACTAATTCAGCATTCATTTCCTCTTCATCTTCCAGCCTCTCAGTCAGCT
CTTTGGCTTTGGCCTCAAGCTGGATCTTACTTTTGATTAGACCCTCACATCTCTCCTCAGC
ATCGACAAGATTATCTTGCTCAGACTGCACTGCAAGCTGCAGGTCATTCTTCTCTTGGAGA
AGGGAAACCATTGTTCTCAGTGGCATGTTTCTCCTTCTCCACTTTGGCCAGAGTGAGCTCC
AGATCATCAATGTCTTTCTTGAGTTCAGAACATTCATC]CTCCAGCTTTCTCTT

Full sequence of *circ-smyhc* is demarcated within square brackets []. Part of the full sequence is repeated in the flanking region outside the brackets.

Bibliography

G. Abe, A. Schouw, M. L. Suster, and K. Kawakami. Transposon-mediated BAC transgenesis in zebrafish. *Nature Protocols*, 6(12):1998–2021, 2011. ISSN 1750-2799. doi: 10.1038/nprot.2011.416.

K. Adelman and J. T. Lis. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *eLife*, 3(e02407):1–25, 2014. ISSN 1471-0056. doi: http://dx.doi.org/10.7554/eLife.02407.

T. Alam, Y. A. Medvedeva, H. Jia, J. B. Brown, L. Lipovich, and V. B. Bajic. Promoter Analysis Reveals Globally Differential Regulation of Human Long Non-Coding RNA and Protein-Coding Genes. *PloS one*, 9(10):e109443, 1 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0109443.

A. E. Almada, X. Wu, A. J. Kriz, C. B. Burge, and P. A. Sharp. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, 499(7458):360–363, 2013. ISSN 0028-0836. doi: 10.1038/nature12349.

C. P. Alves, A. S. Fonseca, B. R. Muys, R. de Barros E Lima Bueno, M. C. Bürger, J. E. S. de Souza, V. Valente, M. A. Zago, and W. A. Silva. The lincRNA Hotair is required for epithelial-to-mesenchymal transition and stemness maintenance of cancer cells lines. *Stem cells (Dayton, Ohio)*, pages 1–12, 9 2013. ISSN 1549-4918. doi: 10.1002/stem.1547.

A. R. Amândio, A. Necsulea, E. Joye, B. Mascrez, and D. Duboule. Hotair Is Dispensible for Mouse Development. *PLoS Genetics*, 12(12):1–27, 2016. ISSN 15537404. doi: 10.1371/journal.pgen.1006232.

N. M. Amin, T. M. Greco, L. M. Kuchenbrod, M. M. Rigney, M.-I. Chung, J. B. Wallingford, I. M. Cristea, and F. L. Conlon. Proteomic profiling of cardiac tissue by isolation of nuclei tagged in specific cell types (INTACT). *Development (Cambridge, England)*, 141(4):962–73, 2014. ISSN 1477-9129. doi: 10.1242/dev.098327.

A. Amsterdam, K. C. Sadler, K. Lai, S. Farrington, R. T. Bronson, J. A. Lees, and N. Hopkins. Many ribosomal protein genes are cancer genes in zebrafish. *PLoS Biology*, 2(5), 2004. ISSN 15449173. doi: 10.1371/journal.pbio.0020139.

S. Anders, P. T. Pyl, and W. Huber. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, 2015. ISSN 14602059. doi: 10.1093/bioinformatics/btu638.

D. M. Anderson, K. M. Anderson, C. L. Chang, C. A. Makarewich, B. R. Nelson, J. R. McAnally, P. Kasaragod, J. M. Shelton, J. Liou, R. Bassel-Duby, and E. N. Olson. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, 160(4):595–606, 2015. ISSN 10974172. doi: 10.1016/j.cell.2015.01.009.

K. M. Anderson, D. M. Anderson, J. R. McAnally, J. M. Shelton, R. Bassel-Duby, and E. N. Olson. Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature*, 2:1–13, 2016. ISSN 0028-0836. doi: 10.1038/nature20128.

R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, a. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. a. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller,

A. R. R. Forrest, P. Carninci, M. Rehli, and A. Sandelin. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493):455–61, 2014. ISSN 1476-4687. doi: 10.1038/nature12787.

R. Andersson, A. Sandelin, and C. G. Danko. A unified architecture of transcriptional regulatory elements. *Trends in Genetics*, 31(8):426–433, 2015. ISSN 13624555. doi: 10.1016/j.tig.2015.05.007.

E. Arner, C. O. Daub, K. Vitting-Seerup, and R. Andersson. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*, 347(6225):1010–1015, 2015.

R. Ashwal-Fluss, M. Meyer, N. R. Pamudurti, A. Ivanov, O. Bartok, M. Hanan, N. Evantal, S. Memczak, N. Rajewsky, and S. Kadener. circRNA Biogenesis Competes with Pre-mRNA Splicing. *Molecular Cell*, 56(1):55–66, 2014. ISSN 10972765. doi: 10.1016/j.molcel.2014.08.019.

T. Auer, K. Duroure, A. D. Cian, J. Concordet, and F. Del Bene. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Research*, pages 142–153, 2014. doi: 10.1101/gr.161638.113.

D. E. Ayer. Histone deacetylases: Transcriptional repression with SINers and NuRDs. *Trends in Cell Biology*, 9(5):193–198, 1999. ISSN 09628924. doi: 10.1016/S0962-8924(99)01536-6.

I. Bahm, E. H. Barriga, A. Frolov, E. Theveneau, P. Frankel, and R. Mayor. PDGF controls contact inhibition of locomotion by regulating N-cadherin during neural crest migration. *Development*, page dev.147926, 2017. ISSN 0950-1991. doi: 10.1242/dev.147926.

R. Bajpai, D. a. Chen, A. Rada-Iglesias, J. Zhang, Y. Xiong, J. Helms, C.-P. Chang, Y. Zhao, T. Swigut, and J. Wysocka. CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, 463(7283):958–62, 2010. ISSN 1476-4687. doi: 10.1038/nature08733.

S. P. Barrett and J. Salzman. Circular RNAs: analysis, expression and potential functions. *Development*, 143(11):1838–1847, 2016. ISSN 0950-1991. doi: 10.1242/dev.128074.

A. R. Bassett, A. Akhtar, D. P. Barlow, A. P. Bird, N. Brockdorff, D. Duboule, A. Ephrussi, A. C. Ferguson-Smith, T. R. Gingeras, W. Haerty, D. R. Higgs, E. a. Miska, and C. P. Ponting. Considerations when investigating lncRNA function in vivo. *eLife*, 3:1–14, 8 2014. ISSN 2050-084X. doi: 10.7554/eLife.03058.

P. W. Becker, N. Sacilotto, S. Nornes, A. Neal, M. O. Thomas, K. Liu, C. Preece, I. Ratnayaka, B. Davies, G. Bou-Gharios, and S. De Val. An intronic Flk1 enhancer directs arterial-specific expression via RBPJ-mediated venous repression. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 36(6):1209–1219, 2016. ISSN 15244636. doi: 10.1161/ATVBAHA.116.307517.

D. Beckett, E. Kovaleva, and P. J. Schatz. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Science*, 8(4):921–929, 2008. ISSN 09618368. doi: 10.1110/ps.8.4.921.

G. Bejerano, C. B. Lowe, N. Ahituv, B. King, A. Siepel, S. R. Salama, E. M. Rubin, W. J. Kent, and D. Haussler. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, 441(7089):87–90, 2006. ISSN 1476-4687. doi: 10.1038/nature04696.

P. Betancur, M. Bronner-Fraser, and T. Sauka-Spengler. Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8):3570–3575, 2010. ISSN 0027-8424. doi: 10.1073/pnas.0906596107.

O. Biehlmaier, Y. Makhankov, and S. C. F. Neuhauss. Impaired retinal differentiation and maintenance in zebrafish laminin mutants. *Investigative Ophthalmology and Visual Science*, 48(6):2887–2894, 2007. ISSN 01460404. doi: 10.1167/iovs.06-1212.

A. N. Boettiger and M. Levine. Synchronous and Stochastic Drosophila Embryo. *Science*, (22):23–25, 2009.

N. Bray, H. Pimentel, P. Melsted, and L. Pachter. Near-optimal RNA-Seq quantification. 34(5):4–8, 2015. ISSN 1087-0156. doi: 10.1038/nbt.3519.

J. A. Briggs, E. J. Wolvetang, J. S. Mattick, J. L. Rinn, and G. Barry. Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. *Neuron*, 88(5):861–877, 2015. ISSN 10974199. doi: 10.1016/j.neuron.2015.09.045.

J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*, 10 (12):1213–8, 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2688.

J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 2015:1–21, 2015. ISSN 19343647. doi: 10.1002/0471142727.mb2129s109.

V. Buffalo. Scythe - A Bayesian adapter trimmer (version 0.994 BETA), 2011. URL `https://github.com/vsbuffalo/scythe`.

E. Buitrago-Delgado, K. Nordin, A. Rao, L. Geary, and C. LaBonne. Shared regulatory programs suggest retention of blastula-stage potential in neural crest cells. *Science (New York, N.Y.)*, 348(6241):1332–1335, 2015. ISSN 1095-9203. doi: 10.1126/science.aaa3655.

T. Buranda, G. M. Jones, J. P. Nolan, J. Keij, G. P. Lopez, and L. A. Sklar. Ligand Receptor Dynamics at Streptavidin-Coated Particle Surfaces: A Flow Cytometric

and Spectrofluorimetric Study. *The Journal of Physical Chemistry B*, 103(17):3399–3410, 1999. ISSN 1520-6106. doi: 10.1021/jp983842h.

E. Calo and J. Wysocka. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell*, 49(5):825–837, 2013. ISSN 10972765. doi: 10.1016/j.molcel.2013.01.038.

B. Capel, A. Swain, S. Nicolis, A. Hacker, M. Walter, P. Koopman, P. Goodfellow, and R. Lovell-Badge. Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell*, 73(5):1019–1030, 1993. ISSN 00928674. doi: 10.1016/0092-8674(93)90279-Y.

T. J. Carney, K. A. Dutton, E. Greenhill, M. Delfino-Machín, P. Dufourcq, P. Blader, and R. N. Kelsh. A direct role for Sox10 in specification of neural crest-derived sensory neurons. *Development (Cambridge, England)*, 133(23):4619–4630, 2006. ISSN 0950-1991. doi: 10.1242/dev.02668.

P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, ..., H. Suzuki, J. Kawai, and Y. Hayashizaki. The transcriptional landscape of the mammalian genome. *Science (New York, N.Y.)*, 309(5740):1559–1563, 9 2005. ISSN 1095-9203. doi: 10.1126/science.1112014.

S. J. C. Caron, D. Prober, M. Choy, and A. F. Schier. In vivo birthdating by BAPTISM reveals that trigeminal sensory neuron diversity depends on early neurogenesis. *Development*, 135(19):3259–3269, 2008. ISSN 0950-1991. doi: 10.1242/dev.023200.

R. Cerny, M. Cattell, T. Sauka-Spengler, M. Bronner-Fraser, F. Yu, and D. M. Medeiros. Evidence for the prepattern/cooption model of vertebrate jaw evolution. *Proc Natl Acad Sci U S A*, 107(40):17262–17267, 2010. ISSN 1091-6490. doi: 10.1073/pnas.1009304107.

V. Chalei, S. N. Sansom, L. Kong, S. Lee, J. F. Montiel, K. W. Vance, and C. P. Ponting. The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. *eLife*, 3:e04530, 2014. ISSN 2050084X. doi: 10.7554/eLife.04530.

L.-l. Chen and G. G. Carmichael. Decoding the function of nuclear long non-coding RNAs. *Current Opinion in Cell Biology*, 22(3):357–364, 2010. ISSN 0955-0674. doi: 10.1016/j.ceb.2010.03.003.

T. Chen and S. Y. R. Dent. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nature reviews. Genetics*, 15(2):93–106, 2014. ISSN 1471-0064. doi: 10.1038/nrg3607.

G.-L. Chew, A. Pauli, J. L. Rinn, A. Regev, A. F. Schier, and E. Valen. Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. *Development (Cambridge, England)*, 140(13):2828–34, 7 2013. ISSN 1477-9129. doi: 10.1242/dev.098343.

C. Chu, K. Qu, F. L. Zhong, S. E. Artandi, and H. Y. Chang. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Molecular Cell*, 44(4):667–678, 2011. ISSN 10972765. doi: 10.1016/j.molcel.2011.08.027.

B. D. Clarke, D. M. Cummins, K. a. McColl, A. C. Ward, and T. J. Doran. Characterization of Zebrafish Polymerase III Promoters for the Expression of Short-Hairpin RNA Interference Molecules. *Zebrafish*, 10(4):121002061007008, 2012. ISSN 1545-8547. doi: 10.1089/zeb.2012.0782.

M. R. Clay and M. C. Halloran. Cadherin 6 promotes neural crest cell detachment via F-actin regulation and influences active Rho distribution during epithelial-to-mesenchymal transition. *Development (Cambridge, England)*, 141(12):2506–15, 2014. ISSN 1477-9129. doi: 10.1242/dev.105551.

R. Close, S. Toro, J. A. Martial, and M. Muller. Expression of the zinc finger Egr1 gene during zebrafish embryonic development. *Mechanisms of Development*, 118 (1-2):269–272, 2002. ISSN 09254773. doi: 10.1016/S0925-4773(02)00283-6.

S. J. Conn, K. A. Pillman, J. Toubia, V. M. Conn, M. Salmanidis, C. A. Phillips, S. Roslan, A. W. Schreiber, P. A. Gregory, and G. J. Goodall. The RNA Binding Protein Quaking Regulates Formation of circRNAs. *Cell*, 160(6):1125–1134, 2015. ISSN 00928674. doi: 10.1016/j.cell.2015.02.014.

L. J. Core, J. J. Waterfall, and J. T. Lis. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*, 322(DECEMBER): 1845–1848, 2008. ISSN 0036-8075. doi: 10.1126/science.1162228.

L. J. Core, A. L. Martins, C. G. Danko, C. T. Waters, A. Siepel, and J. T. Lis. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, 46(12):1311–20, 2014. ISSN 1061-4036. doi: 10.1038/ng.3142.

S. Cowley, R. Kang, J. Frangioni, J. Yada, A. DeGrand, I. Radhakrishnan, and R. Eisenman. Functional analysis of the Mad1-mSin3A repressor-corepressor interaction reveals determinants of specificity, affinity, and transcriptional response. *Mol Cell Biol*, 24(7):2698–2709, 2004. ISSN 0270-7306. doi: 10.1128/MCB.24.7.2698-2709.2004.

J. Dalcq, V. Pasque, A. Ghaye, A. Larbuisson, P. Motte, J. A. Martial, and M. Muller. RUNX3, EGR1 and SOX9B Form a Regulatory Cascade Required to Modulate BMP-Signaling during Cranial Cartilage Development in Zebrafish. *PLoS ONE*, 7 (11), 2012. ISSN 19326203. doi: 10.1371/journal.pone.0050140.

E. de Boer, P. Rodriguez, E. Bonte, J. Krijgsveld, E. Katsantoni, A. Heck, F. Grosveld, and J. Strouboulis. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proceedings of*

the National Academy of Sciences, 100(13):7480–7485, 2003. ISSN 0027-8424. doi: 10.1073/pnas.1332608100.

F. de Santa, I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B. K. Tusi, H. Muller, J. Ragoussis, C. L. Wei, and G. Natoli. A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biology*, 8(5), 2010. ISSN 15449173. doi: 10.1371/journal.pbio.1000384.

R. B. Deal and S. Henikoff. A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. *Developmental Cell*, 18(6):1030–1040, 2010. ISSN 1534-5807. doi: 10.1016/j.devcel.2010.05.013.

R. B. Deal and S. Henikoff. The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. *Nature protocols*, 6(1):56–68, 2011. ISSN 1754-2189. doi: 10.1038/nprot.2010.175.

T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. a. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow, and R. Guigó. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research*, 22(9): 1775–89, 9 2012. ISSN 1549-5469. doi: 10.1101/gr.132159.111.

H. Dhiman, S. Kapoor, A. Sivadas, S. Sivasubbu, and V. Scaria. zflncRNApedia: A Comprehensive Online Resource for Zebrafish Long Non-Coding RNAs. *PloS one*, 10(6):e0129997, 2015. ISSN 1932-6203 (Electronic). doi: 10.1371/journal.pone.0129997.

Y. Ding, Y. Tang, C. K. Kwok, Y. Zhang, P. C. Bevilacqua, and S. M. Assmann. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, 000(000):1–5, 11 2013. ISSN 1476-4687. doi: 10.1038/nature12756.

S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, ..., P. Carninci, R. Guigo, and T. R. Gingeras. Landscape of transcription in human cells. *Nature*, 489(7414):101–108, 2012. ISSN 0028-0836. doi: 10.1038/nature11233.

P. C. Donoghue, A. Graham, and R. N. Kelsh. The origin and evolution of the neural crest. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 30 (6):530–541, 6 2008. ISSN 1521-1878. doi: 10.1002/bies.20767.

E. Dupin and L. Sommer. Neural crest progenitors and stem cells: From early development to adulthood. *Developmental Biology*, 366(1):83–95, 2012. ISSN 00121606. doi: 10.1016/j.ydbio.2012.02.035.

K. Dutton, L. Abbas, J. Spencer, C. Brannon, C. Mowbray, M. Nikaido, R. N. Kelsh, and T. T. Whitfield. A zebrafish model for Waardenburg syndrome type IV reveals diverse roles for Sox10 in the otic vesicle. *Disease models & mechanisms*, 2(1-2):68–83, 2009. ISSN 1754-8403. doi: 10.1242/dmm.001164.

K. A. Dutton, A. Pauliny, S. S. Lopes, S. Elworthy, T. J. Carney, J. Rauch, R. Geisler, P. Haffter, and R. N. Kelsh. Zebrafish colourless encodes sox10 and species non-ectomesenchymal neural crest fates. *Development*, 128:4113–4125, 2001. ISSN 0950-1991. doi: 10.1093/hmg/4.12.2407.

C. E. Ellison and D. Bachtrog. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science*, 342:846–850, 2013.

S. Elworthy, J. A. Lister, T. J. Carney, D. W. Raible, and R. N. Kelsh. Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. *Development (Cambridge, England)*, 130:2809–2818, 2003. ISSN 09501991. doi: 10.1242/dev.00461.

A. Emelyanov and S. Parinov. Mifepristone-inducible LexPR system to drive and control gene expression in transgenic zebrafish. *Developmental Biology*, 320(1):113–121, 2008. ISSN 00121606. doi: 10.1016/j.ydbio.2008.04.042.

A. Emelyanov, Y. Gao, N. I. Naqvi, and S. Parinov. Trans-kingdom transposition of the maize Dissociation element. *Genetics*, 174(3):1095–1104, 2006. ISSN 00166731. doi: 10.1534/genetics.106.061184.

A. J. M. Engreitz, J. E. Haines, G. Munson, J. Chen, M. Elizabeth, M. Kane, P. E. Mcdonel, M. Guttman, and E. S. Lander. Neighborhood regulation by lncRNA promoters, transcription, and splicing. *bioRxiv*, pages 1–15, 2016a. ISSN 0028-0836. doi: 10.1101/050948.

J. M. Engreitz, A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath, and M. Guttman. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science (New York, N.Y.)*, 341(6147):1237973, 8 2013. ISSN 1095-9203. doi: 10.1126/science.1237973.

J. M. Engreitz, N. Ollikainen, and M. Guttman. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nature Reviews Molecular Cell Biology*, 17(12):756–770, 2016b. ISSN 1471-0072. doi: 10.1038/nrm.2016.126.

H. C. Etchevers, C. Vincent, N. M. Le Douarin, and G. F. Couly. The cephalic neural crest provides pericytes and smooth muscle cells to all blood vessels of the face and forebrain. *Development (Cambridge, England)*, 128(7):1059–1068, 2001. ISSN 0950-1991. doi: 10.7554/elife.10036.016.

R. R. Fall. Analysis of Microbial Biotin Proteins. *Methods in Enzymology*, 62:390–398, 1979.

G. J. Faulkner, Y. Kimura, C. O. Daub, S. Wani, C. Plessy, K. M. Irvine, K. Schroder, N. Cloonan, A. L. Steptoe, T. Lassmann, K. Waki, N. Hornig, T. Arakawa, H. Takahashi, J. Kawai, A. R. R. Forrest, H. Suzuki, Y. Hayashizaki, D. A. Hume, V. Orlando, S. M. Grimmond, and P. Carninci. The regulated retrotransposon transcriptome of mammalian cells. *Nature genetics*, 41(5):563–71, 2009. ISSN 1546-1718. doi: 10.1038/ng.368.

N. Fedoroff, S. Wessler, and M. Shure. Isolation of the transposable maize controlling elements Ac and Ds. *Cell*, 35(1):235–242, 1983. ISSN 00928674. doi: 10.1016/0092-8674(83)90226-X.

M. Finzsch, S. Schreiner, T. Kichko, P. Reeh, E. R. Tamm, M. R. Bösl, D. Meijer, and M. Wegner. Sox10 is required for Schwann cell identity and progression beyond the immature Schwann cell stage. *Journal of Cell Biology*, 189(4):701–712, 2010. ISSN 00219525. doi: 10.1083/jcb.200912142.

B. Gaertner, J. Johnston, K. Chen, N. Wallaschek, A. Paulson, A. S. Garruss, K. Gaudenz, B. De Kumar, R. Krumlauf, and J. Zeitlinger. Poised RNA Polymerase II Changes over Developmental Time and Prepares Genes for Future Expression. *Cell Reports*, 2(6):1670–1683, 2012. ISSN 22111247. doi: 10.1016/j.celrep.2012.11.024.

Y. Gao, J. Wang, Y. Zheng, J. Zhang, S. Chen, and F. Zhao. Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nature communications*, 7(May):12060, 2016. ISSN 2041-1723. doi: 10.1038/ncomms12060.

J. L. Garcia-Perez, T. J. Widmann, and I. R. Adams. The impact of transposable elements on mammalian development. *Development*, 143(22):4101–4114, 2016. ISSN 0950-1991. doi: 10.1242/dev.132639.

D. Gavriouchkina, R. M. Williams, M. Lukoseviciute, T. Hochgreb-Hägele, V. Chong-Morrison, S. Thongjuea, E. Repapi, and A. Mead. From pioneer to

repressor: Bimodal foxd3 activity dynamically remodels neural crest regulatory landscape in vivo. *bioRxiv*, 2017. doi: 10.1101/213611.

S. Gessert and M. Kühl. The multiple phases and faces of Wnt signaling during cardiac differentiation and development. *Circulation Research*, 107(2):186–199, 2010. ISSN 00097330. doi: 10.1161/CIRCRESAHA.110.221531.

D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. a. Hutchison, H. O. Smith, C. A. H. Iii, and N. America. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, 2009. ISSN 1548-7105. doi: 10.1038/nmeth.1318.

W. D. Gifford, S. L. Pfaff, and T. S. MacFarlan. Transposable elements as genetic regulatory substrates in early development. *Trends in Cell Biology*, 23(5):218–226, 2013. ISSN 09628924. doi: 10.1016/j.tcb.2013.01.001.

L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, and L. S. Qi. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2):442–51, 7 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.06.044.

D. T. Gilmour, H. M. Maischein, and C. Nüsslein-Volhard. Migration and function of a glial subtype in the vertebrate peripheral nervous system. *Neuron*, 34(4):577–588, 2002. ISSN 08966273. doi: 10.1016/S0896-6273(02)00683-9.

D. Giovannone, M. Reyes, R. Reyes, L. Correa, D. Martinez, H. Ra, G. Gomez, J. Kaiser, L. Ma, M. P. Stein, and M. E. De Bellard. Slits affect the timely migration of neural crest cells via robo receptor. *Developmental Dynamics*, 241(8):1274–1288, 2012. ISSN 10588388. doi: 10.1002/dvdy.23817.

J. Göke, X. Lu, Y. S. Chan, H. H. Ng, L. H. Ly, F. Sachs, and I. Szczerbinska. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific

populations of early human embryonic cells. *Cell Stem Cell*, 16(2):135–141, 2015. ISSN 18759777. doi: 10.1016/j.stem.2015.01.005.

M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7):644–652, 2011. ISSN 1087-0156. doi: 10.1038/nbt.1883.

A. Graham. The neural crest. *Current Biology*, 13(10):381–384, 2003.

P. Grote, L. Wittler, D. Hendrix, F. Koch, S. Währisch, A. Beisaw, K. Macura, G. Bläss, M. Kellis, M. Werber, and B. G. Herrmann. The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental Cell*, 24(2):206–214, 2013. ISSN 15345807. doi: 10.1016/j.devcel.2012.12.012.

J. C. Guimaraes and M. Zavolan. Patterns of ribosomal protein expression specify normal and malignant human cells. *Genome Biology*, 17(1):236, 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1104-z.

J. U. Guo, V. Agarwal, H. Guo, and D. P. Bartel. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol*, 15(7):409, 2014. ISSN 1474-760X (Electronic)\r1474-7596 (Linking). doi: 10.1186/s13059-014-0409-z.

R. A. Gupta, N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M.-C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar, and H. Y. Chang. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, 464(7291):1071–6, 4 2010. ISSN 1476-4687. doi: 10.1038/nature08975.

M. Guttman and J. L. Rinn. Modular regulatory principles of large non-coding RNAs. *Nature*, 482(7385):339–346, 2 2012. ISSN 1476-4687. doi: 10.1038/nature10887.

M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, and E. S. Lander. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 458(7235):223–227, 3 2009. ISSN 1476-4687. doi: 10.1038/nature07672.

M. Guttman, M. Garber, J. Z. Levin, J. Donaghey, J. Robinson, X. Adiconis, L. Fan, M. J. Koziol, A. Gnirke, C. Nusbaum, J. L. Rinn, E. S. Lander, and A. Regev. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature biotechnology*, 28(5):503–510, 5 2010. ISSN 1546-1696. doi: 10.1038/nbt.1633.

M. Guttman, J. Donaghey, B. W. Carey, M. Garber, J. K. Grenier, G. Munson, G. Young, A. B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J. L. Rinn, D. E. Root, and E. S. Lander. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, 477(7364):295–300, 9 2011. ISSN 1476-4687. doi: 10.1038/nature10398.

M. Guttman, P. Russell, N. T. Ingolia, J. S. Weissman, and E. S. Lander. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*, 154(1):240–251, 7 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.06.009.

N. Hamazaki, M. Uesaka, K. Nakashima, K. Agata, and T. Imamura. Gene activation-associated long noncoding RNAs function in mouse preimplantation development. *Development*, (January):1–11, 2015. ISSN 0950-1991. doi: 10.1242/dev.116996.

A. Hamilton and D. Baulcombe. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286(5441):1997–2000, 1999. ISSN 00368075. doi: 10.1126/science.286.5441.950.

P. Han, W. Li, C.-H. Lin, J. Yang, C. Shang, S. T. Nurnberg, K. K. Jin, W. Xu, C.-Y. Lin, C.-J. Lin, Y. Xiong, H.-C. Chien, B. Zhou, E. Ashley, D. Bernstein, P.-S. Chen, H.-S. V. Chen, T. Quertermous, and C.-P. Chang. A long noncoding RNA protects the heart from pathological hypertrophy. *Nature*, 514(7520):102–6, 2014. ISSN 1476-4687. doi: 10.1038/nature13596.

T. B. Hansen, T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen, C. K. Damgaard, and J. Kjems. Natural RNA circles function as efficient microRNA sponges. *Nature*, 495 VN -(7441):384–388, 2013. ISSN 0028-0836. doi: 10.1038/nature11993.

T. B. Hansen, M. T. Veno, C. K. Damgaard, and J. Kjems. Comparison of circular RNA prediction tools. *Nucleic Acids Research*, 44(6), 2015. ISSN 13624962. doi: 10.1093/nar/gkv1458.

L. He and G. J. Hannon. MicroRNAs: small RNAs with a big role in gene regulation. *Nature reviews. Genetics*, 5(7):522–531, 2004. ISSN 1471-0056. doi: 10.1038/nrg1415.

Z. He, H. Bammann, D. Han, G. Xie, and P. Khaitovich. Conserved expression of lincRNA during human and macaque prefrontal cortex development and maturation. *RNA (New York, N.Y.)*, 20(7):1103–11, 7 2014. ISSN 1469-9001. doi: 10.1261/rna.043075.113.

M. Heiman, A. Schaefer, S. Gong, J. D. Peterson, M. Day, K. E. Ramsey, M. Suárez-Fariñas, C. Schwarz, D. a. Stephan, D. J. Surmeier, P. Greengard, and N. Heintz. A translational profiling approach for the molecular characterization of CNS cell types. *Cell*, 135(4):738–748, 11 2008. ISSN 1097-4172. doi: 10.1016/j.cell.2008.10.028.

M. Heiman, R. Kulicke, R. J. Fenster, P. Greengard, and N. Heintz. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nature protocols*, 9(6):1282–91, 2014. ISSN 1750-2799. doi: 10.1038/nprot.2014.085.

W. A. Hendrickson, A. Pähler, J. L. Smith, Y. Satow, E. A. Merritt, and R. P. Phizackerley. Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proceedings of the National Academy of Sciences of the United States of America*, 86(7):2190–4, 1989. ISSN 0027-8424. doi: 10.1073/pnas.86.7.2190.

T. Henriques, D. Gilchrist, S. Nechaev, M. Bern, G. Muse, A. Burkholder, D. Fargo, and K. Adelman. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Molecular Cell*, 52(4):517–528, 2013. ISSN 10972765. doi: 10.1016/j.molcel.2013.10.001.

Y. Hisano, T. Sakuma, S. Nakade, R. Ohga, S. Ota, H. Okamoto, T. Yamamoto, and A. Kawahara. Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Scientific reports*, 5:8841, 2015. ISSN 2045-2322. doi: 10.1038/srep08841.

T. Hochgreb-Hägele and M. E. Bronner. A novel FoxD3 gene trap line reveals neural crest precursor movement and a role for FoxD3 in their specification. *Developmental biology*, 374(1):1–11, 2 2013. ISSN 1095-564X. doi: 10.1016/j.ydbio.2012.11.035.

T. Hochgreb-Hägele, C. Yin, D. E. S. Koo, M. E. Bronner, and D. Y. R. Stainier. Laminin β1a controls distinct steps during the establishment of digestive organ laterality. *Development*, 140(13):2734–2745, 2013. ISSN 1477-9129. doi: 10.1242/dev.097618.

C.-C. Hon, J. A. Ramilowski, J. Harshbarger, N. Bertin, O. J. L. Rackham, J. Gough, E. Denisenko, S. Schmeier, T. M. Poulsen, J. Severin, M. Lizio, H. Kawaji, T. Kasukawa, M. Itoh, A. M. Burroughs, S. Noma, S. Djebali, T. Alam, Y. A. Medvedeva, A. C. Testa, L. Lipovich, C.-W. Yip, I. Abugessaisa, M. Mendez, A. Hasegawa,

D. Tang, T. Lassmann, P. Heutink, M. Babina, C. A. Wells, S. Kojima, Y. Nakamura, H. Suzuki, C. O. Daub, M. J. L. de Hoon, E. Arner, Y. Hayashizaki, P. Carninci, and A. R. R. Forrest. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, 543(7644):199–204, 2017. ISSN 0028-0836. doi: 10.1038/nature21374.

A. K. Hopper, H. M. Traglia, and R. W. Dunst. The yeast RNA1 gene product necessary for RNA processing is located in the cytosol and apparently excluded from the nucleus. *Journal of Cell Biology*, 111(2):309–321, 1990. ISSN 00219525. doi: 10.1083/jcb.111.2.309.

M. P. Housley, S. Reischauer, M. Dieu, M. Raes, D. Y. R. Stainier, and B. Vanhollebeke. Translational profiling through biotinylation of tagged ribosomes in zebrafish. *Development (Cambridge, England)*, 141(20):3988–93, 10 2014. ISSN 1477-9129. doi: 10.1242/dev.111849.

M. Howarth, Wenhao Liu, S. Puthenveetil, Yi Zheng, L. F. Marshall, M. M. Schmidt, K. D. Wittrup, M. G. Bawendi, and A. Y. Ting. Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nature Methods*, 5(5):397–399, 2008. ISSN 15487091. doi: 10.1038/nmeth.1206.

K. Howe, M. D. Clark, ..., J. Rogers, and D. L. Stemple. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446): 498–503, 2013. ISSN 1476-4687. doi: 10.1038/nature12111.

C.-L. Hsieh, T. Fei, Y. Chen, T. Li, Y. Gao, X. Wang, T. Sun, C. J. Sweeney, G.-S. M. Lee, S. Chen, S. P. Balk, X. S. Liu, M. Brown, and P. W. Kantoff. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20):7319–24, 2014. ISSN 1091-6490. doi: 10.1073/pnas.1324151111.

C. Y. Hu, C. H. Yang, W. Y. Chen, C. J. Huang, H. Y. Huang, M. S. Chen, and H. J. Tsai. Egr1 gene knockdown affects embryonic ocular development in zebrafish. *Mol Vis*, 12(September 2005):1250–1258, 2006. ISSN 1090-0535. doi: v12/a142 [pii].

G. Hu, Q. Tang, S. Sharma, F. Yu, T. M. Escobar, S. a. Muljo, J. Zhu, and K. Zhao. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nature immunology*, 14(11):1190–8, 11 2013. ISSN 1529-2916. doi: 10.1038/ni.2712.

M. Huarte, M. Guttman, D. Feldser, M. Garber, M. J. Koziol, D. Kenzelmann-Broz, A. M. Khalil, O. Zuk, I. Amit, M. Rabani, L. D. Attardi, A. Regev, E. S. Lander, T. Jacks, and J. L. Rinn. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, 142(3):409–419, 2010. ISSN 00928674. doi: 10.1016/j.cell.2010.06.040.

N. E. IIott, J. a. Heward, B. Roux, E. Tsitsiou, P. S. Fenwick, L. Lenzi, I. Goodhead, C. Hertz-Fowler, A. Heger, N. Hall, L. E. Donnelly, D. Sims, and M. a. Lindsay. Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. *Nature communications*, 5: 3979, 1 2014. ISSN 2041-1723. doi: 10.1038/ncomms4979.

V. B. Indjeian, G. A. Kingman, F. C. Jones, C. A. Guenther, J. Grimwood, J. Schmutz, R. M. Myers, and D. M. Kingsley. Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell*, 164(1-2):45–56, 2016. ISSN 10974172. doi: 10.1016/j.cell.2015.12.007.

M. Ingaramo and D. Beckett. Biotinylation, a post-translational modification controlled by the rate of protein-protein association. *Journal of Biological Chemistry*, 286(15):13071–13078, 2011. ISSN 00219258. doi: 10.1074/jbc.M110.183624.

N. T. Ingolia. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature reviews. Genetics*, 15(3):205–13, 2014. ISSN 1471-0064. doi: 10.1038/nrg3645.

N. T. Ingolia, G. a. Brar, S. Rouskin, A. M. McGeachy, and J. S. Weissman. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing

of ribosome-protected mRNA fragments. *Nature Protocols*, 7(8):1534–1550, 8 2012. ISSN 1750-2799. doi: 10.1038/nprot.2012.086.

N. T. Ingolia, G. A. Brar, N. Stern-Ginossar, M. S. Harris, G. J. Talhouarne, S. E. Jackson, M. R. Wills, and J. S. Weissman. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Reports*, pages 1365–1379, 8 2014. ISSN 22111247. doi: 10.1016/j.celrep.2014.07.045.

A. Ivanov, S. Memczak, C. Dieterich, N. Rajewsky, E. Wyler, F. Torti, H. T. Porath, M. R. Orejuela, M. Piechotta, E. Y. Levanon, and M. Landthaler. Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals. *Cell Reports*, 10:170–177, 2015. ISSN 2211-1247. doi: 10.1016/j.celrep.2014.12.019.

Y. W. Iwasaki, M. C. Siomi, and H. Siomi. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annual review of biochemistry*, 84:405–33, 2015. ISSN 1545-4509. doi: 10.1146/annurev-biochem-060614-034258.

B. T. Jacques-Fricke and L. S. Gammill. Neural crest specification and migration independently require NSD3-related lysine methyltransferase activity. *Molecular Biology of the Cell*, pages 13–12, 2014. ISSN 1059-1524, 1939-4586. doi: 10.1091/mbc.E13-12-0744.

P. Ji, S. Diederichs, W. Wang, S. Boïng, R. Metzger, P. M. Schneider, N. Tidow, B. Brandt, H. Buerger, E. Bulk, M. Thomas, W. E. Berdel, H. Serve, and C. Muüler-Tidow. MALAT-1, a novel noncoding RNA, and thymosin b4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, 22(June):8031–8041, 2003. ISSN 0950-9232. doi: 10.1038/sj.onc.1206928.

L. Jia, L. Cheng, and J. Raper. Slit/Robo signaling is necessary to confine early neural crest cells to the ventral migratory pathway in the trunk. *Developmental Biology*, 282(2):411–421, 2005. ISSN 00121606. doi: 10.1016/j.ydbio.2005.03.021.

J. Joseph, S. H. Tan, T. S. Karpova, J. G. McNally, and M. Dasso. SUMO-1 targets RanGAP1 to kinetochores and mitotic spindles. *Journal of Cell Biology*, 156(4):595–602, 2002. ISSN 00219525. doi: 10.1083/jcb.200110109.

N. Joshi and J. Fass. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33), 2011. URL `https://github.com/najoshi/sickle`.

M. U. Kaikkonen, N. J. Spann, S. Heinz, C. E. Romanoski, K. a. Allison, J. D. Stender, H. B. Chun, D. F. Tough, R. K. Prinjha, C. Benner, and C. K. Glass. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular Cell*, 51(3):310–325, 2013. ISSN 10972765. doi: 10.1016/j.molcel.2013.07.010.

D. Kalderon, B. L. Roberts, W. D. Richardson, and A. E. Smith. A short amino acid sequence able to specify nuclear location. *Cell*, 39(3):499–509, 1984. ISSN 00928674. doi: 10.1016/0092-8674(84)90457-4.

R. O. Karlstrom, T. Trowe, S. Klostermann, H. Baier, M. Brand, a. D. Crawford, B. Grunewald, P. Haffter, H. Hoffmann, S. U. Meyer, B. K. Müller, S. Richter, F. J. van Eeden, C. Nüsslein-Volhard, and F. Bonhoeffer. Zebrafish mutations affecting retinotectal axon pathfinding. *Development (Cambridge, England)*, 123:427–438, 1996. ISSN 0950-1991.

K. Kawakami. Transgenesis and Gene Trap Methods in Zebrafish by Using the Tol2 Transposable Element. *Methods in Cell Biology*, 77:201–222, 2004. ISSN 0091679X. doi: 10.1016/S0091-679X(04)77011-9.

K. Kawakami. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biology*, 8(Suppl 1):S7, 2007. ISSN 14656906. doi: 10.1186/gb-2007-8-s1-s7.

D. Kelley and J. Rinn. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology*, 13(11):R107, 2012. ISSN 1465-6906. doi: 10.1186/gb-2012-13-11-r107.

R. N. Kelsh. Sorting out Sox10 functions in neural crest development. *BioEssays*, 28 (8):788–798, 2006. ISSN 02659247. doi: 10.1002/bies.20445.

W. J. Kent. BLAT - The BLAST-like alignment tool. *Genome Research*, 12(4):656–664, 2002. ISSN 10889051. doi: 10.1101/gr.229202. Article published online before March 2002.

A. Kenyon, D. Gavriouchkina, J. Zorman, G. Napolitani, V. Cerundolo, and T. Sauka-Spengler. Active nuclear transcriptome analysis reveals inflammasome-dependent mechanism for early neutrophil response to Mycobacterium marinum. *Scientific Reports*, 7(1):6505, 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-06099-x.

A. M. Khalil, M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B. E. Bernstein, A. van Oudenaarden, A. Regev, E. S. Lander, and J. L. Rinn. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28):11667–11672, 2009.

C. H. Kim, E. Ueshima, O. Muraoka, H. Tanaka, S. Y. Yeo, T. L. Huh, and N. Miki. Zebrafish elav/HuC homologue as a very early neuronal marker. *Neuroscience Letters*, 216(2):109–112, 1996. ISSN 03043940. doi: 10.1016/S0304-3940(96)13021-4.

D. Kim, B. Langmead, and S. L. Salzberg. HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4):357–360, 2015. ISSN 1548-7105. doi: 10.1038/nmeth.3317.

J. Kim, A. B. Cantor, S. H. Orkin, and J. Wang. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem

cells. *Nat Protoc*, 4(4):506–517, 2009. ISSN 1754-2189. doi: nprot.2009.23 [pii]\r10.1038/nprot.2009.23.

T. K. Kim and R. Shiekhattar. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell*, 162(5):948–959, 2015. ISSN 10974172. doi: 10.1016/j.cell.2015.08.008.

T.-K. Kim, M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, and M. E. Greenberg. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465(7295):182–7, 5 2010. ISSN 1476-4687. doi: 10.1038/nature09033.

C. A. Klattenhoff, J. C. Scheuermann, L. E. Surface, R. K. Bradley, P. A. Fields, M. L. Steinhauser, H. Ding, V. L. Butty, L. Torrey, S. Haas, R. Abo, M. Tabebordbar, R. T. Lee, C. B. Burge, and L. A. Boyer. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*, 152(3):570–583, 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.01.003.

H. Knaut, P. Blader, U. Strähle, and A. F. Schier. Assembly of trigeminal sensory ganglia by chemokine signaling. *Neuron*, 47(5):653–666, 2005. ISSN 08966273. doi: 10.1016/j.neuron.2005.07.014.

A. K. Knecht and M. Bronner-Fraser. Induction of the neural crest: a multigene process. *Nature reviews. Genetics*, 3(6):453–461, 6 2002. ISSN 1471-0056. doi: 10.1038/nrg819.

R. D. Knight, S. Nair, S. S. Nelson, A. Afshar, Y. Javidan, R. Geisler, G.-J. Rauch, and T. F. Schilling. Lockjaw Encodes a Zebra Sh Tfap2a Required for Early Neural Crest Development. *Development*, 130(23):5755–5768, 2003. ISSN 0950-1991. doi: 10.1242/dev.00575.

S. Konermann, M. D. Brigham, A. E. Trevino, P. D. Hsu, M. Heidenreich, L. Cong, R. J. Platt, D. a. Scott, G. M. Church, and F. Zhang. Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, 500(7463):472–6, 2013. ISSN 1476-4687. doi: 10.1038/nature12466.

S. Konermann, M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg, H. Nishimasu, O. Nureki, and F. Zhang. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517(7536):583–8, 2015. ISSN 0028-0836. doi: 10.1038/nature14136.

M. Kozak and A. J. Shatkin. Sequences and properties of two ribosome binding sites from the small size class of reovirus messenger RNA. *Journal of Biological Chemistry*, 252(19):6895–6908, 1977. ISSN 00219258.

P. H. L. Krijger and W. de Laat. Regulation of disease-associated gene expression in the 3D genome. *Nature Reviews Molecular Cell Biology*, 17(12):771–782, 2016. ISSN 14710080. doi: 10.1038/nrm.2016.138.

K. M. Kwan, E. Fujimoto, C. Grabher, B. D. Mangum, M. E. Hardy, D. S. Campbell, J. M. Parant, H. J. Yost, J. P. Kanki, and C. B. Chien. The Tol2kit: A multisite gateway-based construction Kit for Tol2 transposon transgenesis constructs. *Developmental Dynamics*, 236(11):3088–3099, 2007. ISSN 10588388. doi: 10.1002/dvdy.21343.

F. Lai, U. A. Orom, M. Cesaroni, M. Beringer, D. J. Taatjes, G. A. Blobel, and R. Shiekhattar. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*, 494(7438):497–501, 2013. ISSN 1476-4687. doi: 10.1038/nature11884.

M. T. Y. Lam, H. Cho, H. P. Lesch, D. Gosselin, S. Heinz, Y. Tanaka-Oishi, C. Benner, M. U. Kaikkonen, A. S. Kim, M. Kosaka, C. Y. Lee, A. Watt, T. R. Grossman, M. G.

Rosenfeld, R. M. Evans, and C. K. Glass. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, 498(7455):511–5, 2013. ISSN 1476-4687. doi: 10.1038/nature12209.

K. Landgraf, F. Bollig, M.-O. Trowe, B. Besenbeck, C. Ebert, D. Kruspe, A. Kispert, F. Hänel, and C. Englert. Sipl1 and Rbck1 are novel Eya1-binding proteins with a role in craniofacial development. *Molecular and cellular biology*, 30(24):5764–75, 2010. ISSN 1098-5549. doi: 10.1128/MCB.01645-09.

B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 19(3):R25, 2009. ISSN 1465-6914. doi: 10.1186/gb-2009-10-3-r25.

N. Le Douarin and C. Kalcheim. *The Neural Crest*. 2nd edition, 1999. ISBN 0521620104.

H. S. Lee, J. A. Simon, and J. T. Lis. Structure and expression of ubiquitin genes of Drosophila melanogaster. *Mol Cell Biol*, 8(11):4727–4735, 1988.

C. Lepoivre, M. Belhocine, A. Bergon, A. Griffon, M. Yammine, L. Vanhille, J. Zacarias-Cabeza, M.-A. Garibal, F. Koch, M. A. Maqbool, R. Fenouil, B. Loriod, H. Holota, M. Gut, I. Gut, J. Imbert, J.-C. Andrau, D. Puthier, and S. Spicuglia. Divergent transcription is associated with promoters of transcriptional regulators. *BMC genomics*, 14:914, 2013. ISSN 1471-2164. doi: 10.1186/1471-2164-14-914.

N. Léveillé, C. A. Melo, K. Rooijers, A. Díaz-Lagares, S. A. Melo, G. Korkmaz, R. Lopes, F. A. Moqadam, A. R. Maia, P. J. Wijchers, G. Geeven, M. L. den Boer, R. Kalluri, W. de Laat, M. Esteller, and R. Agami. Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. *Nature Communications*, 6:6520, 2015. ISSN 2041-1723. doi: 10.1038/ncomms7520.

M. Levine and E. H. Davidson. Gene regulatory networks for development. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14): 4936–4942, 2005. ISSN 0027-8424. doi: 10.1073/pnas.0408031102.

H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009. ISSN 13674803. doi: 10.1093/bioinformatics/btp352.

W. Li, D. Notani, Q. Ma, B. Tanasa, E. Nunez, A. Y. Chen, D. Merkurjev, J. Zhang, K. Ohgi, X. Song, S. Oh, H.-S. Kim, C. K. Glass, and M. G. Rosenfeld. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 498(7455):516–20, 2013. ISSN 1476-4687. doi: 10.1038/nature12210.

Z. Li, C. Huang, C. Bao, L. Chen, M. Lin, X. Wang, G. Zhong, B. Yu, W. Hu, L. Dai, P. Zhu, Z. Chang, Q. Wu, Y. Zhao, Y. Jia, P. Xu, H. Liu, and G. Shan. Exon-intron circular RNAs regulate transcription in the nucleus. *Nature Structural & Molecular Biology*, 22(3):256–264, 2015. ISSN 1545-9993. doi: 10.1038/nsmb.2959.

D. Liang and J. E. Wilusz. Short intronic repeat sequences facilitate circular RNA production. *Genes & Development*, pages 2233–2247, 2014. doi: 10.1101/gad.251926.114.scribed.

C. Liles, M. A. Kumar, and D. Weinshenker. Use of Gene-Specific Primer Cocktails for First-Strand cDNA Synthesis With a Reverse Transcriptase Kit. *American Biotechnology Laboratory*, 2(December):20–21, 2004. ISSN 07493223.

N. Lin, K. Y. Chang, Z. Li, K. Gates, Z. A. Rana, J. Dang, D. Zhang, T. Han, C. S. Yang, T. J. Cunningham, S. R. Head, G. Duester, P. D. S. Dong, and T. M. Rana. An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Molecular Cell*, 53(6):1005–1019, 2014. ISSN 10974164. doi: 10.1016/j.molcel.2014.01.021.

S. J. Liu, S. J. Liu, M. A. Horlbeck, S. W. Cho, H. S. Birk, M. Malatesta, F. J. Attenello, J. E. Villalta, M. Y. Cho, Y. Chen, A. Mandegar, M. P. Olvera, L. A. Gilbert, B. R. Conklin, H. Y. Chang, J. S. Weissman, and D. A. Lim. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. 7111 (December), 2016. ISSN 0036-8075. doi: 10.1126/science.aah7111.

O. Livnah, E. A. Bayer, M. Wilchek, and J. L. Sussman. Three-dimensional structures of avidin and the avidin-biotin complex. *Proceedings of the National Academy of Sciences of the United States of America*, 90(11):5076–5080, 1993. ISSN 0027-8424. doi: 10.1073/pnas.90.11.5076.

M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8.

J. Lu, E. Peatman, H. Tang, J. Lewis, and Z. Liu. Profiling of gene duplication patterns of sequenced teleost genomes: evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications. *BMC Genomics*, 13(1):246, 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-246.

S. Luo, J. Y. Lu, L. Liu, Y. Yin, C. Chen, X. Han, B. Wu, R. Xu, W. Liu, P. Yan, W. Shao, Z. Lu, H. Li, J. Na, F. Tang, J. Wang, Y. E. Zhang, and X. Shen. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell*, 18(5):637–652, 2016. ISSN 18759777. doi: 10.1016/j.stem.2016.01.024.

S. Mader, H. Lee, a. Pause, and N. Sonenberg. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Molecular and cellular biology*, 15(9): 4990–4997, 1995. ISSN 0270-7306. doi: 10.1128/MCB.15.9.4990.

R. Mahajan, C. Delphin, T. Guan, L. Gerace, and F. Melchior. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex

protein RanBP2. *Cell*, 88(1):97–107, 1997. ISSN 00928674. doi: 10.1016/S0092-8674(00)81862-0.

P. Mali, J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, and G. M. Church. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*, 31(9): 833–8, 9 2013. ISSN 1546-1696. doi: 10.1038/nbt.2675.

Y. Mamane, E. Petroulakis, L. Rong, K. Yoshida, L. W. Ler, and N. Sonenberg. eIF4E-from translation to transformation. *Oncogene*, 23(18):3172–3179, 2004. ISSN 0950-9232. doi: 10.1038/sj.onc.1207549.

M. L. Martik and M. E. Bronner. Regulatory Logic Underlying Diversification of the Neural Crest. *Trends in Genetics*, 33(10):715–727, 2017. ISSN 13624555. doi: 10.1016/j.tig.2017.07.015.

S. J. Matkovich, J. R. Edwards, T. C. Grossenheider, C. de Guzman Strong, and G. W. n. Dorn. Epigenetic coordination of embryonic heart transcription by dynamically regulated long noncoding RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 111(33):12264–12269, 2014. ISSN 1091-6490 (Electronic). doi: 10.1073/pnas.1410622111.

M. J. Matunis, E. Coutavas, and G. Blobel. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *Journal of Cell Biology*, 135(6):1457–1470, 1996. ISSN 00219525. doi: 10.1083/jcb.135.6.1457.

M. J. Matunis, J. Wu, and G. Blobel. SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. *Journal of Cell Biology*, 140(3):499–509, 1998. ISSN 00219525. doi: 10.1083/jcb.140.3.499.

B. McClintock. The Origin and Behavior of Mutable Loci in Maize. *Genetics*, 36(6): 344–355, 1950. ISSN 0027-8424. doi: 10.1073/pnas.36.6.344.

C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, and G. Bejerano. GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology*, 28(5):495–501, 2010. ISSN 1087-0156. doi: 10.1038/nbt.1630.

M. Melé, K. Mattioli, W. Mallard, D. M. Shechner, C. Gerhardinger, and J. L. Rinn. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Research*, 27(1):27–37, 2017. ISSN 15495469. doi: 10.1101/gr.214205.116.

C. A. Melo, J. Drost, P. J. Wijchers, H. van de Werken, E. de Wit, J. a. F. O. Vrielink, R. Elkon, S. a. Melo, N. Léveillé, R. Kalluri, W. de Laat, and R. Agami. ERNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. *Molecular Cell*, 49(3):524–535, 2013. ISSN 10972765. doi: 10.1016/j.molcel.2012.11.021.

S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. le Noble, and N. Rajewsky. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495(7441):333–8, 3 2013. ISSN 1476-4687. doi: 10.1038/nature11928.

P. Monnier, C. Martinet, J. Pontis, I. Stancheva, S. Ait-Si-Ali, and L. Dandolo. H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *Proceedings of the National Academy of Sciences of the United States of America*, 110(51):20693–8, 12 2013. ISSN 1091-6490. doi: 10.1073/pnas.1310201110.

M. Montero-Balaguer, M. R. Lang, S. W. Sachdev, C. Knappmeyer, R. A. Stewart, A. De La Guardia, A. K. Hatzopoulos, and E. W. Knapik. The mother superior mutation ablates foxd3 activity in neural crest progenitor cells and depletes neural crest derivatives in zebrafish. *Developmental Dynamics*, 235(12):3199–3212, 2006. ISSN 10588388. doi: 10.1002/dvdy.20959.

C. Mosimann, C. K. Kaufman, P. Li, E. K. Pugach, O. J. Tamplin, and L. I. Zon. Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development (Cambridge, England)*, 138(1):169–177, 1 2011. ISSN 1477-9129. doi: 10.1242/dev.059345.

K. Mousavi, H. Zare, S. Dell'Orso, L. Grontved, G. Gutierrez-Cruz, A. Derfoul, G. Hager, and V. Sartorelli. ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Molecular Cell*, 51(5):606–617, 2013. ISSN 10972765. doi: 10.1016/j.molcel.2013.07.022.

A. Necsulea, M. Soumillon, M. Warnefors, A. Liechti, T. Daish, U. Zeller, J. C. Baker, F. Grützner, and H. Kaessmann. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, 505(7485):635–40, 1 2014. ISSN 1476-4687. doi: 10.1038/nature12943.

C. Nepal, Y. Hadzhiev, C. Previti, V. Haberle, N. Li, H. Takahashi, A. M. M. Suzuki, Y. Sheng, R. F. Abdelhamid, S. Anand, J. Gehrig, A. Akalin, C. E. M. Kockx, A. a. J. van der Sloot, W. F. J. van Ijcken, O. Armant, S. Rastegar, C. Watson, U. Strähle, E. Stupka, P. Carninci, B. Lenhard, and F. Müller. Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome research*, 23(11):1938–50, 11 2013. ISSN 1549-5469. doi: 10.1101/gr.153692.112.

M. Nikaido, E. W. P. Law, and R. N. Kelsh. A Systematic Survey of Expression and Function of Zebrafish frizzled Genes. *PLoS ONE*, 8(1), 2013. ISSN 19326203. doi: 10.1371/journal.pone.0054833.

N. Nikitina, T. Sauka-Spengler, and M. Bronner-Fraser. Dissecting early regulatory relationships in the lamprey neural crest gene network. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51):20083–20088, 2008. ISSN 0027-8424. doi: 10.1073/pnas.0806009105.

J. H. Notwell, T. Chung, W. Heavner, and G. Bejerano. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nature communications*, 6(Mc 5329):6644, 2015. ISSN 2041-1723. doi: 10.1038/ncomms7644.

A. C. Oates, A. E. E. Bruce, and R. K. Ho. Too much interference: injection of double-stranded RNA has nonspecific effects in the zebrafish embryo. *Developmental biology*, 224:20–28, 2000. ISSN 00121606. doi: 10.1006/dbio.2000.9761.

S. L. Ooi, J. G. Henikoff, and S. Henikoff. A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Research*, 38(4): 1–14, 2009. ISSN 03051048. doi: 10.1093/nar/gkp1090.

U. A. Ørom and R. Shiekhattar. Long non-coding RNAs and enhancers. *Current opinion in genetics & development*, 21(2):194–198, 2011. ISSN 0959437X. doi: 10.1016/j.gde.2011.01.020.

U. A. Ørom, T. Derrien, M. Beringer, K. Gumireddy, A. Gardini, G. Bussotti, F. Lai, M. Zytnicki, C. Notredame, Q. Huang, R. Guigo, and R. Shiekhattar. Long noncoding RNAs with enhancer-like function in human cells. *Cell*, 143(1):46–58, 2010. ISSN 00928674. doi: 10.1016/j.cell.2010.09.001.

M. Ouedraogo, C. Bettembourg, A. Bretaudeau, O. Sallou, C. Diot, O. Demeure, and F. Lecerf. The Duplicated Genes Database: Identification and Functional Annotation of Co-Localised Duplicated Genes across Genomes. *PLoS ONE*, 7(11):1–8, 2012. ISSN 19326203. doi: 10.1371/journal.pone.0050653.

S. Ounzain, I. Pezzuto, R. Micheletti, F. Burdet, R. Sheta, M. Nemir, C. Gonzales, A. Sarre, M. Alexanian, M. J. Blow, D. May, R. Johnson, J. Dauvillier, L. a. Pennacchio, and T. Pedrazzini. Functional importance of cardiac enhancer-associated noncoding RNAs in heart development and disease. *Journal of Molecular and Cellular Cardiology*, 76:55–70, 2014. ISSN 00222828. doi: 10.1016/j.yjmcc.2014.08.009.

S. Ounzain, R. Micheletti, C. Arnan, I. Plaisance, D. Cecchi, B. Schroen, F. Reverter, M. Alexanian, C. Gonzales, S. Y. Ng, G. Bussotti, I. Pezzuto, C. Notredame, S. Heymans, R. Guigó, R. Johnson, and T. Pedrazzini. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *Journal of Molecular and Cellular Cardiology*, 89: 98–112, 2015. ISSN 10958584. doi: 10.1016/j.yjmcc.2015.09.016.

K. C. Pang, M. C. Frith, and J. S. Mattick. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in genetics : TIG*, 22(1):1–5, 2006.

V. R. Paralkar, C. C. Taborda, P. Huang, Y. Yao, A. V. Kossenkov, R. Prasad, J. Luan, J. O. Davies, J. R. Hughes, R. C. Hardison, G. A. Blobel, and M. J. Weiss. Unlinking an lncRNA from Its Associated cis Element. *Molecular Cell*, 62(1):104–110, 2016. ISSN 10974164. doi: 10.1016/j.molcel.2016.02.029.

K.-S. Park and B. M. Gumbiner. Cadherin 6B induces BMP signaling and de-epithelialization during the epithelial mesenchymal transition of the neural crest. *Development (Cambridge, England)*, 137(16):2691–701, 2010. ISSN 1477-9129. doi: 10.1242/dev.050096.

A. Pauli, J. L. Rinn, and A. F. Schier. Non-coding RNAs as regulators of embryogenesis. *Nature reviews. Genetics*, 12(2):136–149, 2 2011. ISSN 1471-0064. doi: 10.1038/nrg2904.

A. Pauli, E. Valen, M. F. Lin, M. Garber, N. L. Vastenhouw, J. Z. Levin, L. Fan, A. Sandelin, J. L. Rinn, A. Regev, and A. F. Schier. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome research*, 22 (3):577–91, 3 2012. ISSN 1549-5469. doi: 10.1101/gr.133009.111.

A. Pauli, T. G. Montague, K. A. Lennox, M. A. Behlke, and A. F. Schier. Antisense oligonucleotide-mediated transcript knockdown in zebrafish. *PLoS ONE*, 10(10): 1–14, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0139504.

M. Pavlicev, K. Hiratsuka, K. A. Swaggart, C. Dunn, and L. Muglia. Detecting endogenous retrovirus-driven tissue-specific gene transcription. *Genome Biology and Evolution*, 7(4):1082–1097, 2015. ISSN 17596653. doi: 10.1093/gbe/evv049.

M. Pertea, G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, and S. L. Salzberg. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3):290–295, 2015. ISSN 1087-0156. doi: 10.1038/nbt.3122.

S. Petruk, Y. Sedkov, K. M. Riley, J. Hodgson, F. Schweisguth, S. Hirose, J. B. Jaynes, H. W. Brock, and A. Mazo. Transcription of bxd Noncoding RNAs Promoted by Trithorax Represses Ubx in cis by Transcriptional Interference. *Cell*, 127(6):1209–1221, 2006. ISSN 00928674. doi: 10.1016/j.cell.2006.10.039.

C. A. Piccirillo, E. Bjur, I. Topisirovic, N. Sonenberg, and O. Larsson. Translational control of immune responses: from transcripts to translatomes. *Nature immunology*, 15(6):503–511, 2014. ISSN 1529-2916. doi: 10.1038/ni.2891.

J. Picher, B. Budeus, O. Wafzig, C. Krüger, S. García-Gómez, M. I. Martínez-Jiménez, A. Díaz-Talavera, D. Weber, L. Blanco, and A. Schneider. TruePrime is a novel method for whole-genome amplification from single cells based on TthPrimPol. *Nature Communications*, 7, 2016. ISSN 2041-1723. doi: 10.1038/ncomms13296.

A. Pichler, A. Gast, J. S. Seeler, A. Dejean, and F. Melchior. The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell*, 108(1):109–120, 2002. ISSN 00928674. doi: 10.1016/S0092-8674(01)00633-X.

H. J. Pimentel, N. Bray, S. Puente, P. Melsted, and L. Pachter. Differential analysis of RNA-Seq incorporating quantification uncertainty. *Nat Methods*, 14(7):058164, 2016. ISSN 1548-7105. doi: 10.1101/058164.

M. S. Prasad, T. Sauka-Spengler, and C. LaBonne. Induction of the neural crest state: Control of stem cell attributes by gene regulatory, post-transcriptional and

epigenetic interactions. *Developmental Biology*, 366(1):10–21, 7 2012. ISSN 1095-564X. doi: 10.1016/j.ydbio.2012.03.014.

S. L. Prescott, R. Srinivasan, M. C. Marchetto, I. Grishina, I. Narvaiza, L. Selleri, F. H. Gage, T. Swigut, and J. Wysocka. Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. *Cell*, 163(1):68–84, 2015. ISSN 10974172. doi: 10.1016/j.cell.2015.08.036.

L. S. Qi, M. H. Larson, L. a. Gilbert, J. a. Doudna, J. S. Weissman, A. P. Arkin, and W. a. Lim. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–83, 2 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.02.022.

C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1):590–596, 2013. ISSN 03051048. doi: 10.1093/nar/gks1219.

J. J. Quinn and H. Y. Chang. Unique features of long non-coding RNA biogenesis and function. *Nature reviews. Genetics*, 17(1):47–62, 2015. ISSN 1471-0064. doi: 10.1038/nrg.2015.10.

J. J. Quinn, I. A. Ilik, K. Qu, P. Georgiev, C. Chu, A. Akhtar, and H. Y. Chang. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nature biotechnology*, (November 2013), 7 2014. ISSN 1546-1696. doi: 10.1038/nbt.2943.

M. A. Rabadán, A. Herrera, L. Fanlo, S. Usieto, C. Carmona-Fontaine, E. H. Barriga, R. Mayor, S. Pons, and E. Martí. Delamination of neural crest cells requires transient and reversible Wnt inhibition mediated by Dact1/2. *Development (Cambridge, England)*, 143(12):2194–205, 2016. ISSN 1477-9129. doi: 10.1242/dev.134981.

A. Rada-Iglesias, R. Bajpai, S. Prescott, S. A. Brugmann, T. Swigut, and J. Wysocka. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell*, 11(5):633–648, 2012. ISSN 19345909. doi: 10.1016/j.stem.2012.07.006.

J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. a. Brugmann, L. H. Goodnough, J. a. Helms, P. J. Farnham, E. Segal, and H. Y. Chang. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 129(7):1311–23, 6 2007. ISSN 0092-8674. doi: 10.1016/j.cell.2007.05.022.

T. C. Roberts, K. V. Morris, and M. J. A. Wood. The role of long non-coding RNAs in neurodevelopment, brain function and neurological disease. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1652), 2014. ISSN 1471-2970. doi: 10.1098/rstb.2013.0507.

J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. Integrative Genomics Viewer. *Nature Biotechnology*, 29(1):24–26, 2011. ISSN 1087-0156. doi: 10.1038/nbt.1754.

L. E. E. Roth, R. Kalev-Altman, E. Monsonego-Ornan, and D. Sela-Donenfeld. A new role of the membrane-type matrix metalloproteinase 16 (MMP16 / MT3-MMP) in neural crest cell migration. *International Journal of Developmental Biology*, 256: 245–256, 2017. ISSN 1696-3547. doi: 10.1387/ijdb.160286ds.

J. Salzman. Circular RNA Expression : Its Potential Regulation and Function. *Trends in Genetics*, 32(5):309–316, 2016. ISSN 0168-9525. doi: 10.1016/j.tig.2016.03.002.

J. Salzman, C. Gawad, P. L. Wang, N. Lacayo, and P. O. Brown. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE*, 7(2), 2012. ISSN 19326203. doi: 10.1371/journal.pone.0030733.

M. E. Samuels, P. Schedl, and T. W. Cline. The complex set of late transcripts from the Drosophila sex determination gene sex-lethal encodes multiple related polypeptides. *Molecular and cellular biology*, 11(7):3584–602, 1991. ISSN 0270-7306. doi: 10.1128/MCB.11.7.3584.

N. Sanjana, J. Wright, K. Zheng, O. Shalem, P. Fontanillas, J. Joung, C. Cheng, A. Regev, and F. Zhang. High-resolution interrogation of functional elements in the noncoding genome. *Science*, 353(6307):1545–1549, 2016.

F. Santagati and F. M. Rijli. Cranial neural crest and the building of the vertebrate head. *Nature reviews. Neuroscience*, 4(10):806–818, 10 2003. ISSN 1471-003X. doi: 10.1038/nrn1221.

F. Santoro, D. Mayer, R. M. Klement, K. E. Warczok, A. Stukalov, D. P. Barlow, and F. M. Pauler. Imprinted Igf2r silencing depends on continuous Airn lncRNA expression and is not restricted to a developmental window. *Development (Cambridge, England)*, 140(6):1184–95, 3 2013. ISSN 1477-9129. doi: 10.1242/dev.088849.

T. Sauka-Spengler and M. Bronner-Fraser. A gene regulatory network orchestrates neural crest formation. *Nature reviews. Molecular cell biology*, 9(7):557–568, 2008. ISSN 1471-0072. doi: 10.1038/nrm2428.

T. Sauka-Spengler, D. Meulemans, M. Jones, and M. Bronner-Fraser. Ancient Evolutionary Origin of the Neural Crest Gene Regulatory Network. *Developmental Cell*, 13(3):405–420, 2007. ISSN 15345807. doi: 10.1016/j.devcel.2007.08.005.

M. Sauvageau, L. a. Goff, S. Lodato, B. Bonev, A. F. Groff, C. Gerhardinger, D. B. Sanchez-Gomez, E. Hacisuleyman, E. Li, M. Spence, S. C. Liapis, W. Mallard, M. Morse, M. R. Swerdel, M. F. D'Ecclessis, J. C. Moore, V. Lai, G. Gong, G. D. Yancopoulos, D. Frendewey, M. Kellis, R. P. Hart, D. M. Valenzuela, P. Arlotta, and J. L. Rinn. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife*, 2:e01749, 1 2013. ISSN 2050-084X. doi: 10.7554/eLife.01749.

P. J. Schatz. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Bio/technology (Nature Publishing Company)*, 11(10):1138–1143, 1993. ISSN 0733-222X. doi: 10.1038/nbt1093-1138.

K. Schaukowitch, J.-Y. Joo, X. Liu, J. K. Watts, C. Martinez, and T.-K. Kim. Enhancer RNA Facilitates NELF Release from Immediate Early Genes. *Molecular Cell*, 56(1): 29–42, 2014. ISSN 10972765. doi: 10.1016/j.molcel.2014.08.023.

A. F. Schier, S. C. Neuhauss, M. Harvey, J. Malicki, L. Solnica-Krezel, D. Y. Stainier, F. Zwartkruis, S. Abdelilah, D. L. Stemple, Z. Rangini, H. Yang, and W. Driever. Mutations affecting the development of the embryonic zebrafish brain. *Development (Cambridge, England)*, 123:165–178, 1996. ISSN 0950-1991.

A. T. Schiffmacher, V. Xie, and L. A. Taneyhill. Cadherin-6B proteolysis promotes the neural crest cell epithelial-to-mesenchymal transition through transcriptional regulation. *Journal of Cell Biology*, 215(5):735–747, 2016. ISSN 15408140. doi: 10.1083/jcb.201604006.

S. Schmitt, M. Prestel, and R. Paro. Intergenic transcription through a Polycomb group response element counteracts silencing. *Genes and Development*, 19(6):697–708, 2005. ISSN 08909369. doi: 10.1101/gad.326205.

M. P. Schnetz, L. Handoko, B. Akhtar-Zaidi, C. F. Bartels, C. Filipe Pereira, A. G. Fisher, D. J. Adams, P. Flicek, G. E. Crawford, T. LaFramboise, P. Tesar, C. L. Wei, and P. C. Scacheri. CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genetics*, 6(7):1–15, 2010. ISSN 15537390. doi: 10.1371/journal.pgen.1001023.

M. Schorpp, R. Jäger, K. Schellander, J. Schenkel, E. F. Wagner, H. Weiher, and P. Angel. The human ubiquitin C promoter directs high ubiquitous expression of transgenes in mice. *Nucleic Acids Research*, 24(9):1787–1788, 1996. ISSN 03051048. doi: 10.1093/nar/24.9.1787.

B. S. Scruggs, D. A. Gilchrist, S. Nechaev, G. W. Muse, A. Burkholder, D. C. Fargo, and K. Adelman. Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular Cell*, 58(6):1101–1112, 2015. ISSN 10972765. doi: 10.1016/j.molcel.2015.04.006.

Y. Shen, X. Guo, and W. Wang. Identification and characterization of circular RNAs in zebrafish. *FEBS Letters*, 591:213–220, 2016. ISSN 00145793. doi: 10.1002/1873-3468.12500.

A. Sigova, B. Abraham, X. Ji, B. Molinie, N. Hannett, Y. E. Guo, M. Jangi, C. Giallourakis, P. Sharp, and R. A. Young. Transcription factor trapping by RNA in gene regulatory elements. *Science*, 350(6263):978–982, 2015.

M. Simoes-Costa and M. E. Bronner. Reprogramming of avian neural crest axial identity and cell fate. *Science (New York, N.Y.)*, 352(6293):1570–3, 2016. ISSN 1095-9203. doi: 10.1126/science.aaf2729.

M. D. Simon, C. I. Wang, P. V. Kharchenko, J. A. West, B. A. Chapman, A. A. Alekseyenko, M. L. Borowsky, M. I. Kuroda, and R. E. Kingston. The genomic binding sites of a noncoding RNA. *Proceedings of the National Academy of Sciences*, 108(51): 20497–20502, 12 2011. ISSN 1091-6490. doi: 10.1073/pnas.1113536108.

M. Simões-Costa and M. E. Bronner. Establishing neural crest identity: a gene regulatory recipe. *Development (Cambridge, England)*, 142(2):242–57, 2015. ISSN 1477-9129. doi: 10.1242/dev.105445.

M. Simões-Costa, J. Tan-Cabugao, I. Antoshechkin, T. Sauka-Spengler, and M. E. Bronner. Transcriptome analysis reveals novel players in the cranial neural crest gene regulatory network. *Genome research*, 1 2014. ISSN 1549-5469. doi: 10.1101/gr.161182.113.

M. S. Simões-Costa, S. J. McKeown, J. Tan-Cabugao, T. Sauka-Spengler, and M. E. Bronner. Dynamic and Differential Regulation of Stem Cell Factor FoxD3 in

the Neural Crest Is Encrypted in the Genome. *PLoS Genetics*, 8(12), 2012. ISSN 15537390. doi: 10.1371/journal.pgen.1003142.

S. P. Sripathy, J. Stevens, and D. C. Schultz. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Molecular and cellular biology*, 26(22):8623–38, 2006. ISSN 0270-7306. doi: 10.1128/MCB.00487-06.

S. Starke, I. Jost, O. Rossbach, T. Schneider, S. Schreiner, L.-H. Hung, and A. Bindereif. Exon Circularization Requires Canonical Splice Signals. *Cell Reports*, 10(1):103–111, 2015. ISSN 2211-1247. doi: 10.1016/j.celrep.2014.12.002.

F. A. Steiner, P. B. Talbert, S. Kasinathan, R. B. Deal, and S. Henikoff. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Research*, 22(4):766–777, 2012. ISSN 10889051. doi: 10.1101/gr.131748.111.

J. A. Steitz. Polypeptide Chain Initiation: Nucleotide sequences of the three ribosomal binding sites of bacteriophage R17 RNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 34:621–630, 1969. ISSN 00917451. doi: 10.1101/SQB.1969.034.01.072.

D. L. Stemple, L. Solnica-Krezel, F. Zwartkruis, S. C. Neuhauss, A. F. Schier, J. Malicki, D. Y. Stainier, S. Abdelilah, Z. Rangini, E. Mountcastle-Shah, and W. Driever. Mutations affecting development of the notochord in zebrafish. *Development (Cambridge, England)*, 123:263–273, 1996. ISSN 0950-1991. doi: 10.5167/uzh-229.

R. A. Stewart, B. L. Arduini, S. Berghmans, R. E. George, J. P. Kanki, P. D. Henion, and A. T. Look. Zebrafish foxd3 is selectively required for neural crest specification, migration and survival. *Developmental biology*, 292(1):174–88, 4 2006. ISSN 0012-1606. doi: 10.1016/j.ydbio.2005.12.035.

C. C. Stolt, S. Rehberg, M. Ader, P. Lommes, D. Riethmacher, M. Schachner, U. Bartsch, and M. Wegner. Terminal differentiation of myelin-forming oligodendrocytes depends on the transcription factor Sox10 service Terminal differentiation of myelin-forming oligodendrocytes depends on the transcription factor Sox10. pages 165–170, 2002. ISSN 08909369. doi: 10.1101/gad.215802.

L. R. Strachan and M. L. Condic. Neural crest motility and integrin regulation are distinct in cranial and trunk populations. *Developmental Biology*, 259(2):288–302, 2003. ISSN 00121606. doi: 10.1016/S0012-1606(03)00187-8.

V. Sundaram, Y. Cheng, Z. Ma, V. Sundaram, Y. Cheng, Z. Ma, D. Li, X. Xing, P. Edge, M. P. Snyder, and T. Wang. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research*, 24:1963–1976, 2014. ISSN 15495469. doi: 10.1101/gr.168872.113.

S. Swaminathan, F. Kiendl, R. Körner, R. Lupetti, L. Hengst, and F. Melchior. RanGAP1*SUMO1 is phosphorylated at the onset of mitosis and remains associated with RanBP2 upon NPC disassembly. *Journal of Cell Biology*, 164(7):965–971, 2004. ISSN 00219525. doi: 10.1083/jcb.200309126.

L. Szabo, R. Morey, N. J. Palpant, P. L. Wang, N. Afari, C. Jiang, M. M. Parast, C. E. Murry, L. C. Laurent, and J. Salzman. Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome biology*, 16:126, 2015. ISSN 1474-760X. doi: 10.1186/s13059-015-0690-5.

L. A. Taneyhill. To adhere or not to adhere: the role of Cadherins in neural crest development. *Cell adhesion & migration*, 2(4):223–230, 2008. ISSN 19336926. doi: 10.1038/nrm2825.

L. A. Taneyhill, E. G. Coles, and M. Bronner-Fraser. Snail2 directly represses cadherin6B during epithelial-to-mesenchymal transitions of the neural crest. *Development (Cambridge, England)*, 134(8):1481–1490, 2007. ISSN 0950-1991. doi: 10.1242/dev.02834.

P. I. Thakore, A. M, D'Ippolito, L. Song, A. Safi, N. K. Shivakumar, A. M. Kabadi, T. E. Reddy, G. E. Crawford, and C. A. Gersbach. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods*, 12(12):1143–1149, 2015. ISSN 1548-7091. doi: 10.1038/nmeth.3630.

The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447 (7146):799–816, 2007. ISSN 1476-4687. doi: 10.1038/nature05874.

E. Theveneau and R. Mayor. Neural crest delamination and migration: From epithelium-to-mesenchyme transition to collective cell migration. *Developmental Biology*, 366(1):34–54, 2012. ISSN 00121606. doi: 10.1016/j.ydbio.2011.12.041.

N. A. Thomas, M. Koudijs, F. J. M. van Eeden, A. L. Joyner, and D. Yelon. Hedgehog signaling plays a cell-autonomous role in maximizing cardiac developmental potential. *Development (Cambridge, England)*, 135(22):3789–99, 2008. ISSN 0950-1991. doi: 10.1242/dev.024083.

D. Tollervey and T. Kiss. Function and synthesis of small nucleolar RNAs. *Current Opinion in Cell Biology*, (9):337–342, 1997.

L. A. Trinh, V. Chong-Morrison, D. Gavriouchkina, T. Hochgreb-Hägele, U. Senanayake, S. E. Fraser, and T. Sauka-Spengler. Biotagging of Specific Cell Populations in Zebrafish Reveals Gene Regulatory Logic Encoded in the Nuclear Transcriptome. *Cell Reports*, 19(2):425–440, 2017. ISSN 22111247. doi: 10.1016/j.celrep.2017.03.045.

R. C. Tryon, N. Pisat, S. L. Johnson, and J. D. Dougherty. Development of translating ribosome affinity purification for zebrafish. *Genesis (New York, N.Y. : 2000)*, 51(3): 187–92, 3 2013. ISSN 1526-968X. doi: 10.1002/dvg.22363.

M.-C. Tsai, O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, Y. Shi, E. Segal, and H. Y. Chang. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science*, 329(5992):689–693, 2010. doi: 10.1126/science.1192002.Long.

I. Ulitsky, A. Shkumatava, C. H. Jan, H. Sive, and D. P. Bartel. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*, 147(7):1537–1550, 2011. ISSN 00928674. doi: 10.1016/j.cell.2011.11.055. URL `http://dx.doi.org/10.1016/j.cell.2012.10.002`.

I. Ulitsky, A. Shkumatava, C. H. Jan, H. Sive, and D. P. Bartel. Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell*, 151(3):684–686, 12 2012. ISSN 1097-4172. doi: 10.1016/j.cell.2011.11.055.

S. van Heesch, M. van Iterson, J. Jacobi, S. Boymans, P. B. Essers, E. de Bruijn, W. Hao, A. W. Macinnes, E. Cuppen, and M. Simonis. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome biology*, 15(1):R6, 2014. ISSN 1465-6914. doi: 10.1186/gb-2014-15-1-r6.

L. Väremo, J. Nielsen, and I. Nookaew. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*, 41(8):4378–4391, 2013. ISSN 03051048. doi: 10.1093/nar/gkt111.

H. Wada, H. Tanaka, S. Nakayama, M. Iwasaki, and H. Okamoto. Frizzled3a and Celsr2 function in the neuroepithelium to regulate migration of facial motor neurons in the developing zebrafish hindbrain. *Development (Cambridge, England)*, 133 (23):4749–59, 2006. ISSN 0950-1991. doi: 10.1242/dev.02665.

C. Wang, B. Gong, P. R. Bushel, J. Thierry-Mieg, D. Thierry-Mieg, J. Xu, H. Fang, H. Hong, J. Shen, Z. Su, J. Meehan, X. Li, L. Yang, H. Li, P. P. Łabaj, D. P. Kreil, D. Megherbi, S. Gaj, F. Caiment, J. van Delft, J. Kleinjans, A. Scherer, V. Devanarayan, J. Wang, Y. Yang, H.-R. Qian, L. J. Lancashire, M. Bessarabova, Y. Nikolsky, C. Furlanello, M. Chierici, D. Albanese, G. Jurman, S. Riccadonna, M. Filosi, R. Visintainer, K. K. Zhang, J. Li, J.-H. Hsieh, D. L. Svoboda, J. C. Fuscoe, Y. Deng, L. Shi, R. S. Paules, S. S. Auerbach, and W. Tong. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nature biotechnology*, 32(9):926–32, 2014. ISSN 1546-1696. doi: 10.1038/nbt.3001.

D. Wang, I. Garcia-Bassets, C. Benner, W. Li, X. Su, Y. Zhou, J. Qiu, W. Liu, M. U. Kaikkonen, K. a. Ohgi, C. K. Glass, M. G. Rosenfeld, and X.-D. Fu. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, 474(7351):390–394, 2011a. ISSN 0028-0836. doi: 10.1038/nature10006.

J. Wang, X. Li, L. Wang, J. Li, Y. Zhao, G. Bou, Y. Li, G. Jiao, X. Shen, R. Wei, S. Liu, B. Xie, L. Lei, W. Li, Q. Zhou, and Z. Liu. A novel long intergenic noncoding RNA indispensable for the cleavage of mouse two-cell embryos. *EMBO Rep*, 17(10):1452–1470, 2016. ISSN 1469-3178 (Electronic)\r1469-221X (Linking). doi: 10.15252/embr.201642051.

K. C. Wang, Y. W. Yang, B. Liu, A. Sanyal, R. Corces-Zimmerman, Y. Chen, B. R. Lajoie, A. Protacio, R. a. Flynn, R. a. Gupta, J. Wysocka, M. Lei, J. Dekker, J. a. Helms, and H. Y. Chang. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, 472(7341):120–4, 4 2011b. ISSN 1476-4687. doi: 10.1038/nature09819.

L. Wang, X. Ma, X. Xu, and Y. Zhang. Systematic identification and characterization of cardiac long intergenic noncoding RNAs in zebrafish. *Scientific Reports*, 7(1): 1250, 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-00823-3.

W. D. Wang, D. B. Melville, M. Montero-Balaguer, A. K. Hatzopoulos, and E. W. Knapik. Tfap2a and Foxd3 regulate early steps in the development of the neural crest progenitor population. *Developmental Biology*, 360(1):173–185, 2011c. ISSN 00121606. doi: 10.1016/j.ydbio.2011.09.019.

Y. Wang and Z. Wang. Efficient backsplicing produces translatable circular mRNAs. *RNA (New York, N.Y.)*, 21(2):172–9, 2015. ISSN 1469-9001. doi: 10.1261/rna.048272.114.

Z. Wang, M. Gerstein, and M. Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1):57–63, 2009.

S. Washietl, I. L. Hofacker, M. Lukasser, A. Hüttenhofer, and P. F. Stadler. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nature biotechnology*, 23(11):1383–1390, 11 2005a. ISSN 1087-0156. doi: 10.1038/nbt1144.

S. Washietl, I. L. Hofacker, and P. F. Stadler. Fast and reliable prediction of noncoding RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7):2454–2459, 2005b.

A. E. Webb, J. Sanderford, D. Frank, W. S. Talbot, W. Driever, and D. Kimelman. Laminin $\alpha 5$ is essential for the formation of the zebrafish fins. *Developmental Biology*, 311(2):369–382, 2007. ISSN 00121606. doi: 10.1016/j.ydbio.2007.08.034.

P. C. Weber, D. H. Ohlendorf, J. J. Wendoloski, and F. R. Salemme. Structural origins of high-affinity biotin binding to streptavidin. *Science (New York, N.Y.)*, 243(4887): 85–88, 1989. ISSN 0036-8075. doi: 10.1126/science.2911722.

C. F. Weil and R. Kunze. Transposition of maize Ac/Ds transposable elements in the yeast Saccharomyces cerevisiae. *Nature genetics*, 26(2):187–90, 2000. ISSN 1061-4036. doi: 10.1038/82827.

M. S. Werner and A. J. Ruthenburg. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell Reports*, pages 1–10, 2015. ISSN 22111247. doi: 10.1016/j.celrep.2015.07.033.

J. O. Westholm, P. Miura, S. Olson, S. Shenker, B. Joseph, P. Sanfilippo, S. E. Celniker, B. R. Graveley, and E. C. Lai. Genome-wide Analysis of Drosophila Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation. *Cell Reports*, 9(5):1966–1981, 2014. ISSN 22111247. doi: 10.1016/j.celrep.2014.10.062.

T. T. Whitfield, G. M., F. J. M. van Eeden, U. Schach, M. Brand, M. Furutani-Seiki, P. Haffter, M. Hammerschmidt, C. P. Heisenberg, Y.-J. Jiang, D. A. Kane, R. N. Kelsh, M. C. Mullins, J. Odenthal, and C. Nusslein-Volhard. Mutations affecting development of the zebrafish inner ear and lateral line. *Dev*, 123:240–254, 1995.

R. M. Williams, U. Senanayake, M. Artibani, G. O. Taylor, D. Wells, A. A. Ahmed, and T. Sauka-Spengler. Genome and epigenome engineering CRISPR toolkit for probing in vivo cis-regulatory interactions in the chicken embryo. *bioRxiv*, 2017. doi: 10.1101/135525.

H. Wu, A. S. Nord, J. A. Akiyama, M. Shoukry, V. Afzal, E. M. Rubin, L. A. Pennacchio, and A. Visel. Tissue-Specific RNA Expression Marks Distant-Acting Developmental Enhancers. *PLoS Genetics*, 10(9):e1004610, 9 2014. ISSN 1553-7404. doi: 10.1371/journal.pgen.1004610.

M. Xie, C. Hong, B. Zhang, R. F. Lowdon, X. Xing, D. Li, X. Zhou, H. J. Lee, C. L. Maire, K. L. Ligon, P. Gascard, M. Sigaroudinia, T. D. Tlsty, T. Kadlecek, A. Weiss, H. O'Geen, P. J. Farnham, P. A. F. Madden, A. J. Mungall, A. Tam, B. Kamoh, S. Cho, R. Moore, M. Hirst, M. A. Marra, J. F. Costello, and T. Wang. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nature genetics*, 45(7):836–41, 2013. ISSN 1546-1718. doi: 10.1038/ng.2649.

H. Xu, S. Guo, W. Li, and P. Yu. The circular RNA Cdr1as, via miR-7 and its targets, regulates insulin transcription and secretion in islet cells. *Scientific Reports*, 5 (October 2014):12453, 2015. ISSN 2045-2322. doi: 10.1038/srep12453.

Z. Xue, S. Hennelly, B. Doyle, A. A. Gulati, I. V. Novikova, K. Y. Sanbonmatsu, and L. A. Boyer. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Molecular Cell*, 64(1): 37–50, 2016. ISSN 10974164. doi: 10.1016/j.molcel.2016.08.010.

T. Ye, A. R. Krebs, M. A. Choukrallah, C. Keime, F. Plewniak, I. Davidson, and L. Tora. seqMINER: An integrated ChIP-seq data interpretation platform. *Nucleic Acids Research*, 39(6):1–10, 2011. ISSN 03051048. doi: 10.1093/nar/gkq1287.

L. Yin, L. A. Maddison, M. Li, N. Kara, M. C. Lafave, G. K. Varshney, S. M. Burgess, J. G. Patton, and W. Chen. Multiplex conditional mutagenesis using transgenic expression of Cas9 and sgRNAs. *Genetics*, 200(2):431–441, 2015. ISSN 19432631. doi: 10.1534/genetics.115.176917.

X. You, I. Vlatkovic, A. Babic, T. Will, I. Epstein, G. Tushev, G. Akbalik, M. Wang, C. Glock, C. Quedenau, X. Wang, J. Hou, H. Liu, W. Sun, S. Sambandan, T. Chen, E. M. Schuman, and W. Chen. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nature Neuroscience*, 18(4): 603–610, 2015. ISSN 1097-6256. doi: 10.1038/nn.3975.

R. S. Young, Y. Kumar, W. A. Bickmore, and M. S. Taylor. Bidirectional transcription marks accessible chromatin and is not specific to enhancers. *bioRxiv*, page 048629, 2016. doi: 10.1101/048629.

J. G. Zalatan, M. E. Lee, R. Almeida, L. A. Gilbert, E. H. Whitehead, M. La Russa, J. C. Tsai, J. S. Weissman, J. E. Dueber, L. S. Qi, and W. A. Lim. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, 160(1-2): 339–350, 2015. ISSN 10974172. doi: 10.1016/j.cell.2014.11.052.

B. Zhang, G. Arun, Y. S. Mao, Z. Lazar, G. Hung, G. Bhattacharjee, X. Xiao, C. J. Booth, J. Wu, C. Zhang, and D. L. Spector. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell reports*, 2(1):111–23, 7 2012. ISSN 2211-1247. doi: 10.1016/j.celrep.2012.06.003.

K. Zhang, K. Huang, Y. Luo, and S. Li. Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. *BMC genomics*, 15(1):845, 2014a. ISSN 1471-2164. doi: 10.1186/1471-2164-15-845.

X.-O. O. Zhang, H.-B. B. Wang, Y. Zhang, X. Lu, L.-L. L. Chen, and L. Yang. Complementary Sequence-Mediated Exon Circularization. *Cell*, 159(1):134–147, 2014b. ISSN 00928674. doi: 10.1016/j.cell.2014.09.001.

Y. Zhang, W. Xue, X. Li, J. Zhang, S. Chen, J. L. Zhang, L. Yang, and L. L. Chen. The Biogenesis of Nascent Circular RNAs. *Cell Reports*, 15(3):611–624, 2016. ISSN 22111247. doi: 10.1016/j.celrep.2016.03.058.

S. Zhao, W. P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, 9(1), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0078644.

T. Zhao, J. Xu, L. Liu, J. Bai, L. Wang, Y. Xiao, X. Li, and L. Zhang. Computational identification of epigenetically regulated lncRNAs and their associated genes based on integrating genomic data. *FEBS Letters*, 589(4):521–531, 2015. ISSN 00145793. doi: 10.1016/j.febslet.2015.01.013.

P. Zhou, F. Gu, L. Zhang, B. N. Akerberg, Q. Ma, K. Li, A. He, Z. Lin, S. M. Stevens, B. Zhou, and W. T. Pu. Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *eLife*, 6:1–29, 2017. ISSN 2050084X. doi: 10.7554/eLife.22039.

S. Zhu, W. Li, J. Liu, C.-H. Chen, Q. Liao, P. Xu, H. Xu, T. Xiao, Z. Cao, J. Peng, P. Yuan, M. Brown, X. S. Liu, and W. Wei. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nature Biotechnology*, 34(12):1279–1286, 2016. ISSN 1087-0156. doi: 10.1038/nbt.3715.

Y. Zhu, L. Sun, Z. Chen, J. W. Whitaker, T. Wang, and W. Wang. Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Research*, 41(22):10032–10043, 2013. ISSN 03051048. doi: 10.1093/nar/gkt826.