

Transcriptional regulation in early development:
identification of downstream targets through time series
perturbation in *Xenopus tropicalis*

Brook Rosie Donaldson Cooper

UCL

This thesis is submitted for the degree of
Doctor of Philosophy

I, Brook Rosie Donaldson Cooper, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

A handwritten signature in cursive script, appearing to read "Brook Cooper", is written above a horizontal line.

Brook Cooper

For my wonderful parents

Acknowledgements

Thanks to my friends and family for being so amazing and making these four years not just bearable, but really great. There are too many people to list but you know who you are and I love you all. Special thanks and huge love to David Allen for supporting me all the way through, and being there when things got really tough!

Thanks to my friends at NIMR Lesley Gray, Sarah Caswell, Hannah Stanforth and Sissy Wamaitha for all the cups of coffee, moral support and drinks at the bar.

Thanks to Mike Gilchrist for supervision and guidance. Thanks to everyone in the Gilchrist and Elgar labs for being great company and helping me throughout my PhD. Special thanks to Nick Owens for tireless analysis and explanations! Also special thanks to Elena De Domenico and Rosa Faria for making frog days fun!

Thanks for scientific advice and from members of the Smith lab, especially Kevin Dingwell, George Gentsch, Anna Strobl and Alex Watson.

Thanks to Abdul Sesay and Leena Bhaw-Rosun for your tireless running of the sequencing facility and for being great fun!

Abstract

This thesis aims to improve our understanding of zygotic gene regulation during early vertebrate development using *Xenopus tropicalis* as a model system. The aim of the first part of the work was to differentiate between maternal polyadenylation and zygotic transcription as mechanisms of gene activation in the early embryo. This work used non-polyA+ selective gene expression analysis to systematically discriminate *de novo* zygotic transcription from polyadenylation of maternal transcripts. It concludes that immediate post-fertilisation transcripts are activated by polyadenylation and shortly before mid-blastula transition (MBT) transcription becomes the dominant mechanism of activation. The motivation then was to explore the gene regulatory network downstream of early activated transcription factors. A Morpholino screen of early, transiently activated transcription factors was carried out and Mix1 was selected for its early, penetrant phenotype. The aims of the following work were to identify the targets of Mix1, and to evaluate the use of knockdown time-series RNA-seq as a means of determining transcription factor targets. To identify downstream targets of Mix1, the transcriptomes of knockdown and control embryos were compared over a time course. This work concludes that the time-series RNA-seq approach can be used to identify candidate Mix1 targets, but that these may include morpholino-specific off-target effects. The final aim was to control for off-target morpholino effects and to validate candidate Mix1 targets. To do so, two additional knockdown time-series were generated using two different translation-blocking morpholinos targeting Mix1 and Mixer. In addition, *mix1*-expressing animal cap explants were transcriptionally profiled to validate candidate Mix1 targets. This work concludes that nine Mix1 validated activatory targets are found, most of which are transcription factors which are enriched for functions in neural development indicating a novel role for Mix1, as well as functions in antero-posterior and dorsal-ventral patterning, supporting previous publications linking Mix1 to dorso-anterior development.

Contents

<u>List of Tables and Figures</u>	17
<u>List of abbreviations</u>	21
<hr/>	
<u>Chapter 1: Introduction</u>	
<u>1.1 Aims of this thesis</u>	22
<u>1.2 Regulating the polyA tail length of gene transcripts in the early embryo</u>	23
<u>1.2.1 Specific transcript deadenylation in the oocyte and post-fertilisation embryo</u>	23
<u>1.2.2 Regulation of polyadenylation in the oocyte and early embryo</u>	23
<u>1.2.3 Polyadenylation affects translation efficiency</u>	27
<u>1.2.4 Transcript degradation at MBT</u>	27
<u>1.2.5 Large scale studies into transcript polyadenylation during embryogenesis</u>	28
<u>1.3 The Mid-blastula transition and zygotic genome activation</u>	28
<u>1.3.1 The Mid-blastula transition</u>	28
<u>1.3.2 Epigenetic regulation of zygotic genome activation</u>	30
<u>1.3.3 The regulation of zygotic genome activation by transcription factors</u>	32
<u>1.4 Studying gene regulatory networks</u>	34
<u>1.4.1 Experimental methods for identifying transcription factor targets</u>	34
<u>1.4.2 Determining direct transcription factor targets</u>	35
<u>1.4.3 Computational methods of investigating transcription factor targets</u>	36
<u>1.4.4 Methods for constructing gene regulatory networks</u>	37
<u>1.5 Patterning the early <i>Xenopus</i> embryo</u>	39

<u>1.5.1 The establishment of embryonic axes</u>	39
<u>1.5.2 Formation of the Spemann organiser</u>	39
<u>1.5.3 BMPs antagonise neural development</u>	42
<u>1.5.4 Mesendoderm specification</u>	43
<u>1.6 The homeobox transcription factor Mix1</u>	44
<u>1.6.1 The activation of <i>mix1</i></u>	44
<u>1.6.2 The localisation of <i>mix1</i></u>	45
<u>1.6.3 Mix1 protein structure</u>	45
<u>1.6.4 Mix1 in endoderm development</u>	46
<u>1.6.5 Mix1 in anterior development</u>	47
<u>1.6.6 Mix1 in mesoderm development</u>	48
<u>1.6.7 Mix1 in blood development</u>	49
<u>1.7 Other Mix family genes present in <i>X. tropicalis</i></u>	50
<u>1.7.1 Mixer induction and localisation</u>	50
<u>1.7.2 Mixer in endoderm and mesoderm specification</u>	51
<u>1.7.3 Bix1.1 and Bix1.2</u>	52
<u>1.8 Mix family genes in other vertebrates</u>	52
<u>1.8.1 Mix genes in zebrafish</u>	52
<u>1.8.2 Mix1 in mouse</u>	53
<u>1.8.3 CMIX the chicken Mix1 orthologue</u>	54
<u>1.9 Morpholino oligonucleotides</u>	54

Chapter 2- Materials and Methods

<u>2.1 <i>Xenopus</i> techniques</u>	57
<u>2.1.1 <i>Xenopus tropicalis</i> embryo generation</u>	57
<u>2.1.2 <i>Xenopus Laevis</i> embryo generation</u>	57
<u>2.1.3 Morpholino Microinjection</u>	58
<u>2.1.4 <i>In vivo</i> morpholino assay</u>	58
<u>2.1.5 Time course knockdown experiment</u>	59
<u>2.1.6 RNA extraction from <i>Xenopus</i> tissue</u>	59
<u>2.1.7 Animal Cap Explants</u>	59
<u>2.1.8 Over-expression experiments</u>	60
<u>2.1.9 Whole Mount In Situ Hybridisation</u>	60
<u>2.1.10 β-Galactosidase Staining</u>	61
<u>2.2 Sequencing Techniques</u>	62
<u>2.2.1 Illumina RNA-seq library preparation</u>	62
<u>2.2.2 RiboZero and accompanying PolyA+ library preparation</u>	66
<u>2.2.3 Quality control of RNA and libraries</u>	69
<u>2.3 Molecular biology techniques</u>	69
<u>2.3.1 Synthesis of cDNA</u>	69
<u>2.3.2 Quantitative PCR</u>	70
<u>2.3.3 PCR and Agarose Gel Electrophoresis</u>	71
<u>2.3.4 RT-PCR to test activity of Mix1 splice blocking morpholino</u>	72

2.3.5 In Vitro transcription of mRNA	73
2.3.6 Molecular Cloning	73
2.3.7 Gateway Cloning	74
2.3.8 Bacterial Transformation and plasmid DNA isolation	74
2.3.9 Vector Linearisation	75
2.3.10 Antisense RNA probe synthesis	76
2.3.11 NanoString transcript quantification	76
2.4 Western blot techniques	77
2.4.1 Extraction of protein from embryos	77
2.4.2 Immunoprecipitation of proteins	77
2.4.3 Western Blot	78
2.4.4 <i>In vitro</i> protein translation	79
2.5 Data Analysis	80
2.5.1 Analysis of time-course RNA-seq data	80
2.5.2 Analysing splicing of RNA-seq reads	81
2.5.3 Detecting gene activations	81
2.5.4 NanoString data analysis	81
2.5.5 Statistics for detecting differential expression in animal cap sequencing data	81
2.5.6 Mapping <i>X. tropicalis</i> and <i>X. laevis</i> genes	82
2.5.7 Gene Ontology Analysis	82
2.5.8 Comparing differentially expressed genes from different experimental conditions	82
2.5.9 Transcription factor identification	83

2.6 Solutions and Reagents	83
2.7 Tables	86

Chapter 3: Mechanisms of gene activation

3.1 Introduction	89
3.1.1 Background and aims	89
3.1.2 Gene activation and gene expression profiles	89
3.1.3 Polyadenylation-independent transcript quantification	90
3.2 Identifying mechanisms of gene activation using NanoString technology	90
3.3 Identifying global activation using Ribo-Zero RNA-seq	95
3.3.1 Polyadenylation is the initial mechanism of gene activation and transcription is dominant shortly before MBT	95
3.3.2 Timing of gene activations are later than in the original time-series	99
3.4. Discussion	101
3.4.1 Transcription becomes the dominant mechanism of gene activation shortly before MBT	101
3.4.2 Early polyadenylation is necessary for progression through MBT	102
3.4.3 Detection of early transcription	103
3.4.4 Conclusion	105

Chapter 4: Morpholino screen of early-activated genes

4.1 Introduction	106
4.2 Screening the candidate zygotic transcription factors	108

4.2.1 Choice of genes	106
4.2.2 Phenotypes arising from the screen	108
4.2.3 Genes with strong and early defective phenotypes are interesting candidates for further study	110
4.3 Screening the candidate maternal transcription factors	110
4.3.1 Choice of genes	111
4.3.2 Phenotypes from the MO screen	112
4.4 Discussion	114
4.4.1 MOs as a knockdown tool	114
4.4.2 <i>Mix1</i> and <i>Mixer</i> are interesting candidates for further investigation	114
4.4.3 Several MOs do not generate a phenotype	115
4.4.4 The translation blocking MOs induced greater phenotypic effects than splice blocking MOs for genes which have maternal and zygotic transcripts	115
4.4.5 Comparison of phenotypes from my screen to previous knockdown studies	116
4.4.6 Comparing the zygotic gene screen to previous studies	117
4.4.7 Comparing the maternal gene screen to previous studies	119
4.4.8 Conclusion	119

Chapter 5: Investigating downstream targets of Mix1

5.1. Introduction	120
5.2 Results	120
5.2.1 Phenotypes and validation of the Mix1 splice blocking MO	120

5.2.1.1 Mix1 splice blocking MO-induced phenotypes	120
5.2.2.1 Mix1 splice blocking MO validation	121
5.2.2 Expression of <i>mix1</i> in the early embryo	123
5.2.2.1 Expression of <i>mix1</i>	123
5.2.3 Time-series analysis of Mix1 splice blocking MO-injected knockdown embryos reveals candidate Mix1 targets	124
5.2.3.1 Generating knockdown and control time-series	124
5.2.3.2 Data normalisation and differential expression calling	125
5.2.3.3 Activation of <i>mix1</i>	126
5.2.3.4 Assigning divergence times	126
5.2.3.5 Many candidate Mix1 targets were identified	129
5.2.4 The divergence timings in the splice blocking MO time-series	130
5.2.5 Gene expression perturbations are reproducible in Mix1 knockdown embryos	131
5.2.5.1 Gene expression perturbations found in the splice blocking MO knockdown time-series were replicated and measured by qPCR	131
5.2.5.2 Confirmation of differentially expressed genes in Mix1 depleted embryos	132
5.2.6 Overexpression of <i>mix1</i> disrupts embryonic development	135
5.2.7 The Mix1 splice blocking MO knockdown phenotype and downstream gene expression perturbations were not rescued	136
5.2.7.1 The Mix1 splice blocking MO knockdown phenotype was not rescued	136
5.2.7.2 The downstream gene expression perturbations in Mix1 knockdown embryos were not rescued	137

<u>5.2.8 The expression of <i>mix1</i> is perturbed in the Mix1 knockdown time-series</u>	138
<u>5.3 Discussion</u>	139
<u>5.3.1 Time-series analysis reveals candidate Mix1 targets</u>	139
<u>5.3.2 Mix1 knockdown phenotype</u>	139
<u>5.3.2.1 The Mix1 splice blocking MO phenotype</u>	139
<u>5.3.2.2 Differences to the Mix1 knockdown phenotype found in <i>X. laevis</i></u>	140
<u>5.3.3 Divergences give insight into the dynamic regulation by Mix1</u>	140
<u>5.3.4 Reproducibility of Mix1 target expression perturbations in splice blocking MO-injected embryos</u>	141
<u>5.3.5 Mix1 overexpression</u>	141
<u>5.3.6 Failure of Mix1 rescue</u>	141
<u>5.3.7 The limitations of sequencing whole embryos</u>	142
<u>5.3.8 Conclusions</u>	143

Chapter 6 – Controlling for MO off-target effects to determine Mix1 targets

<u>6.1 Introduction</u>	144
<u>6.2 Results</u>	146
<u>6.2.1 Translation blocking MO1 validation</u>	146
<u>6.2.2 Translation blocking MO2 validation</u>	147
<u>6.2.3 Phenotypes generated by the two MOs</u>	148
<u>6.2.4 Time-series analysis of gene expression in translation blocking MO1 and MO2-injected embryos</u>	150

<u>6.2.5 Comparison of the two translation blocking MO time-series</u>	153
<u>6.2.6 Comparison to the splice blocking MO time-series</u>	153
<u>6.2.7 <i>Mix1</i>-expressing animal caps were sequenced to validate <i>Mix1</i> target genes</u>	155
<u>6.2.8 Comparison of differentially expressed genes in the three knockdown time-series and <i>mix1</i>-expressing animal caps</u>	158
<u>6.2.9 Conflicting differential expression of <i>gsc</i> in the MO time-series</u>	161
<u>6.2.10 <i>Mix1</i> targets are enriched for transcription factors</u>	162
<u>6.2.11 Gene ontology analysis of <i>Mix1</i> targets reveals potential functions of <i>Mix1</i></u>	163

6.3 Discussion

<u>6.3.1 Genes are validated through perturbation in various experimental conditions</u>	164
<u>6.3.2 Limitations of using sequencing data from <i>mix1</i>-expressing animal caps to identify <i>Mix1</i> targets in this analysis</u>	165
<u>6.3.3 Combining data from three conditions controls for MO off-target effects</u>	166
<u>6.3.4 Targets not differentially expressed in all three MO time-series</u>	168
<u>6.3.5 <i>Mix1</i> targets not found in this work</u>	168
<u>6.3.6 Gene ontology analysis of <i>Mix1</i> targets</u>	169
<u>6.3.6.1 Analysis of <i>Mix1</i> targets reveals a likely function for <i>Mix1</i> in neural development</u>	169
<u>6.3.6.2 <i>Mix1</i> as a regulator of dorso-anterior development</u>	171
<u>6.3.6.3 <i>Mix1</i> as a regulator of endoderm and mesoderm development</u>	172
<u>6.3.7 The translation blocking MOs target <i>Mix1</i> and <i>Mixer</i></u>	172

<u>6.3.8 Some known Mixer targets were consistently differentially expressed in the two translation blocking MO time-series but not in the Mix1 splice blocking MO time-series</u>	174
<u>6.3.9 The Mix1 translation blocking MO phenotypes</u>	175
<u>6.3.10 Conclusions</u>	175

Chapter 7: Final discussion and future directions

<u>7.1 Chapter 3: mechanisms of gene activation</u>	176
<u>7.2 Chapter 4: Morpholino screen of early-activated genes</u>	176
<u>7.3 Chapter 5: Investigating downstream targets of Mix1</u>	176
<u>7.4 Chapter 6 – Controlling for MO off-target effects to determine Mix1 targets</u>	177
<u>7.5 Criticisms of the approaches used in this thesis</u>	178
<u>7.5.1 The morpholino screen</u>	178
<u>7.5.2 Additional experiments to complement the Mix1 splice blocking morpholino time-series</u>	178
<u>7.5.3 Additional experiments to investigate Mix1 function</u>	180
<u>7.5.4 Limitations of control morpholinos</u>	181
<u>7.6 Conclusion</u>	181
<u>8. Bibliography</u>	182

Appendix can be found in a spreadsheet on the attached CD

Tab 1 – Chapter 3 - Genes activated in RiboZero and polyA+ RNA-seq time-series

Tab 2 – Chapter 5 - Mix1 splice blocking MO time-series candidate targets

Tab 3 – Chapter 6 - Translation blocking MO1 time-series candidate targets

Tab 4 - Chapter 6 - Translation blocking MO2 time-series candidate targets

Tab 5 - Chapter 6 - *mix1*-expressing animal cap activatory candidate targets

Tab 6 – Chapter 6 – Group 1 strong candidate Mix1 targets and Group 2 validated Mix1 targets

Tab 7 – Chapter 6 – Group 1 strong candidate Mix1 targets GO enrichments for biological processes

Tab 8 – Chapter 6 – Group 2 validated Mix1 targets GO enrichments for biological processes

List of Tables and Figures

Chapter 1

<u>Figure 1.1 Maskin mediates inhibition of translation in the immature oocyte</u>	<u>25</u>
<u>Figure 1.2 The early stages of <i>Xenopus</i> development</u>	<u>29</u>
<u>Figure 1.3 Signalling centres at the blastula and gastrula stages</u>	<u>40</u>
<u>Figure 1.4 Mix1 protein domains</u>	<u>46</u>

Chapter 2

<u>Table 2.1 : Morpholino sequences</u>	<u>86</u>
<u>Table 2.2: <i>X. tropicalis</i> qPCR Primers</u>	<u>86</u>
<u>Table 2.3: <i>X. laevis</i> qPCR Primers for animal caps</u>	<u>87</u>
<u>Table 2.4: Cloning primers for Mix1 targets for use making <i>in situ</i> hybridisation probes</u>	<u>88</u>
<u>Table 2.5: Cloning primers for expression of genes of interest</u>	<u>88</u>

Chapter 3

<u>Figure 3.1. Early and late onset gene expression profiles</u>	<u>92</u>
<u>Figure 3.2. NanoString time-series data</u>	<u>94</u>
<u>Figure 3.3. All activated genes in the polyA+ and Ribo-Zero RNA-seq time-series</u>	<u>96</u>
<u>Figure 3.4. Ribo-Zero and polyA+ RNA-seq time-series data</u>	<u>98</u>
<u>Figure 3.5. Time shift of gene activations between the original and the Ribo-Zero time-series data</u>	<u>100</u>

Chapter 4

<u>Figure 4.1 Zygotic genes for MO screen</u>	107
<u>Figure 4.2. Phenotypes from the zygotic MO screen</u>	108
<u>Table 4.1. Zygotic MO screen phenotypes</u>	109
<u>Figure 4.3. Mix1 and Mixer MO phenotypes</u>	110
<u>Figure 4.4. Maternal genes for MO screen</u>	112
<u>Figure 4.5. Phenotypes from the maternal MO screen</u>	113
<u>Table 4.2. Maternal MO screen phenotypes</u>	114
<u>Table 4.3. Knockdown phenotypes in previous studies</u>	116

Chapter 5

<u>Figure 5.1. Mix1 splice blocking MO phenotype</u>	121
<u>Figure 5.2. Mix1 splice blocking MO activity</u>	122
<u>Figure 5.3. Localisation of <i>mix1</i></u>	124
<u>Fig. 5.4. Mix1 knockdown time-series collection</u>	125
<u>Fig. 5.5. The <i>mix1</i> expression profile</u>	126
<u>Figure 5.6. Candidate Mix1 target expression profiles</u>	128
<u>Figure 5.7. Total candidate Mix1 targets</u>	129
<u>Figure 5.8. Divergence times of candidate Mix1 targets</u>	130
<u>Figure 5.9. Candidate Mix1 target expression is reproducible</u>	132

Figure 5.10. Altered expression of candidate Mix1 activatory targets in response to Mix1 depletion	134
Figure 5.11. Overexpression of <i>mix1</i> disrupts development	136
Figure 5.12. Candidate Mix1 target expression was not rescued	138
Figure 5.13. Mix1 is up-regulated in knockdown embryos	139

Chapter 6

Fig. 6.1. Mixer and Mix1 share 5'UTR sequences	144
Table 6.1. Group 1 strong candidate and Group 2 validated Mix1 targets	146
Figure 6.2. Translation blocking MO1 activity	147
Figure 6.3. Translation blocking MO2 activity	148
Figure 6.4. Phenotypes induced by translation blocking MO1 and MO2	150
Figure 6.5. Time-series and animal cap collection times	151
Figure 6.6. Total differentially expressed genes in the translation blocking MO1 and MO2 time-series	152
Figure 6.7. Comparison of candidate targets from the three MO time-series	155
Figure 6.8. Mix1 induces <i>gsc</i> expression in animal cap explants	156
Figure 6.9. Total differentially expressed genes in <i>mix1</i>-expressing animal caps	157
Figure 6.10. Comparison of Group 1 strong candidate Mix1 targets and <i>mix1</i>-expressing animal cap targets	159
Figure 6.11. Group 2 validated Mix1 target expression profiles	160

Figure. 6.12. Differential expression of *gsc* reveals different regulation in the splice and translation blocking MO time-series 162

Figure 6.13. Many consistently differentially regulated genes in Group 1 and Group 2 are transcription factors 163

Figure 6.14. Annotated GO terms for biological processes enriched in the Group 1 and Group 2 Mix1 targets 164

List of abbreviations

BCNE centre	-	blastula chordin and noggin expressing centre
ChIP	-	chromatin immunoprecipitation
GFP	-	green fluorescent protein
GO	-	gene ontology
HA	-	hemagglutinin
Hpf	-	hours post fertilisation
MBT	-	mid-blastula transition
MO	-	morpholino oligonucleotide
mRNA	-	messenger RNA
NLS	-	nuclear localisation signal
ORF	-	open reading frame
PCR	-	polymerase chain reaction
qPCR	-	quantitative PCR
RNA-seq	-	RNA sequencing
RT-PCR	-	reverse transcriptase PCR
TF	-	transcription factor
UTR	-	un-translated region
WT	-	wild type

Chapter 1: Introduction

1.1 Aims of this thesis

This PhD project focuses on the early regulation of gene expression. First I investigate the regulation of maternal transcript polyadenylation and zygotic gene activation in the embryo of *Xenopus tropicalis*. The aim was to distinguish between polyadenylation and transcription as mechanisms of gene activation in order to understand when maternal transcripts are activated and when zygotic transcription becomes the dominant mechanism.

Next I investigate the early gene regulatory networks that regulate development, by investigating transcription factor targets. I carried out a morpholino screen for knockdown phenotypes in order to identify developmentally important transcription factors for further investigation. From this screen I selected Mix1 due to the early and penetrant phenotype found in knockdown embryos.

In the final chapters of the thesis I investigate the transcriptional targets of the homeobox transcription factor Mix1, which is amongst the earliest genes to be transcribed in the *X. tropicalis* embryo. I compared data from different morpholino knockdown RNA-seq time-series to control for off-target effects. In order to validate candidate Mix1 targets, *mix1*-expressing animal caps were sequenced to identify targets. The developmental functions of these targets were used to understand the likely functions of Mix1 in development.

The key aims of the thesis are:

- To distinguish between the early mechanisms of gene activation; polyadenylation and transcription.
- To identify targets of the zygotic transcription factor Mix1
- To investigate the likely functions of Mix1 through analysis of identified targets.

The computational analysis in this project was carried out by Nick Owens (See Materials and Methods for details).

The following introduction is subdivided into (i) the regulation of transcript polyadenylation in the oocyte and early embryo, (ii) the regulation of zygotic genome activation and mid-blastula transition (MBT), (iii) the methods used for investigating transcription factor targets and gene regulatory networks, (iv) examples of gene regulatory networks which regulate embryonic patterning, (v) the transcription factor Mix1 and related proteins and (vi) morpholino oligonucleotides as a tool for knocking down gene function.

1.2 Regulating the polyA tail length of gene transcripts in the early embryo

1.2.1 Specific transcript deadenylation in the oocyte and post-fertilisation embryo

The shrinkage of the polyA tail (deadenylation) depends on some non-coding regulatory elements frequently located towards the 3' of the transcript. These are AU-rich elements (AREs) with the sequence motif AUUUA and the U(G/A)-rich embryo deadenylation element (EDEN) (Ueno and Sagata 2002). Deadenylation is slower in the oocyte than in the early embryo. The degradation of deadenylated transcripts does not occur until MBT, so deadenylation does not destabilise transcripts (Voeltz and Steitz 1998).

In the oocyte and early embryo, EDEN-BP (embryo deadenylation element binding protein) binds to the EDEN and directs deadenylation of specific transcripts at the 3' end. EDEN-BP changes from a phosphorylated state in the maturing oocyte to a dephosphorylated state after fertilisation. This shift in phosphorylation state is calcium dependent and is accompanied by an increase in the rate of EDEN-dependent deadenylation (Detivaud et al. 2003). The deadenylation and translational arrest of *mos* and *eg2* mRNAs is driven by an EDEN in the 3'UTR. This effect is enhanced by an ARE located distally from the EDEN, also within the 3' UTR (Ueno and Sagata 2002). *Eg2* is among the first deadenylated mRNAs and has cis-regulatory motifs that instruct its rapid deadenylation. The deadenylation of *eg2* is sufficient for its degradation at the blastula stage and forced polyadenylation was sufficient to block transcript degradation (Audic, Omilli, and Osborne 1997). A large scale study revealed that 158 maternal mRNAs were targets of deadenylation mediated by EDEN and these genes were enriched for factors regulating oocyte maturation in *X. tropicalis* (Graindorge et al. 2008).

1.2.2 Regulation of polyadenylation in the oocyte and early embryo

During oocyte maturation, transcripts which regulate early oogenesis are deadenylated as they are no longer required. For example, mRNAs encoding cytoskeletal actin, ribosomal proteins and translation elongation factor 1a are deadenylated which reduces their

translational efficiency (Varnum and Wormington 1990). Also during meiotic maturation, mRNAs encoding factors that function in the establishment of mitosis are polyadenylated, these include c-Mos kinase and mitotic cyclins (Radford, Meijer, and de Moor 2008).

Two elements regulate polyadenylation; these are the hexanucleotide AAUAAA sequence and the cytoplasmic polyadenylation element (CPE). The CPE must be bound by CPE-binding protein (CPEB) for polyadenylation to occur (Hake, Mendez, and Richter 1998). The hexanucleotide sequence is bound by the cleavage and polyadenylation specificity factor (CPSF) and functions as part of the polyadenylation machinery (Dickson et al. 1999). The distance between these two elements affects the timing of polyadenylation. Bringing these two elements closer together causes precocious completion of transcript polyadenylation (Simon, Tassan, and Richter 1992). Pre-fertilisation polyadenylation is specific to the transcripts which contain a CPE. This protects them from the default deadenylation pathway (Legagneux, Omilli, and Osborne 1995).

Several maternal mRNAs with a role in meiosis II and mitosis such as *cyclins a, b1* and *b2*, and *c-mos* are amongst those polyadenylated during oocyte maturation. The polyadenylation of such transcripts is necessary for the progression through meiosis II and for the establishment of mitosis (Weeks, Walder, and Dagle 1991; Barkoff, Ballantyne, and Wickens 1998; Groisman et al. 2000).

Maskin is a factor which can block translation of CPE-containing mRNAs in the oocyte by associating with CPEB and the translation initiation factor eIF4E (Fig. 1.1) (Cao and Richter 2002). This blocks the association between eIF4E and eIF4G which is necessary for the formation of the translational machinery. During oocyte maturation, progesterone induces the phosphorylation of CPEB by Aurora kinase A, which promotes the association of CPSF with the hexanucleotide element. Once CPSF is bound to the mRNA, polyA polymerase (PAP) associates with the 3' end of the mRNA and activates polyadenylation. Once the polyA tail is elongated, several PABP molecules bind to it causing Maskin to dissociate from eIF4E which permits formation of the translation complex on the mRNA (Fig. 1.1) (Cao and Richter 2002).

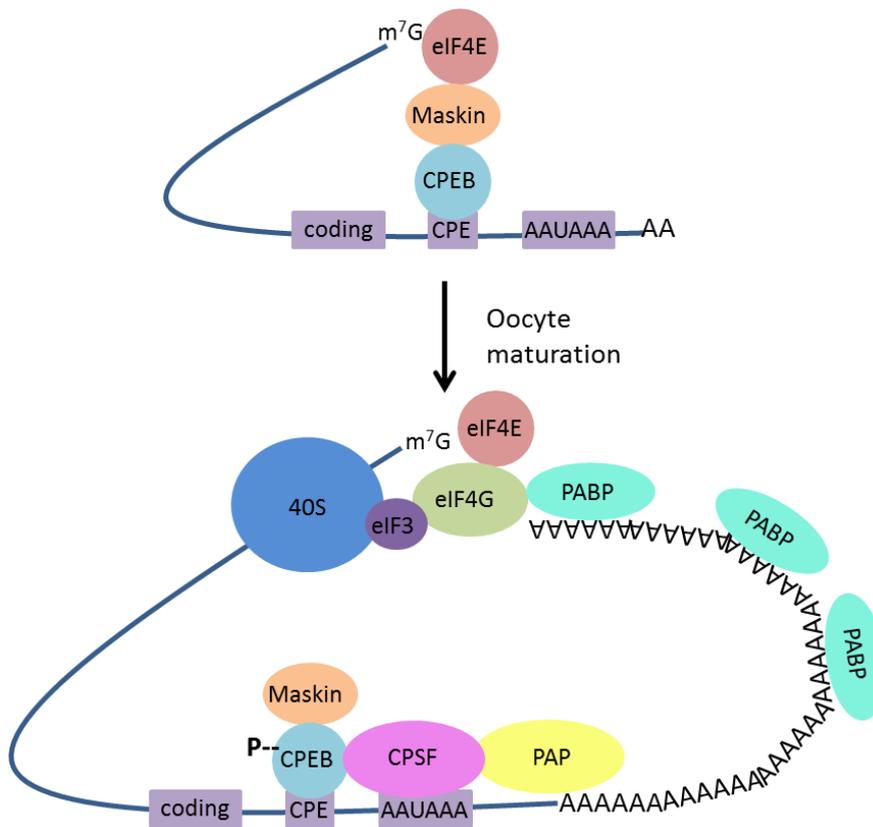


Figure 1.1. Maskin mediates inhibition of translation in the immature oocyte. Following a progesterone signal during oocyte maturation, phosphorylation of CPEB by Aurora kinase A mediates elongation of the polyA tail. PABP associates with the polyA tail and associates with eIF4G which in turn associates with eIF4E, displacing it from Maskin. The 40S ribosomal subunit is then bound by eIF3 which associates with eIF4G. Figure adapted from (Cao and Richter 2002).

In the *Xenopus* embryo, maternally-deposited mRNAs are polyadenylated from fertilisation through to the blastula stage (Sagata, Shiokawa, and Yamana 1980). A class of mRNAs which includes *cl2* and *c-raf* was identified, these transcripts are present in a deadenylated state in the oocyte and become polyadenylated after fertilisation (Paris et al. 1988). The same as in the oocyte, two cis elements in the 3'UTR of maternal mRNAs; the CPE and hexanucleotide, instruct polyadenylation in the embryo, leading to translation. The *cl2* and *c-raf* mRNAs also have a masking element which prevents their polyadenylation in the oocyte (Simon and Richter 1994).

In the mouse oocyte, the necessity for the CPE and hexanucleotide elements in the regulation of cytoplasmic polyadenylation has been demonstrated for several mRNAs including *C-mos* and *CyclinB1*. The CPEB proteins are highly homologous in *Xenopus* and *mouse* and both undergo phosphorylation during oocyte maturation, and in mouse this may also mediate polyA tail elongation (Richter 1999). In contrast to *Xenopus*, polyA tail elongation and

translation do not appear to be coupled in the mouse oocyte. Instead, an unknown mRNA binding protein associates with the CPE during oocyte maturation and this is removed to unmask sequences and allow translation to occur in the presence of a short polyA tail (Stutz et al. 1998). In the post-fertilisation mouse embryo, polyadenylation of maternal transcripts occurs, as is found in *Xenopus* embryos. This indicates that maternal factors may be important for early development in mouse, even though zygotic transcription is established at the 1-cell stage and maternal transcripts are deadenylated at the 2-cell stage (Richter 1999).

Bicoid, Toll and Torso are essential to axis formation in *Drosophila* embryos and are present as maternal transcripts which undergo polyadenylation as a means of translational activation, as is found in *Xenopus* (Salles et al. 1994). The *in vivo* function of the CPE and hexanucleotide elements have not been demonstrated in *Drosophila*. However Orb, an orthologue of CPEB, along with an elongated polyA tail, are essential for Oskar translation during oocyte maturation in *Drosophila* (Castagnetti and Ephrussi 2003). This suggests that the CPE and hexanucleotide-mediated pathway of polyadenylation that is found in *Xenopus* is also active in *Drosophila*. *Xenopus cyclinB1* transcripts were used to demonstrate the ability of *Drosophila* to regulate polyadenylation through the CPE and hexanucleotide element (Coll et al. 2010). Other mechanisms of regulation of polyadenylation exist, Toll mRNA contains a CPE hexanucleotide element which is not required for its polyadenylation. Instead, the polyadenylation of Toll was found to be regulated by a region of 183 nucleotides in the 3'UTR known as the polyadenylation region (PR) (Coll et al. 2010).

The mechanism of cytoplasmic polyadenylation in zebrafish appears to be much the same as in *Xenopus*, requiring a CPE and hexanucleotide sequence. However, the CPE sequences in *cyclinB1* mRNA have diverged, resulting in higher efficiency of polyadenylation and translation in *Xenopus* than in zebrafish, most likely due to differing requirements for levels of CyclinB1 protein (Zhang and Sheets 2009). The specific transcripts that were detected as being regulated by cytoplasmic polyadenylation in the early zebrafish and *Xenopus* embryo were compared. Of the early polyadenylated genes in zebrafish which had an orthologue in *X. tropicalis*, 20% were also polyadenylated shortly after fertilisation in *X. tropicalis* embryos indicating some level of conservation in the targets for maternal polyadenylation (Collart et al. 2014; Aanes et al. 2011).

1.2.3 Polyadenylation affects translation efficiency

In the oocyte and early embryo, polyadenylation is used to control translation of specific transcripts (Paris and Philippe 1990). The process of polyadenylation directly induces translation by stimulating methylation of the ribose cap. This methylation event is permissive for translation of maternal transcripts polyadenylated during oocyte maturation (Kuge and Richter 1995). The methylated 5' cap can then be bound by a protein complex which regulates the association of the small ribosomal subunit to the mRNA. A long polyA tail stimulates translation by the formation of a closed loop between the 5' and 3' end of the transcript attracting major components of the translation machinery (Piccioni, Zappavigna, and Verrotti 2005). The polyadenylation of the maternal mRNA *c-mos* is sufficient to trigger its translational activation, which is required for meiotic maturation.

Transcriptome-wide sequencing was used to monitor polyA tail length in pre-MBT and gastrula stage *Xenopus* and zebrafish embryos. Before MBT longer polyA tails correlate with higher translational efficiencies. At the gastrula stage, this correlation is lost and the polyA tail appears to protect against transcript degradation (Subtelny et al. 2014).

1.2.4 Transcript degradation at MBT

The introduction of the degradation pathway at MBT means that polyadenylation then stabilises transcripts and removal of the polyA tail causes rapid degradation of transcripts (Duval et al. 1990). To address whether deadenylated transcripts can persist in the embryo until degradation is established at the MBT, Audic et al. injected β -globin mRNAs comprising just the 3' UTR and ORF into *X. laevis* embryos in polyA+ and polyA- forms. After MBT, only the polyA+ transcripts were resistant to degradation and deadenylation was necessary to promote their degradation (Audic, Omilli, and Osborne 1997). The zygotically transcribed miR-427 directs deadenylation of maternal transcripts after MBT to regulate transcript clearance. This is mediated through a miR-427 binding sequence found in the 3'UTRs of mRNAs such as *cyclin a1* and *b2* (Lund et al. 2009).

Degradation of maternal transcripts is regulated by maternal factors in *Xenopus*. The post-MBT degradation pathway is blocked in embryos without de novo protein synthesis. However, blocking transcription does not inhibit the degradation of *aurka* (aurora kinase A) mRNA. Together these results indicate that the translation of maternal factors is necessary for the degradation pathway to function (Bouvet et al. 1991; Duval et al. 1990).

Possible reasons for the absence of degradation in pre-MBT embryos are: (i) A limiting factor of the degradation machinery is not translated before MBT, (ii) RNA binding proteins protect transcripts from degradation before MBT, or (iii) A component of the de-capping pathway is synthesised at MBT (Audic, Omilli, and Osborne 1997). De-capping of mRNAs is necessary for their degradation in yeast where the enzyme Dcp1 is required for de-capping activity (Beelman et al. 1996).

1.2.5 Large scale studies into transcript polyadenylation during embryogenesis

Large-scale studies have uncovered a detailed picture of the genes that are regulated post-transcriptionally in early development. Graindorge et al. used microarray technology to identify 500 genes that undergo changes in polyA tail length during oocyte maturation and early embryogenesis in *X. tropicalis* (Graindorge et al. 2006). They found that during oocyte maturation 142 mRNAs were polyadenylated and 294 were deadenylated. After fertilisation 114 mRNAs were polyadenylated and 122 were deadenylated. Interestingly, in around 75% of the transcripts polyadenylated during oocyte maturation no CPE was found, which would render these transcripts permissive to the default deadenylation pathway (Graindorge et al. 2006). RNA-seq was used to reveal maternal polyadenylation of 551 genes in the early wave of gene activation which occurs over the first two hours of development (Collart et al. 2014).

To investigate the dynamic changes in polyA tail length and transcript abundance, *X. tropicalis* embryos were profiled at several developmental stages for polyA⁺ and total RNA (Paranjpe et al. 2013). Over 9000 transcripts were found to undergo changes to their polyA tail length (Paranjpe et al. 2013).

1.3 The Mid-blastula transition and zygotic genome activation

1.3.1 The Mid-blastula transition

The mid-blastula transition (MBT) is characterised by the switch of maternal to zygotic control of development. At MBT maternal gene products are degraded and zygotic transcriptional regulation is established (Newport and Kirschner 1982a). In *Xenopus*, rapid cell cycles lacking gap phases become asynchronous and more highly regulated to include gap phases and cell cycle checkpoints. In addition, cell motility is established at this stage (Newport and Kirschner 1982b; Shimuta et al. 2002).

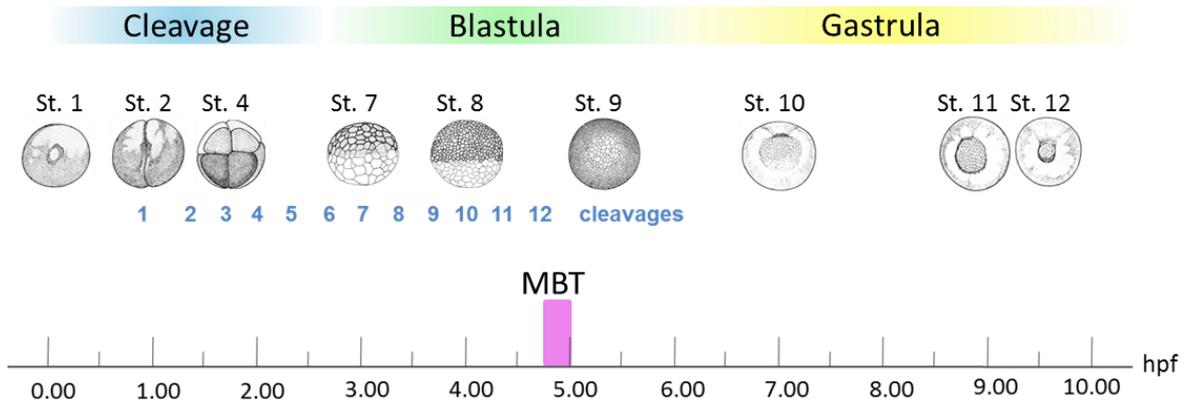


Figure 1.2. The early stages of *Xenopus* development. The mid-blastula transition (MBT) occurs after the first twelve rapid cleavages in *Xenopus*. The developmental stages in the blastula and gastrula stages occur at irregular intervals. *Xenopus* stage images reproduced from Nieuwkoop and Faber (Nieuwkoop and Faber 1994) with permission Garland Science/Taylor & Francis, LLC.

The MBT occurs independently of transcription and the number of cell cycles completed (Newport and Kirschner 1982a). The onset of MBT was proposed to be regulated by the down-titration of an unknown factor as its cytoplasmic concentration decreases in proportion to cellular DNA levels over successive cell divisions, known as the nucleocytoplasmic ratio. Transcription can be induced prematurely in *X. laevis* embryos by injecting enough RNA to recapitulate the level of DNA that is found at MBT (Newport and Kirschner 1982b). As well as transcriptional activation, the nucleocytoplasmic ratio regulates the lengthening of the cell cycles at MBT and the establishment of cellular motility (Clute and Masui 1995; Newport and Kirschner 1982b). It has been proposed that the cell cycle checkpoint is regulated by developmental timing rather than nucleocytoplasmic ratio (Clute and Masui 1995).

The induction of transcription at MBT is thought to be triggered by the depletion of a DNA-binding repressive factor (Newport and Kirschner 1982b). A gradual increase in genomic DNA that causes the levels of free histones to deplete leading up to MBT was reported in *Xenopus*. As the DNA concentration increases, the transcriptional machinery is able to outcompete histones for DNA occupancy (Hair et al. 1998). The histone proteins H3 and H4 form tetramers and have been shown to regulate the activation of transcription as well as the elongation of cell cycles which occurs at MBT (Amodeo et al. 2015). Depletion of H3 induces early transcription and increased cell cycle length, whereas addition of H3/H4 decreases cell

cycle length. The high histone to chromatin ratio in the early cell cycles creates a densely occupied chromatin which is inaccessible to transcription factors and transcriptional machinery. As the ratio decreases over successive cell cycles, histones are titrated and the chromatin becomes more accessible (Amodeo et al. 2015). However, pre-MBT gene transcription occurs which indicates that some genomic regions are accessible earlier than MBT (Collart et al. 2014). Histone modifications at specific regions may facilitate transcription factor binding before MBT.

The increasing nuclear concentration of DNA leads to down-titration of four DNA replication factors RecQ4, Treslin, Drf1, and Cut5 (Collart et al. 2013). They are implicated in regulating the increased initiation of origin firing at MBT, which controls the rate of DNA synthesis. When all four factors are overexpressed, the post MBT cell cycles do not lengthen and rapid cell divisions continue beyond the twelfth cycle leading to an increased cell number. Furthermore, the checkpoint protein Chk1 is prematurely activated by phosphorylation, extensive MBT transcription is delayed and embryos fail to gastrulate normally (Collart et al. 2013). This is a convincing explanation for the regulation of the number of rapid cell cycles occurring, but it is unclear how transcription could be controlled by these DNA replication factors.

1.3.2 Epigenetic regulation of zygotic genome activation

Epigenetic marks come in the form of chemical modifications of the DNA and histones, such as methylation of cytosine within DNA and the addition of methyl and acetyl groups to amino acids of histone proteins. Such modifications are usually found at promoter, enhancer and coding regions and act to alter chromatin packaging to control its availability for gene expression events (Bogdanovic, van Heeringen, and Veenstra 2012). The ability of a cell to specify genes as active or inactive enables the emergence of distinct cell lineages each with their own repertoire of expression events (Bogdanovic, van Heeringen, and Veenstra 2012). By altering the epigenetics within individual cells, organisms can adapt to changes by regulating gene expression at the chromatin level (Bruce et al. 2003).

In the pre-MBT embryo, genes with developmental and homeostatic roles are regulated at the level of chromatin modification. They can be given permissive H3K4me3 and/or repressive H3K27me3 or H3K9me3 histone marks, specifying different states of gene expression from the MBT and onwards (Lindeman et al. 2011). The H3K27me3 histone modification alters the chromatin to a closed and repressive state through recruitment of

PRC1 to the histone tail which in turn recruits SWI/SNF chromatin remodelling factors (Cao et al. 2002).

ChIP-seq revealed enrichment of the permissive H3K4me3 histone modification at the 5' end of genes which is associated with their transcription in *X. tropicalis*. Similarly, the H3K27me3 signal is associated at the 5' end of genes, but corresponds with restricted gene expression. ChIP-seq also revealed enrichment of RNA polymerase II (RNAPII) primarily at the transcriptional start site (TSS) and to a lesser extent along the length of the reading frame (Akkers et al. 2009). The H3K36me3 modification is commonly associated with transcriptional activity and can often be found downstream of the H3K4me3 promoter mark, across the reading frame of active genes (Mikkelsen et al. 2007).

In *X. tropicalis*, no histone modifications were found in the pre-MBT embryo where a repressive chromatin state is maintained (Akkers et al. 2009; Bogdanovic et al. 2011). H3K4me3 marks were present from the MBT on many genes and the introduction of this signal generally preceded or coincided with gene expression. The majority of H3K27me3 modifications were detected from the mid-gastrula stage and were frequently accompanied by a reduction in gene expression (Akkers et al. 2009).

In the *Xenopus* embryo, Beta-catenin is maternally deposited and is activated downstream of asymmetric Wnt signals (Schroeder et al. 1999). Beta-catenin specifies the dorsal-ventral axis by the 4th cleavage even though its first downstream targets *nodal5* and *nodal6* are not expressed until the 8th cleavage (Blythe et al. 2010). The implication is that chromatin modifications induced by Beta-catenin in the early embryo regulate later transcriptional events. Beta-catenin is necessary for the establishment of H3K4me3 histone marks at the promoter regions of its targets *sia1* and *nodal3*, establishing a poised transcriptional state (Blythe et al. 2010). Importantly, the interaction between Beta-catenin and the methyltransferase Prmt2 is necessary and sufficient for dorsal specification (Blythe et al. 2010).

Regions co-enriched for H3K4me3 and H3K27me3 modifications appeared to correlate with low levels of gene expression and RNAPII association in *X. tropicalis* (Akkers et al. 2009). ChIP-seq of embryos dissected into animal-vegetal and dorsal-ventral halves revealed that the H3K27me3 modification is associated with spatially restricted genes along the animal-vegetal or dorso-ventral axes. Different histone marks are present within different cell populations in the embryo to allow localised gene activation (Akkers et al. 2009). For example, at the *vegt*

locus, H3K27me3 was present in the animal pole segment, which is consistent with the localised expression of *vegt* in the vegetal pole (Akkers et al. 2009). This contrasts with the finding in mouse ESCs and zebrafish embryos that bivalent H3K4me3 and H3K27me3 modifications at a single locus are associated with genes poised for transcriptional activation (Lindeman et al. 2011; Bernstein et al. 2006).

The maternally encoded DNA-methyltransferase Dnmt1 regulates the timing of zygotic gene activation. A methylated DNA state is found in the sperm and oocyte but is not maintained after fertilisation and diminishes up to the MBT, reaching a state of hypomethylation (Stancheva et al. 2002). Dnmt1 is required for specific methylation of promoter regions of MBT-transcribed genes in order to prevent premature expression. Depletion of Dnmt1 was associated with loss of promoter methylation and precocious expression of *brachyury*, *gft3a*, and *myc*. The same was not true for genes expressed after the MBT, suggesting a role for Dnmt1 in temporally regulating the earliest zygotic transcription events. As the histone to DNA ratio decreases leading up to MBT, DNA methylation is reduced and transcription is rapidly de-repressed (Stancheva et al. 2002).

When Dnmt1 was depleted by morpholino knockdown, the premature induction of MBT expressed genes was observed 2 cell cycles early even though methylation at promoter sequences was not reduced compared to controls at stage 8. Rescue using mutant human mRNA demonstrates that the methyltransferase activity of Dnmt1 is not required for repression of pre-MBT transcription (Dunican et al. 2008). Dnmt1 seems to function as a transcriptional repressor by binding non-specifically to un-methylated gene promoter regions. Dnmt1 has a dual function to both directly negatively repress gene induction through promoter binding, and to methylate chromatin to generate a general state of transcriptional repression. Both functions are necessary for gene induction, genes such as *sp1* and *oct91* are sensitive to Dnmt1 repression through direct promoter binding, whereas *brachyury* is sensitive to Dnmt1-mediated methylation of its promoter (Dunican et al. 2008).

1.3.3 The regulation of zygotic genome activation by transcription factors

Different “waves” of zygotic transcription were identified in zebrafish, the first wave emerges from 4.0 hpf and is regulated by maternal factors and zygotically-regulated transcription occurs in subsequent waves. In zebrafish, three transcription factors were identified as master regulators of the initial wave of zygotic transcription, these were Nanog, SoxB1 and Pou5f1 (Oct4) (Lee et al. 2013). Depletion of all three factors using morpholinos reduced

expression of zygotic genes activated in the first wave by 77%, and expression of maternal and zygotic genes in the first wave by 50% leading to a greater effect in the reduction of subsequent transcription. The three transcription factors induce transcription of miR-430, leading to maternal transcript degradation once zygotic transcription is established (Lee et al. 2013).

In the mouse embryo Oct4, Sall4 and Nanog are responsible for the activation of widespread zygotic transcription from the 2-cell stage. Over 2000 mRNAs and 120 miRNAs were downregulated in Oct4, Sall4 or Nanog single knockdown embryos. (Tan, Au, Leong, et al. 2013). ChIP-seq data revealed that mouse ESCs have enriched binding of Sox2, Oct4 and Nanog (Lee et al. 2013). There are similarities in the transcription factors that regulate mouse and zebrafish transcriptional activation, as Pou5f1 and Oct4 are orthologous proteins, however no striking similarities to *Xenopus* have been identified.

In *Drosophila* development, the transcription factor Zelda activates early zygotic transcription and maternal transcript depletion. Zelda null embryos underwent extensive down-regulation of zygotic expression. Affected genes were enriched for the CAGGTAG TAGteam motif close to the TSS. Zelda is sufficient to induce expression through the TAGteam motif (Liang et al. 2008). No Zelda orthologue or similarly functioning factor has been identified in *Xenopus* (Skirkanich et al. 2011).

In *Xenopus*, no transcription factors have yet been identified which act as master regulators of zygotic transcription. Ventx factors are ventrally expressed and have been found to be functionally and structurally similar to mammalian Nanog. Similarly to mammalian Nanog, *Xenopus* Ventx factors interact with Pou5f1 factors such Oct91. *Ventx* overexpression represses differentiation to committed cell fates in all germ layers, indicating a function for Ventx factors in promoting pluripotency. The pluripotency factor *Oct91* was repressed when Ventx was knocked down. Morpholino knockdown of Ventx generated a dorsalised phenotype, which was rescued using mouse *Nanog* mRNA (Scerbo et al. 2012).

The timing of the onset of zygotic transcription varies between species. In zebrafish and *Xenopus* embryos, transcription is first detected at the 64-cell stage and in *Drosophila* the earliest transcription is detected at the 1024-cell stage (Wang and Davis 2014). Low level transcription is detected from the 1-cell stage in mouse embryos, followed by the onset of major transcription during the 2-cell stage (Abe et al. 2015; Wang and Davis 2014). There is a correlation between the timing of onset of zygotic transcription and the length of

embryonic cell cycles. In mammals, early cleavages are slow and allow sufficient time for transcription to proceed between cell divisions, whereas animals with faster early cleavage times, such as *Xenopus*, zebrafish and *Drosophila* cannot commence transcription until later (Wang and Davis 2014). Maternally deposited proteins and the post-transcriptional regulation of maternal transcripts occurs in *Xenopus*, zebrafish and *Drosophila* to sustain the early embryo before zygotic transcription is activated (Wang and Davis 2014).

1.4 Studying gene regulatory networks

1.4.1 Experimental methods for identifying transcription factor targets

Gene targets can be investigated using loss-of-function strategies e.g. morpholino knockdown and transgenic lines. The control and knockdown samples can be compared in order to examine which genes are differentially expressed in response to depletion of a particular transcription factor. Microarray and RNA-seq are commonly used to identify expression changes in knockdowns (Plouhinec et al. 2014; Kwon et al. 2014). A weakness of knockdown studies is that transcription factors may have redundant functions to activate the same downstream genes, which can mask the identification of transcription factor targets.

In *S. cerevisiae*, an issue with knockdown studies for identifying transcription factor targets is that some transcription factors may not be expressed under certain growth conditions. An alternative strategy was employed in yeast: generating transcription factor overexpression strains and analysing downstream gene expression by microarray to identify transcription factor targets (Chua et al. 2006). The functional classes of identified targets were used to characterise transcription factor functions (Chua et al. 2006).

In *Xenopus*, animal cap explants are commonly used as a gain-of-function assay to measure gene induction in response to injection of a particular transcription factor or signalling molecule mRNA. The animal cap cells of the *Xenopus* embryo retain pluripotency and can be treated or injected with molecular signals to manipulate the caps to express different genes and to differentiate into different tissues (Ariizumi et al. 2009). This technique combined with RNA-seq and RT-PCR has been successfully used to identify gene interactions downstream of the transcriptional regulator *Cited2* (Yoon et al. 2011).

A caveat of overexpression analysis is that when transcription factor protein levels are elevated above normal levels, a transcription factor may indiscriminately target gene expression leading to false identification of targets. To overcome this problem, comparative

analysis of transcription factors targets identified by gain-of-function and loss-of-function analysis of transgenic lines can be applied to provide stringent identification of targets (Horton et al. 2003), much like my approach of using morpholino knockdown and animal-cap sequencing experiments to complement one another.

1.4.2 Determining direct transcription factor targets

There are several techniques which enable us to detect direct TF targets. One of them, commonly used in *Xenopus* is the animal cap assay performed with a translational repressor, and more recently ChIP experiments have been used to determine direct transcription factor targets. *Xenopus* animal cap explants and whole embryos can be treated with translation inhibitor cycloheximide to demonstrate direct transcription factor-target interactions (Yokotal et al. 1995). However cycloheximide treatment has been shown to induce gene expression, therefore results from experiments using cycloheximide must be validated by other means (Hu and Hoffman 1993). Alternatively, the physical binding of a transcription factor to genomic regions can be demonstrated by electrophoretic mobility shift assay (EMSA). This technique tests binding of an isolated transcription factor to specific DNA sequences by assaying the gel size shift of protein-DNA complexes compared to unbound DNA (Garner and Revzin 1981). However, this technique is now outdated and large scale approaches are more commonly used to assay transcription factor binding.

To demonstrate potential global direct transcription factor-target interactions, the genomic binding regions of a particular transcription factor can be determined using Chromatin immunoprecipitation (ChIP) (Robertson et al. 2007). Proteins and DNA are cross-linked and the chromatin complexes are immunoprecipitated using an antibody specific to a transcription factor. Microarray (ChIP-chip) or RNA-seq (ChIP-seq) can then be used to determine genomic binding locations or 'peaks' (Robertson et al. 2007). Candidate target genes are determined by proximity to peaks, and can be validated individually by qPCR, or at a larger scale by comparative transcriptomics (Ramagopalan et al. 2010; Robertson et al. 2007). Typical ChIP experiments may yield several thousand peaks, implicating at least many hundreds of candidate target genes (Robertson et al. 2007; Sakabe et al. 2012). This approach is complicated by the discovery of functional enhancers hundreds of kb from the relevant gene promoter (Hallikas et al. 2006). For this reason, binding proximity of a transcription factor to a transcriptional start site is not a reliable way of identifying transcription factor targets and ChIP should be used in combination with other techniques.

Different ChIP-seq studies have applied various parameters for the identification of transcription factor targets based on binding proximity to the transcriptional start site (TSS). A comparative analysis of the methods used in 68 ChIP-seq studies revealed that for determination of transcription factor targets, all peaks in proximity of a gene should be taken into account and each peak should be scored based on intensity and proximity to the TSS. This means long range interactions cannot be reliably detected using this method (Sikora-Wohlfeld et al. 2013).

These approaches mentioned, when used in combination can be a particularly powerful strategy for identifying transcription factor targets. The targets of transcription factors can be identified through morpholino injection, followed by animal cap dissection, culture and RNA-seq to identify differentially expressed genes. ChIP-seq can then be used to determine transcription binding regions, and this data can then be combined with the RNA-seq data by comparing targets identified in each assay to determine direct transcription factor targets (Kwon et al. 2014). A caveat of using tagged-ChIP is that overexpressing a transcription factor may lead to mis-regulated binding to regions not usually regulated by that transcription factor.

1.4.3 Computational methods of investigating transcription factor targets

Sequence analysis can be used to predict transcription factor binding sites. Bioinformatics was used to predict genomic binding sites for E2F family transcription factors based on the proximity of binding motifs to transcriptional start sites. This method was used to identify numerous known and novel putative targets likely to have a role in cell-cycle processes. One novel target was confirmed through ChIP-PCR to confirm transcription factor binding in mouse, human and hamster (Kel et al. 2001).

TargetOrtho is a bioinformatics tool which carries out a genome-wide search for established transcription factor binding motifs in related species in order to predict transcription factor targets. Novel predicted transcription factor binding sites were validated using a reporter assay, demonstrating the effectiveness of this technique in *C. elegans* (Glenwinkel et al. 2014). Prediction of transcription factor targets through sequence analysis may be prone to errors because transcription factors may not bind to all consensus sequences and binding to sequences may be context dependent. Furthermore, there may be a high frequency of predications for degenerate motifs.

1.4.4 Methods for constructing gene regulatory networks

Combined knockdown analysis of several genes has been used to study GRNs. Embryos individually depleted of three known regulators of endoderm; Mixer, Nodal and Sox17 were analysed by microarray for expression of genes enriched in the vegetal region. Unique and shared targets were identified between Mixer, Nodal and Sox17, providing an understanding of part of the complex regulatory network governing endoderm induction and mesendoderm repression (Sinner et al. 2006).

Data can be taken from multiple gene perturbation analysis studies to map together gene interactions of numerous transcription factors and signalling molecules. A network for endomesoderm development in sea urchin embryos has been developed, which collates information on gene perturbation analysis, gene localisation and directness of gene interactions defined using computational and experimental methods. This reveals feedback loops, and localised mechanisms of activation and repression that function to specify endoderm and mesoderm (Davidson et al. 2002). Loose and Patient carried out a meta-analysis to collate previous work into visual GRNs of *Xenopus* mesendoderm development. Gene interaction networks were constructed based on gene expression changes influenced by altered upstream gene expression and proof of transcription factor binding within regulatory regions, or the demonstration of direct interaction. The GRNs display spatial and temporal organisation of signalling events as well as distinguishing between maternal and zygotic signals (Loose and Patient 2004).

The cell cycle gene regulatory network in *S. cerevisiae* was studied by identifying direct targets of nine known cell cycle transcriptional activators using ChIP-chip (Simon et al. 2001). Transcription factors were found to regulate other transcription factors functioning in successive stages of the cell cycle, resulting in a circuit of gene regulation. The transcription factors were each found to have stage-specific functions to form a highly regulated network controlling all processes involved in the cell cycle (Simon et al. 2001). As a follow up to this study, transgenic strains were individually constructed for 106 transcription factors to contain a myc epitope tag. ChIP-chip was then used to identify likely transcription factor targets. This revealed nearly 4000 gene interactions which were mapped to construct highly interconnected gene regulatory networks for different processes e.g. metabolism and developmental processes. Furthermore, different regulatory motifs within the network were identified e.g. positive feedback loops, giving an insight into the regulatory mechanisms within the network (Lee et al. 2002).

A major advantage of computational methods for building gene regulatory networks is that large datasets can be combined which use different conditions and therefore availability of co-factors to build a richer picture of gene regulation. Gene interactions identified by Lee et al. through ChIP-microarray analysis described above (Lee et al. 2002), were compared to a library of gene expression data collected from hundreds of transcriptomes (Gao, Foat, and Bussemaker 2004). Gene interactions that were correlated in both data sets were assigned a “coupling strength” based on the significance of the measured interactions. Gene ontology analysis revealed functional enrichment amongst the significant gene interactions. This revealed that approximately half of targets identified through ChIP analysis were not validated through expression data. The unconfirmed ChIP interactions may be false positives, but may also reflect complex regulatory scenarios e.g. the presence of inhibitory factors, chromatin inaccessibility and the lack of co-factors (Gao, Foat, and Bussemaker 2004). This approach of combining large-scale datasets is more sensitive for determining direct transcription factor targets than simply combining loss-of-function data with ChIP data from a single study.

In an attempt to understand the networks which regulate human disease, ENCODE ChIP-seq data was compared to functional modules derived from microarray meta-analysis to improve the identification of transcription factor targets (Karczewski et al. 2014). Functional modules are sets of genes which are co-regulated in different experimental conditions. Transcription factor targets identified by ChIP-seq analysis were compared to the functional modules. If many genes share a functional module, they are more likely to be genuine transcription factor targets. The functional modules associated with transcription factors were subjected to gene ontology (GO) analysis and many were associated with known functions of the associated transcription factor. Different functional modules were connected when regulated by the same transcription factors, revealing network interactions. These module connections based on transcription factor regulation were used to generate disease networks. The transcription factors associated with particular diseases clustered into different sub-networks. This method was used to identify thousands of significant transcription factor targets, and associations with 253 diseases (Karczewski et al. 2014).

1.5 Patterning the early *Xenopus* embryo

1.5.1 The establishment of embryonic axes

The early embryo depends on maternally deposited proteins and mRNA to direct the early specification of dorso-ventral and animal-vegetal axes. After zygotic genome activation, a complex gene regulatory network that regulates embryonic development is established (Heasman 2006).

The sperm entry point defines the position of the dorso-ventral axis, Wnt11 accumulates dorsally through differential polyadenylation and translation of *wnt11* mRNA (Schroeder et al. 1999). Wnt11 then signals through Beta-catenin to activate *tcf3* which then drives transcription of *nodal5* and *nodal6* to influence dorsal fates (Yang et al. 2002). The Spemann organiser forms at the gastrula stage on the dorsal side of the embryo and is expressed in the mesoderm and endoderm from the upper blastopore lip to the blastocoel (Sudou et al. 2012).

1.5.2 Formation of the Spemann organiser

At the blastula stage two distinct signalling centres form downstream of the maternal Beta-catenin signal; the Nieuwkoop centre and blastula Chordin and Noggin expressing centre (BCNE centre), which express genes that establish formation of the Spemann organiser at the gastrula stage (Kuroda, Wessely, and De Robertis 2004). The BCNE centre is established in the dorsal animal region of the embryo and expresses *chordin*, *noggin*, *siamois* (*sia1*) and *nodal3*. BCNE cells give rise to anterior neural plate and contribute to brain and retina formation. The Nieuwkoop centre is located in the dorsal vegetal region of the embryo and expresses nodal signalling molecules and Cerberus (Kuroda, Wessely, and De Robertis 2004).

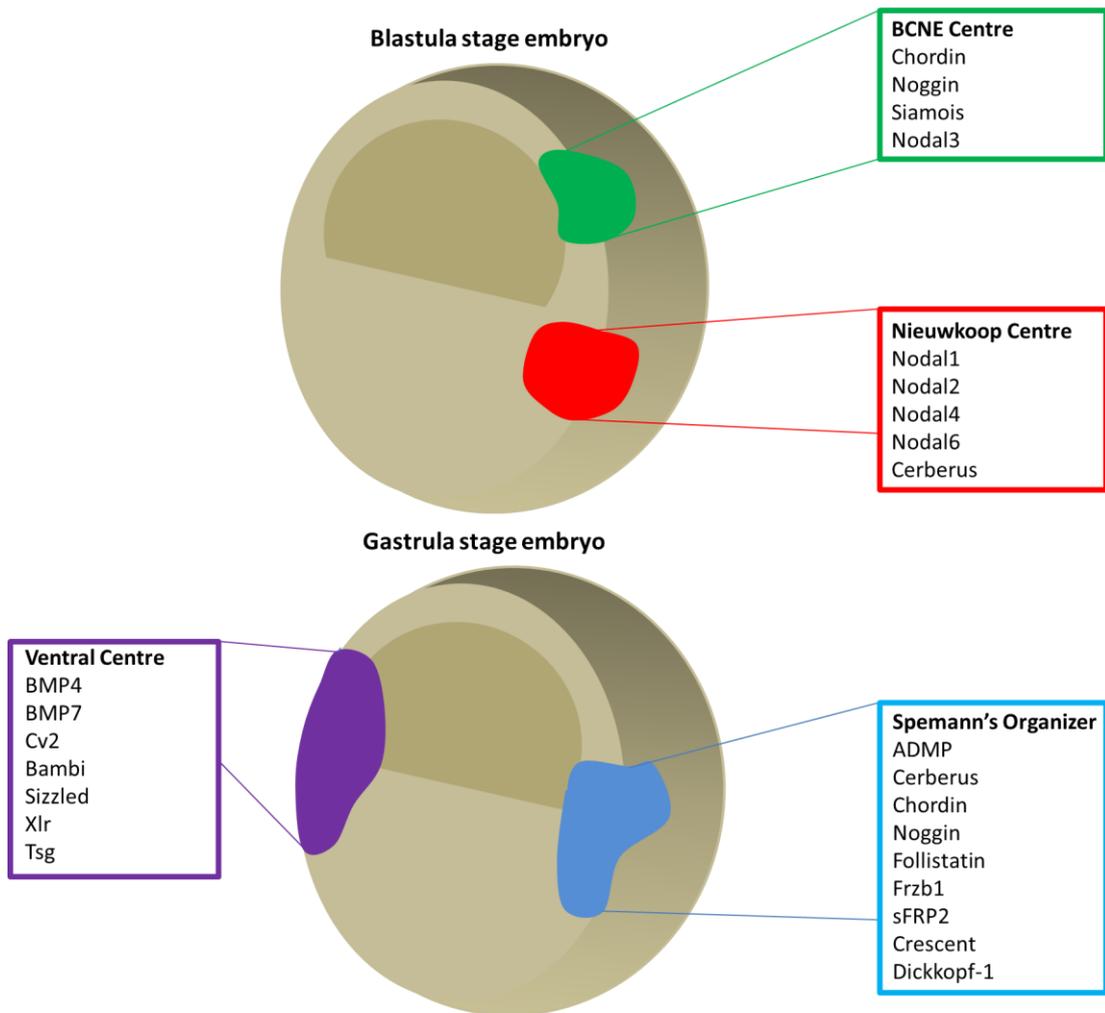


Figure 1.3. Signalling centres at the blastula and gastrula stages. At the blastula stage, the BCNE and Nieuwkoop centres form in the dorsal region and each expresses secreted factors. At the gastrula stage the Spemann organiser forms in the dorsal mesoderm and endoderm and expresses secreted BMP antagonists (Chordin, Noggin and Follistatin), secreted Wnt antagonists (Frzb1, sFRP2, Crescent and Dickkopf-1), and Cerberus which antagonises BMP, Wnt and TGF β signals. ADMP is also expressed in the Spemann organiser, despite being a BMP protein which is activated at low levels of BMP. A ventral centre expresses BMP growth factors and several BMP regulating factors (Cv2, Bambi, Sizzled, Xlr and Tsg). Figure adapted from (Kuroda, Wessely, and De Robertis 2004; De Robertis 2006).

Genes expressed in the two early signalling centres cooperatively induce anterior CNS. At the 8-cell stage Cerberus was knocked down in the dorsal vegetal cells and Chordin was knocked down in the dorsal animal cells. When either morpholino was injected alone this caused a partial loss of anterior CNS formation. When both were knocked down, anterior CNS formation was completely blocked, demonstrating a cooperative effect of genes from the

BCNE and Nieuwkoop centres in driving neural differentiation (Kuroda, Wessely, and De Robertis 2004). This cooperation between BCNE and Nieuwkoop centres demonstrates a partial overlap in function between the presumptive ectoderm and mesendoderm germ layers and also shows that neural induction starts as early as the blastula stage (Kuroda, Wessely, and De Robertis 2004).

The Nieuwkoop centre is located in the dorsal vegetal cells and is induced by the vegetal VegT signal and the dorsal Beta-catenin signal (Agius et al. 2000). Cerberus, a known antagonist of nodal signalling molecules, can block dorsal and ventral mesoderm formation. Dose response experiments using Cer-S; the carboxy-terminal region of Cerberus which specifically inhibits Nodal signalling, revealed that a gradient of Nodal activity is established from the Nieuwkoop centre. This graded expression of the Nodal related proteins from the Nieuwkoop centre drives the specification of dorsal mesoderm. Nodal signalling is sufficient to induce expression of several genes expressed in the Spemann organiser; *follistatin*, *frzb1*, *dkk1*, *cerberus*, *goosecoid (gsc)*, *chordin* and *noggin* (Agius et al. 2000).

At gastrulation, signals from the BCNE and Nieuwkoop centres specify the formation of the Spemann organiser in the dorsal mesoderm which expresses genes essential for anterior development (Lane and Sheets 2000; Kuroda, Wessely, and De Robertis 2004). The assignment of the embryonic axis changes between gastrula and neurula stages so that the embryonic dorsal pole becomes the definitive anterior pole. It is proposed that some of the genes expressed in the Spemann organiser region e.g. *gsc*, *cer1* and *sia1* are markers for anterior development (Lane and Sheets 2000). *Gsc* is expressed in the Spemann organiser and induces head formation. Ectopic expression of *gsc* in the ventral side of *X. laevis* embryos is sufficient to induce a secondary dorsal axis complete with head structures and notochord (Cho et al. 1991).

Ectopic expression of *cerberus* in the ventral vegetal region of the 32-cell stage embryo was sufficient to induce a secondary head structure as well as secondary liver and heart tissue (Bouwmeester et al. 1996). Cerberus is secreted from the endoderm and binds to and antagonises Nodal1, BMP4 and Wnt8a to allow a permissive state for head specification (Piccolo et al. 1999). The head inducing ability of Cerberus through antagonism of BMP, Wnt and Nodal signalling molecules was demonstrated. Defective head formation was observed when the Cerberus antagonistic targets BMP4, Nodal1 and Wnt8a were specifically expressed in the anterior dorsal endoderm under a Cerberus-like promoter. A Cerberus morpholino further potentiated this decrease in anterior neural development however this

morpholino was later shown to be ineffective in targeting all Cerberus alleles in *X. laevis*. Had a more effective morpholino been used, complete loss of head formation would be expected. Although the morpholino was not fully effective, a reduction in forebrain markers and an increase in *krox20*, a posterior neural marker was observed, further confirming the pro-anterior neural function of Cerberus (Silva et al. 2003).

1.5.3 BMPs antagonise neural development

BMP4 and BMP7 signals drive ectoderm maintenance through repression of neural inducing signals. The specification of ectoderm into neural tissue is thought to be the default mechanism, and BMP signals are required to allow epidermis to form on the ventral side (Hawley et al. 1995). Dominant negative BMP4 and BMP7 injected into the ventral marginal zone drive ectopic expression of neural markers and formation of a secondary axis. Expression of dominant negative BMP4 and BMP7 in animal caps up-regulate neural marker expression (Hawley et al. 1995). BMP4 overexpression leads to an increase in ventral mesoderm formation and can induce postero-ventral mesoderm in animal caps (Dale et al. 1992; Jones et al. 1992).

BMP antagonists emanate from the Spemann organiser to regulate the dorso-ventral axis, and the formation of neural tissue. Follistatin is a soluble BMP antagonist which is expressed in the Spemann organiser and notochord which are both neural inducing tissues. Follistatin induces neural formation and blocks mesoderm (Hemmati-Brivanlou, Kelly, and Melton 1994). Noggin is secreted from the Spemann organiser and antagonises BMP4, BMP2 and BMP7 by strongly binding to them and preventing them from binding to BMP receptors (Zimmerman, De Jesus-Escobar, and Harland 1996). Chordin binds with high affinity to BMP4 and inhibits its receptor-binding potential, as a result of this BMP antagonism. Chordin can dorsalise mesoderm and neuralise ectoderm when expressed in ventral marginal zone explants (Piccolo et al. 1996).

The three BMP antagonists Chordin, Follistatin and Noggin were knocked down using morpholinos to give a more severe phenotype than is produced in single or double knockdowns of these genes. This suggests that these three genes have a shared and cooperative function. In triple knockdown embryos ventral tissue formation was increased and dorsal and neural development were blocked. This was caused by a loss of BMP antagonism leading to widespread neural repression. The phenotype can be partially rescued

by injection of *noggin* mRNA or morpholinos targeting BMP4 and BMP7, where the neural plate formation is restored (Khokha et al. 2005).

Simultaneous knockdown of BMP2, BMP4 and BMP7 blocked tail and trunk development, and generated dorsalised embryos. When formation of the Spemann organiser was also blocked in these embryos using UV or knockdown of Beta-catenin, head and neural development were unaffected. This indicates that the ectoderm forms neural tissue by default, and the antagonism between BMP and the Spemann organiser signals regulates the formation of epidermis in regions where BMP signals dominate (Reversade et al. 2005).

1.5.4 Mesendoderm specification

VegT protein and mRNA are maternally deposited and are responsible for establishing endodermal signals in the animal hemisphere by first inducing the TGF β signalling molecules *nodal1*, *nodal2*, *nodal4* and *gdf3* (Xanthos et al. 2001). This leads to activation of the pro-endodermal transcription factors *mixer*, *sox17*, *gata4*, *gata5* and *gata6* (Xanthos et al. 2001). VegT also induces the ventral mesodermal gene *wnt8a* and the mesodermal genes *brachyury*, *mix1* and *eomes* (Stennard, Carnac, and Gurdon 1996). Mixer antagonises mesodermal fates by repression of the mesoderm inducing growth factors *nodal1*, *nodal5*, *fgf3* and *fgf8* (Kofron, Wylie, and Heasman 2004).

Nodal5 and *nodal6* are activated cell-autonomously by maternal Beta-catenin and VegT and induce mesendoderm and the expression of *nodal1* and *nodal2* (Takahashi et al. 2000). Nodal1 and Nodal2 control the convergent extension movements of axial and paraxial mesoderm and mesoderm migration during gastrulation (Luxardi et al. 2010).

FGF signalling is essential to the patterning of the posterior and lateral mesoderm (trunk and tail mesoderm). Blocking FGF signalling through injection of a dominant negative FGF receptor disrupts lateral mesoderm differentiation (Amaya, Musci, and Kirschner 1991). Downstream of Activin signals, FGF induces mesodermal gene expression, including *brachyury* and *xpo1* in the equatorial zone of the blastula stage embryo (Amaya et al. 1993; Cornell, Musci, and Kimelman 1995). FGF is required for the maintenance but not the activation of several organiser genes such as *chordin*, *noggin* and *gsc* (Fletcher and Harland 2008). FGF has been shown to be essential for the initiation and maintenance of *brachyury* expression and the resultant formation of paraxial mesoderm (Fletcher and Harland 2008).

A complex set of signals define the dorso-ventral axis. *Wnt8a* and *gsc* are induced in the mesoderm cell autonomously, but *wnt8a* is primarily expressed in the ventral mesoderm and

endoderm, whereas *gsc* is localised within the dorsal mesoderm (Lemaire and Gurdon 1994). *Wnt8a* has a strong ventralising influence on mesoderm and inhibits dorsal mesoderm development. When ectopically expressed, *wnt8a* can block organiser induction and ventralise prospective notochord cells to form muscle, and can re-specify ectoderm cells to ventral mesoderm (Christian and Moon 1993). As mentioned, Gsc has a strong dorsalising influence and when expressed ventrally, induces a secondary dorsal axis with head and axial structures (Cho et al. 1991).

1.6 The homeobox transcription factor Mix1

Mix1 is a homeobox transcription factor found in *X. tropicalis* which is a member of the Mix family of homeobox genes. The Mix family genes have duplicated and diverged in fish and frog. In *X. tropicalis*, four Mix family genes are found; *mix1*, *mixer*, *bix1.1* and *bix1.2*, whereas six Mix family genes are found in *X. laevis* and four are found in zebrafish. In mammalian species and chicken no duplication events have occurred and just one Mix1 family gene exists in each (Pereira et al. 2012).

1.6.1 The activation of *mix1*

The homeobox transcription factor Mix1 is activated in untreated and cycloheximide-treated embryos indicating that a maternal factor initiates its expression (Yasuo and Lemaire 1999). At NF stages 9 and 10, *mix1* was induced in *vegt*-expressing animal caps. To test whether *mix1* induction is cell contact dependent, embryos were injected with *vegt* mRNA, animal caps were dissected at stage 8, dissociated and cultured to stages 9 and 10. *Mix1* was induced in stage 9 dissociated caps but not in stage 10 dissociated caps indicating that early induction of *mix1* is not dependent on cell communication (Yasuo and Lemaire 1999). *Mix1* was also expressed in animal caps expressing the TGF β family growth factors *nodal1*, *nodal2* or *gdf3* (Yasuo and Lemaire 1999). These results suggest that early maternal signals such as VegT activate *mix1*, and later *mix1* transcription is dependent on TGF beta signalling molecules such as Nodal1, Nodal2 and Gdf3 to signal within the endoderm (Yasuo and Lemaire 1999). This fits with the activation times of these genes in *X. tropicalis*, as *nodal1*, *nodal2* and *gdf3* are activated at approximately the same time as Mix1, so would make implausible initial activators, and *vegt* is present as a maternal transcript throughout the early cleavage stages (Collart et al. 2014).

Several studies have shown that TGF β family growth factors are sufficient for Mix1 induction. Animal caps expressing *bmp4* (Mead et al. 1996), *gdf1* (*vg1*) and *activin* (*inhba*) can induce *mix1* expression (Cornell, Musci, and Kimelman 1995). *Mix1* expression was down-regulated in embryos expressing a dominant negative type II activin receptor (Yasuo and Lemaire 1999). *In vitro*, activin has been shown to affect the expression of *mix1*, but *activin* is first expressed several hours after *mix1* is activated (Collart et al. 2014), therefore maternally deposited Activin protein might regulate *mix1* expression. Gdf1 is more plausible as an initial activator of *mix1* as *gdf1* transcripts are present from fertilisation until gastrulation (Collart et al. 2014). BMP4 is unlikely to be the initial activator of *mix1* as it is activated at the same time as Mix1 and is expressed in the ventral marginal zone, whereas *mix1* is expressed in the dorsal and ventral marginal zone (Knochel, Schuler-Metz, and Knochel 2000; Lemaire et al. 1998).

1.6.2 The localisation of *mix1*

The homeobox transcription factor *mix1* is first expressed in the vegetal hemisphere of the *X. laevis* embryo (Rosa 1989). During early gastrulation its expression intensifies throughout the marginal zone, this region includes prospective endoderm and mesoderm (Rosa 1989). *Mix1* transcripts are present throughout the vegetal hemisphere and are most abundant in the marginal zone at stage 10 (Lemaire et al. 1998). From stage 10.5 *mix1* becomes gradually more enriched in the ventral marginal zone and is last detected at stage 12.5 (Colas et al. 2008).

1.6.3 Mix1 protein structure

The structure of Mix1 protein is conserved over different species, consisting of an N-terminal proline-rich domain, followed by the conserved DNA-binding homeodomain and a C-terminal polar/acidic domain. The proline-rich domain does not have any conserved motifs, but may function as a helix breaking domain, a protein association domain or a transcriptional repression or activation domain. The polar/acidic region is conserved and could potentially form an amphipathic helix which could function as an activation domain (Sahr et al. 2002). A yeast two-hybrid screen was carried out using mouse Mixl1 protein to reveal that the homeodomain and the C-terminal domain, but not the N-terminal domain are necessary for reporter gene induction, indicating an essential role for the polar/acidic domain. Mouse Mixl1 binds to the motif TAATTGAATTA, but not single TAAT motifs, indicating that it homodimerises for DNA-binding to the palindromic TAAT sequences (Sahr et al. 2002).

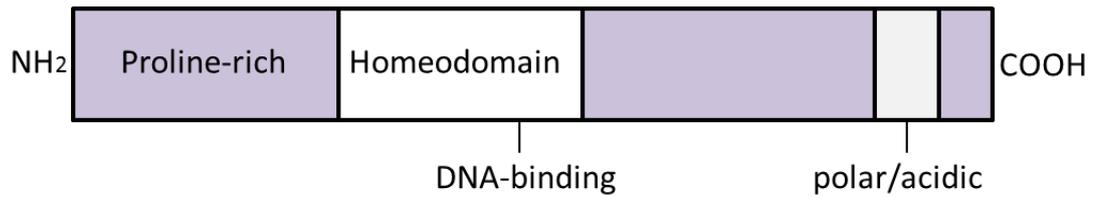


Figure 1.4. Mix1 protein domains. Mix1 protein has an N-terminal proline-rich domain, a homeodomain and C-terminal polar/acidic domain (Sahr et al. 2002).

1.6.4 Mix1 in endoderm development

Mix1 has a general role to promote endoderm formation and inhibit mesoderm differentiation. Overexpression of *mix1* in the dorsal marginal zone induced *a2m* (*endodermin*), an endoderm and dorsal mesoderm marker, but did not perturb expression of *cer1*, an anterior mesendoderm marker. Mix1 may promote endoderm formation in the dorsal region (Lemaire et al. 1998).

Vegetal expression of dominant negative enRMix1 caused a reduction in gut development indicating a requirement for Mix1 in endoderm formation (Lemaire et al. 1998). In support of this, the endoderm marker *sox17a* was repressed in whole embryos expressing enRMix1 (Latinkic and Smith 1999). Animal caps injected with *mix1* mRNA expressed very low levels of the endoderm marker *a2m* but no other endoderm markers were induced (Lemaire et al. 1998; Henry and Melton 1998). The expression of *a2m* in animal caps was synergistically increased by co-expression of *sia1* and *mix1*. Animal caps expressing *sia1* and *mix1* induced expression of the posterior and anterior endoderm markers, respectively, *fabp2* and *pdx1*. Low level induction of the anterior mesendoderm marker gene *cer1* was detected in *sia1* and *mix1*-expressing animal caps. This indicates Mix1 activates endoderm weakly alone, and works in combination with Sia1 and potentially other factors in endoderm induction (Lemaire et al. 1998). Sia1 and Mix1 are able to form stable heterodimers and this association may be key to the synergistic induction of endoderm gene expression (Mead et al. 1996).

Mix1 and Gsc are also able to heterodimerise (Wilson et al. 1993). Mix1 induces Gsc in cycloheximide-treated animal caps (Latinkic and Smith 1999). *Mix1*-expressing animal caps induce low level expression of the endoderm marker *sox17a*, this effect is greatly enhanced by coinjection of *gsc* mRNA (Latinkic and Smith 1999). *Gsc* is expressed in the Spemann organiser region and patterns dorsal mesoderm (Steinbeisser et al. 1995).

The transcription factors Mix1, Sia1 and Gsc have a synergistic effect on the rate of cell motility in the head mesoderm, part of the dorsal mesendoderm, in the direction of the

animal pole during gastrulation (Luu et al. 2008). It is possible that this is linked to the heterodimerisation of Mix1 with Sia1 or Gsc.

1.6.5 Mix1 in anterior development

Mix1 binds to gene regulatory regions in cooperation with other transcription factors to activate downstream gene expression. Mix1, Lim1 and Sia1 synergistically activate *cer1* through binding to the 3xTAAT promoter element (Yamamoto et al. 2003). Mix1 can independently and directly activate *gsc in vitro* by binding to both the proximal and distal elements of its promoter (Latinkic and Smith 1999). An *in vivo* study revealed that this region of the *gsc* promoter is bound cooperatively by Mix1, Lim1, Otx2, Sia1 and VegT at the early gastrula stage (Sudou et al. 2012). These five transcription factors also cooperatively bind to the promoter of *cer1* at this stage in the U1 region which contains a T-box recognition sequence and 5xTAAT elements. These transcription factors all have expression domains that overlap with the Spemann organiser and mediate the expression of their targets *cer1* and *gsc* in this region (Sudou et al. 2012) which in turn promote head development (Bouwmeester et al. 1996; Cho et al. 1991). The cooperative binding was in *X. laevis* but the cis-regulatory elements in the *gsc* and *cer* promoters are conserved in *X. tropicalis* (Sudou et al. 2012).

When dominant negative enRMix1 was expressed in the dorsal-vegetal region, head development was severely disrupted, leading to many embryos developing with cyclopia or no head. Embryos also displayed enlarged axial structures (notochord and somites). This effect was rescued by injection of *mix1* mRNA. *A2m* and *cer1* were down-regulated in enRMix1 expressing embryos and *chordin* was unaffected, indicating that anterior endomesoderm but not dorsal mesoderm is dependent on *mix1* expression (Lemaire et al. 1998).

Mix1 loss-of-function by injection of either dominant negative enR-Mix1 or antisense *mix1* mRNA in *X. laevis* disrupted head formation and led to heart and gut defects implicating Mix1 in regulation of dorso-anterior endoderm and mesoderm development (Latinkic and Smith 1999). Over half of embryos developed without a heart, however this apparent induction of mesoderm tissue by Mix1 may be indirect as dorso-anterior endoderm contributes to heart development (Latinkic and Smith 1999; Nascone and Mercola 1995). Mix1 may repress posterior development as overexpression of *mix1* in the ventral-equatorial region generated embryos with tail defects (Lemaire et al. 1998).

1.6.6 Mix1 in mesoderm development

Several studies have implicated Mix1 in the repression of mesoderm development. *Mix1* mRNA was injected dorsally, laterally or ventrally and repressed expression of the mesodermal activator *brachyury* within the targeted region. *Brachyury* overexpression caused localised down-regulation of *mix1* (Lemaire et al. 1998; Conlon et al. 1996). *Mix1* and *brachyury* have overlapping expression domains at stage 10, and by stage 10.5 a negative feedback loop is established and their expression domains within the marginal zone segregate. Lemaire et al. hypothesise that the exclusion of *mix1* from the *brachyury* domain in the marginal zone is required to generate a permissive state for mesoderm to form (Lemaire et al. 1998). Mix1 was shown to block *brachyury* induction in FGF expressing animal caps, however this may not be a direct effect as Mix1 is known to repress FGF expression (Latinkic et al. 1997; Colas et al. 2008).

Previous work suggests that *brachyury* may be an indirect target of Mix1 (Latinkic and Smith 1999). Two possible indirect mechanisms of *brachyury* repression by Mix1 exist; through repression of *fgf8* and *fgf4* which both activate *brachyury* (Fletcher, Baker, and Harland 2006; Isaacs, Pownall, and Slack 1994), and through activation of *gsc* which directly represses *brachyury* (Latinkic and Smith 1999).

Mix1 was fused to a viral transcription activation domain to make VP16Mix.1, a dominant positive Mix1 that would induce all Mix1 targets, both repressive and activatory. VP16Mix.1 blocked mesoderm development indicating that Mix1 functions as an activator to repress mesoderm (Lemaire et al. 1998). A luciferase reporter assay and the induction of *gsc* in animal caps further demonstrate the ability of Mix1 to activate gene expression (Latinkic and Smith 1999).

Mix1 overexpression can block mesoderm differentiation. Ventral overexpression of *mix1* down-regulated the ventral mesoderm markers, *xpo* and *ventx1.2*, whereas dorsal *mix1* overexpression down-regulated the dorsal mesoderm marker *chordin*. *Mix1* overexpression in the marginal zone can block dorsal and ventral mesoderm tissue formation. The phenotypic effects of *mix1* overexpression were severe, beginning with blastopore closure failure, leading to defective notochord and muscle formation and in some cases a complete loss of these tissues, furthermore embryos were highly truncated along the anterior-posterior axis (Lemaire et al. 1998). A dominant positive Mix1 construct was used to generate similar phenotypic effects characterised by a reduction in axial structures, leading to the

conclusion that Mix1 blocks mesoderm induction through transcriptional activation. Lemaire et al. oppose the proposition by Mead et al. that Mix1 represses dorsal mesoderm because they demonstrated that Mix1 can also repress ventral mesoderm (Lemaire et al. 1998; Mead et al. 1996).

When implanted at the early gastrula stage, *mix1*-expressing cells can integrate normally into somites, indicating that Mix1 repression of mesoderm observed by Lemaire et al. is non-cell autonomous. This indicates that once FGF signalling has been established, Mix1 may be permissive to mesoderm development (Colas et al. 2008).

Mix1 is thought to repress *fgf3*, *fgf4* and *fgf8* expression. The expression domains of the mesodermal genes *fgf4*, *fgf8* and of the mesodermal FGF targets *brachyury* and *myod* were enlarged at stage 10 in response to morpholino knockdown of Mix1 and of both *X. laevis* homeologs Mix1 and Mix2 (Colas et al. 2008). The same result was found using dominant negative Mix.1, although the targets were induced to a lesser extent. Morpholino knockdown of Mix1 or both Mix1 and Mix2 lead to repression of the pronephros mesoderm markers *lim-1* and *pax-8*, but this effect was rescued by inhibition of the FGF receptor. This indicates that elevated expression of *fgf* can repress pronephros induction. *Mix1* overexpression in the marginal zone repressed the mesodermal marker genes *brachyury*, *myod*, *fgf3*, *fgf4* and *fgf8* at the mid-gastrula stage (Colas et al. 2008).

1.6.7 Mix1 in blood development

Several studies have linked Mix1 to blood development, blood forms in the ventral mesoderm (Kumano and Smith 2000). Overexpression of *mix1* at low doses caused a reduction in the number of α -globin expressing cells, and at high doses blocked their production entirely (Lemaire et al. 1998). Mead et al. report the opposite scenario, in which injection of *mix1* mRNA leads to the expression of globin throughout the dorsal-ventral marginal zone and in animal caps (Mead et al. 1996). Injection of the same dose of mRNA at the 2-cell stage was later reported to ventralise the embryo and induce elevated expression of the blood marker *scl* (Mead, Kelley, et al. 1998). These conflicting effects on globin expression may stem from different dosing and localisation of *mix1* expression in different overexpression experiments. Lemaire et al. injected 100pg or 400pg *mix1* mRNA ventrally at the 4 cell stage, whereas Mead et al. injected 1ng at the 1 cell stage (Mead et al. 1996; Lemaire et al. 1998). When injected at 1ng, Mix1 may repress downstream FGF signals, which confine primary blood islands to the ventral region and can repress globin expression

(Kumano and Smith 2000; Colas et al. 2008). Loss of FGF signals can therefore lead to delocalised blood island formation in the dorsal region, which would explain why *mix1*-induced dorsal marginal cells expressed globin (Mead et al. 1996).

It is interesting that ventral overexpression of *mix1* leads to loss of blood formation, as one would expect elevated globin expression in the ventral region if FGF signals were suppressed (Lemaire et al. 1998; Kumano and Smith 2000). Lemaire et al. suggest that Mix1 acts to block terminal differentiation of ventral mesoderm cells (Lemaire et al. 1998). Mouse ESCs expressing *Mixl1* develop embryoid bodies which do not express hematopoietic progenitor markers but have elevated endoderm gene expression (Lim et al. 2009). Mix1 may block mesoderm in favour of endoderm expression.

The opposite result was derived using doxycycline-inducible *Mixl1*-expressing mouse ESCs, which formed embryoid bodies expressing mesoderm markers, including elevated hematopoietic progenitor expression (Willey et al. 2006). *MIXL1* is expressed in human T and B progenitor cells in the bone marrow and is expressed in various leukemic cell lines. Expression of human *MIXL1* in *X. laevis* animal caps induced α -globin which demonstrates a role for human *MIXL1* in hematopoiesis (Guo et al. 2002). Clearly the role of Mix1 in blood formation is complex and not fully understood. Mix1 may repress blood formation in a dose dependent manner and interact with nearby genes to establish the localisation of blood islands.

1.7 Other Mix family genes present in *X. tropicalis*

1.7.1 Mixer induction and localisation

Mixer is a homeodomain transcription factor in the Mix family (Pereira et al. 2012). Mix1 and Mixer are able to form heterodimers, which reduces the ability of Mixer to induce endoderm (Mead, Zhou, et al. 1998).

Maternal VegT activates endodermal genes such as *gata4*, *gata5*, *gata6*, *sox17a*, *a2m* and all of the Mix family genes including *mixer*. Nodal signals downstream from VegT activate *mixer* and *gata5* (Xanthos et al. 2001). The nodal-related genes *cyc* and *sqt* are necessary for appreciable levels of Mixer induction in zebrafish embryos and for endoderm and mesoderm formation (Alexander and Stainier 1999).

Mixer is thought to be induced by zygotic factors, as cycloheximide treated embryos do not express *mixer* (Yasuo and Lemaire 1999). *Mixer* is activated shortly before the MBT so a zygotic activator must function rapidly to activate it (Collart et al. 2014). *Mixer* was expressed in animal caps expressing *nodal1*, *nodal2* or *gdf3* (Yasuo and Lemaire 1999). However, *nodal1* and *nodal2* are activated after *mixer*, and *gdf3* is activated 30 minutes before *mixer* (Collart et al. 2014). It may be plausible that *mixer* activation is not dependent on Nodal1, Nodal2 for its activation, but these genes may maintain its expression, and Gdf3 may be a zygotic activator of *mixer* (Yasuo and Lemaire 1999). In conflict with the finding that *mixer* is zygotically activated, *Mixer* was found to act downstream of a maternal transcription factor. In animal caps, *mixer* is activated downstream of the maternally deposited growth factor *gdf1* (*vg1*), and the induction of endoderm genes downstream of Gdf1 was blocked in animal caps expressing dominant negative *Mixer* (Henry and Melton 1998; Collart et al. 2014).

Mixer was expressed at stage 10 in animal caps injected with *vegt*, furthermore this induction was cell-contact dependent. *Vegt* is expressed as both a maternal and zygotic transcript in the early embryo. A zygotic factor is thought to be responsible for *mixer* induction and this could be zygotic *Vegt* (Collart et al. 2014; Yasuo and Lemaire 1999).

1.7.2 Mixer in endoderm and mesoderm specification

Dominant negative *Mixer-enR* injected embryos displayed gastrulation defects, anterior truncation, abnormal head and gut formation as well as loss of endodermal markers. *Mixer-enR* was used to show that *mixer* is essential for endoderm induction and for the expression of *sox17a* and *sox17b* (Henry and Melton 1998). *Mixer* is activated shortly after *sox17a* and *sox17b* in *X. tropicalis* therefore *Mixer* is unlikely to be the initial activator of the *sox17* genes, but may be necessary to maintain their expression (Collart et al. 2014). Morpholino knockdown of *Mixer* in *X. laevis* delayed gastrulation and caused anterior and gut defects at the tailbud stage. Vegetal cells showed increased mesoderm differentiation. Knockdown embryos showed reduced expression of endodermal genes *gata5/6*, *cer1*, *a2m* and *sox17* and increased expression of mesodermal genes e.g. *brachyury*, *eomes*, *bix1*, *bix4* and *fgf8* and the TGF beta genes *nodal1* and *nodal5*. This shows that *Mixer* can drive endoderm and block mesoderm development (Kofron, Wylie, and Heasman 2004).

The carboxyl terminus and homeodomain of *Mixer* are both necessary to induce endoderm, and a mutant *Mixer* protein comprised of just these two regions can drive endoderm specification in animal caps. *X. laevis* embryos injected with mutant *mixer* mRNA were

ventralised with underdeveloped heads and tails (Doherty et al. 2006). Embryos overexpressing *mix1*, *mixer* or *mix4* were ventralised and with increased blood formation. Overexpression of *mixer* in the whole embryo induced ectopic endoderm expression and caused severe axial abnormalities (Mead, Zhou, et al. 1998).

Microarray analysis of separate Mixer, Nodal and Sox17 depleted embryos revealed common and unique downstream targets. Mixer was found to have 66 unique targets, and 31 shared with Sox17 and 35 shared with Nodal (Sinner et al. 2006). This contrasts with previous work which implicates Sox17 as the major transducer of Mixer signalling (Alexander and Stainier 1999). The majority of Mixer targets were repressed and localised in the equatorial region, suggesting that Mixer may repress mesoderm gene expression (Sinner et al. 2006). *Mixer*- and *sox17b*-expressing animal caps were analysed by microarray to reveal 71 targets, 10 of which were targets of both genes indicating separate and shared functions in endoderm specification (Dickinson, Leonard, and Baker 2006).

1.7.3 Bix1.1 and Bix1.2

The majority of work on *bix1* has been carried out in *X. laevis*, this work has shown that *bix1* is activated by Brachyury and Vegt, and promotes endoderm and ventral mesoderm development (Tada et al. 1998). When *bix1* is expressed in animal caps at low concentrations, ventral mesoderm markers *wnt8* and *ventx1.2* were upregulated, and at higher concentrations of *bix1* expression, the endoderm markers *sox17a*, *a2m* and *fabp2* were induced in animal caps. At all doses of *bix1* expression in animal caps, the anterior mesendoderm markers *gsc* and *cer1* were induced (Tada et al. 1998). This indicates that Bix1.1 and Bix1.2 may have a similar function to Mix1 in inducing endoderm and anterior development.

1.8 Mix family genes in other vertebrates

1.8.1 Mix genes in zebrafish

The zebrafish gene *mixl1*, previously known as *bon*, is a Mix-related gene (Pereira et al. 2012). Zebrafish *mixl1* is expressed in the mesoderm and endoderm and, downstream of nodal signals, induces *sox17* expression and endoderm development (Kikuchi et al. 2000). Mixl1

knockout embryos have a smaller forebrain than wild types and the expression domain of the neural plate marker *otx2* is reduced, implicating Mixl1 in neural patterning. Differentiation is disrupted in the anterior axial mesoderm, demonstrated by reduced *gsc* expression in knockout embryos at 90% epiboly (Trinh, Meyer, and Stainier 2003). Sebox (previously known as Mezzo) is another zebrafish Mix family gene which is activated by activin and nodal signals in the prospective mesendoderm. Sebox has a role in endoderm specification and activates the endodermal genes *sox17* and *sox32/cas* and the mesodermal gene *ntl* (*brachyury*). Reduced endodermal markers were expressed in morpholino injected embryos, and endoderm and heart defects were observed. Mixl1 and Sebox may work cooperatively to pattern mesoderm as *sebox* mRNA can partially rescue the Mixl1 mutant heart defect phenotype (Poulain and Lepage 2002).

1.8.2 Mixl1 in mouse

Mixl1 is the only Mix family gene present in mouse and is orthologous to *Xenopus mix1* (Hart et al. 2002). *Mixl1* is expressed in the primitive streak in early gastrula stage mouse embryos, and later is localised to each end of the primitive streak in tissues fated to become mesoderm and endoderm (Wolfe and Downs 2014). Mixl1 knockout embryos exhibit a shortened axis, abnormal neural folds and heart and gut defects. Mixl1 is thought to regulate cell movements that occur at gastrulation which allow spatial patterning of mesoderm and endoderm. Mixl1 null embryonic stem cells were ineffective at inducing endoderm tissues in chimeras indicating an essential role in endoderm differentiation. (Hart et al. 2002). Mixl1 homozygous knockout mice were found to have reduced recruitment of cells from the primitive streak to the prospective endoderm and increased recruitment of primitive streak cells to mesoderm when compared to heterozygous mice. Furthermore, the endoderm cells remained stationary in the knockdown whilst the mesoderm cells retained their motility, suggesting that Mixl1 functions to drive cell movement in the prospective endoderm, leading to recruitment of cells from the primitive streak into endoderm (Tam et al. 2007).

Mixl1-expressing mouse stem cells expressed elevated levels of the endodermal genes *foxa2*, *sox17*, *cer1* and *e-cadherin*, and the anterior primitive streak marker gene *gsc*, further implicating Mixl1 in endoderm and anterior specification. The finding that Mixl1 binds to and activate the promoters of *gsc*, *sox17* and *e-cadherin* supports the notion that Mixl1 promotes endoderm formation (Lim et al. 2009). Mouse Mixl1 mRNA was injected into *X. laevis* animal

caps to induce the endodermal markers *endodermin*, *gata5* and *pdx1*. Similarly to Mixer in *X. laevis*, Mixl1 promotes endoderm specification in mouse (Mohn et al. 2003).

1.8.3 CMIX the chicken Mix1 orthologue

The Mix1 orthologue found in chicken is CMIX. The *CMIX* homeodomain has 71% sequence similarity to that of *X. laevis mix1* and is expressed in the posterior marginal zone in the early embryo. It later expands to occupy the antero-posterior axis of the primitive streak in ectoderm and mesoderm cells. Its expression pattern is similar to that of *brachyury* in the early and primitive streak stages (Peale, Sugden, and Bothwell 1998).

1.9 Morpholino oligonucleotides

Morpholino oligonucleotides (MOs) are commercially available antisense oligonucleotides of approximately 25 bases which bind to mRNA by base pairing. MOs are resistant to nucleases because they consist of a morpholine ring rather than the ribose ring normally found in nucleic acids (Summerton and Weller 1997). Translation blocking MOs inhibit translation by binding to targeted regions upstream of the start codon, and splice blocking MOs bind to a region overlapping a splice junction which blocks the association of the translational or splicing machinery with the mRNA. Blocking translation or splicing prevents mature protein being processed and therefore gene function is blocked (Eisen and Smith 2008). MOs disperse quickly throughout the stage 1 *Xenopus* embryo and are a convenient tool for inducing loss of gene function without having to produce a transgenic line (Nutt et al. 2001).

MOs have non-ionic backbones, rather than the negative backbones naturally occurring in nucleic acids which often interact with proteins. This means that MOs are unlikely to cause toxic and off-target effects other than those caused by non-specific RNA binding. MOs are also highly stable meaning no degradation products are present which could be toxic to cells (Summerton 2007). MOs have a high affinity for RNA which means they can associate even in the presence of secondary structures and can diffuse between the nucleus and cytoplasm to block both splicing and translation (Summerton 2007). The optimum temperature for morpholino activity is 37°C, which is higher than the incubation temperature for *Xenopus* embryos which is approximately 25°C for *X. tropicalis* and 22°C for *X. laevis*. Target sequences with a higher A-U content are thought to be more amenable to specific interactions at low temperatures (Summerton 2007).

MOs can cause off-target effects by mis-targeting mRNA sequences and causing depletion of the incorrect protein (Eisen and Smith 2008). It can be difficult to differentiate between effects caused by gene knockdown of an intended gene and off-target effects. For this reason, validation steps are very important in order to be confident of any results. The most reliable way to validate a MO is by rescue as this proves that knockdown is caused by depletion of the intended protein. This is done by injecting both a MO and MO-resistant mRNA into the embryo to see if phenotypic or downstream gene expression effects can be reversed (Fletcher and Harland 2008). Partial rescue is an acceptable validation, whereby a certain percentage of embryos are restored to the normal phenotype, or the phenotype is less severe in rescue embryos than in knockdown embryos. Rescue is often difficult as replicating the dose, timing and amount of mRNA can be problematic (Eisen and Smith 2008). Antibodies can be used to demonstrate depletion of protein as a result of MO knockdown.

A useful control is a 5-base mismatch MO because it is designed to resemble the original MO without blocking translation or splicing of the target gene. A standard control MO is available from Gene Tools and targets the human β -globin pre-mRNA and can be used as an injected control (Eisen and Smith 2008). Both of these controls are limited as the specific sequence of a MO is likely to cause any off-target effects.

There are cases where MO experiments seem to indicate a particular gene function, backed up by rescue, for example in zebrafish, *Prox1b* was found to be essential for lymphatic development, and the MO phenotype was rescued by mRNA injection (Del Giacco, Pistocchi, and Ghilardi 2010). A later study using two MOs targeting *Prox1b* and a *Prox1b* zebrafish mutant were found to have no lymphatic defects, contradicting the previous finding that this gene is essential for normal lymphatic development (Tao et al. 2011). The lymphatic defect was possibly an off-target effect of the *Prox1b* MO used in the earlier study. Alternatively, the first result could be genuine, and compensatory mechanisms may be acting in mutant embryos.

A study in zebrafish set out to recapitulate knockdown phenotypes found in MO-injected embryos by generating mutants using zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the clustered, regularly interspaced, short palindromic repeats (CRISPR)-Cas9 system (Kok et al. 2015). All three approaches are used to generate a targeted gene disruption to generate mutant lines. 24 mutant lines were generated and the phenotypes were compared to those reported for the MO knockdown of these genes, and only three phenotypes were consistent in both (Kok et al. 2015). A wider comparison of

mutant and morphant phenotypes was carried out by comparing mutants reported in the Sanger Zebrafish Mutation Project (ZMP), where 24 cases were found in which a MO knockdown generated an overt phenotype (Kok et al. 2015). Only five of these 24 genes had consistent defects induced by both loss-of-function strategies. These two comparisons reveal that over 80% of MO phenotypes were not recapitulated in mutant lines (Kok et al. 2015).

Interestingly, five of the genes for which the phenotype was not recapitulated in mutant embryos had been validated through phenotypic rescue in MO knockdown studies (Kok et al. 2015). New evidence suggests that compensatory gene regulatory networks can be activated in knockout lines, which essentially rescue the effect of gene deletion (Rossi et al. 2015). Therefore, MOs may be a more useful tool for loss-of-function analysis in certain cases. Caution should be taken when using MOs to infer gene functions. A useful control would be the comparison of morphant and mutant phenotypes as a confirmation of their specific effect (Schulte-Merker and Smith 1995).

Chapter 2- Materials and Methods

2.1 Xenopus techniques

2.1.1 Xenopus tropicalis embryo generation

Animal procedures were performed under license, as required by Animals (Scientific Procedures) Act 1986 (UK). *Xenopus tropicalis* embryos were generated through *in vitro* fertilisation (IVF). Female frogs were primed by subcutaneous injection of 10 units of hCG (Human chorionic gonadotropin Sigma CG10) in 100ul sterile water into the dorsal lymph sac 20 hours prior to IVF. The following day the female frogs were boosted by injection of 100 units of hCG in 100ul sterile water to induce ovulation. Approximately 4 hours later the frogs were held over a petri dish containing a drop of 1XMMR and gently squeezed on the lower abdomen to collect eggs. To prepare testes for IVF, male frogs were killed according to schedule 1 guidelines by immersion in 0.2% Ethyl 3-aminobenzoate methanesulfonate (MS222, Sigma E10521) for 15 minutes followed by decapitation. The testes were dissected and crushed with a pestle in an Eppendorf tube containing Leibovitz's L-15 Medium (Sigma L5520) + 10% fetal bovine serum (FBS, Sigma 12003C). The testes solution was mixed with the eggs and after 4 minutes the dish was flooded with 0.05X MMR; a low salt solution for egg activation. Approximately 10 minutes later the embryos may show signs of fertilisation such as cortical rotation and contraction of the pigmented region. The embryos were de-jellied in 2.2% cysteine in 0.05X MMR for 3-5 minutes before being washed in 0.05X MMR. Embryos were then incubated in 0.05X MMR + gentamicin (Sigma G1914) at 25-28°C. All solutions were pH7.8.

2.1.2 Xenopus Laevis embryo generation

Female *X. laevis* were primed 3-7 days prior to use by injection of 50 units of hCG. The day before use they were boosted by injection of a further 500 units of hCG. Female frogs were kept in tanks of 1xMMR to lay eggs directly into a high salt buffer to preserve egg quality. To prepare testes for IVF, male frogs were killed according to schedule 1 guidelines by immersion in 0.2% Ethyl 3-aminobenzoate methanesulfonate (MS222, Sigma E10521) for 30 minutes followed by decapitation. Testes were then dissected and could be stored at 4°C for up to one week in 70% L-15. To fertilise eggs *in vitro*, all 1X MMR was removed from a petri dish containing eggs, and a small piece (approximately 1/6) of testis was dissected and homogenised in an Eppendorf tube of 1XMMR. The sperm media was mixed in with the eggs,

incubated at room temperature for 6 minutes and the plate was then flooded with 0.1X MMR to lower the salt concentration of the buffer. This allows the eggs to activate and induces cortical rotation in those which are fertilised and takes around 15 minutes. Embryos were then de-jellied in 1% cysteine in distilled water for approximately 5 minutes, when the jelly coat has visibly gone. Embryos were then washed twice on 0.1X MMR and injected in small plates containing 2.5% ficoll in 0.1XMMR embryos were then incubated at 14°C overnight.

2.1.3 Morpholino Microinjection

Glass capillaries (WPI 1B100F-4) were pulled using a micropipette puller (Sutter p97) to make needles for microinjection. For *X. tropicalis* needles the heat was determined by a heat ramp and the settings were; pull 180, velocity 80, time 60. For *X. laevis* needles the heat was determined by a heat ramp and the settings were; pull 55, velocity 55, time 10.

Embryos were injected in plates containing filter sterilised 2.5% ficoll in 0.05X MMR. Morpholino oligonucleotides can be injected directly into embryos to knock down gene expression by specifically targeting mRNAs to block splicing or translation. Morpholinos were supplied by gene tools and stock solutions diluted in distilled water at 15ng/nl. For injection 10ul of each morpholino stock was diluted to 10ng/nl in 5ul 10% Dextran Alexa Fluor 488 10,000MW dye (Life Technologies D-22910) or 5ul 10% Dextran Texas Red 10,000MW (Life Technologies D-1863). 1nl was injected for a 10ng dose into each embryo. Embryos were injected at the 1 cell or 2 cell stage, following injection they were kept in 2.5% ficoll for 1 hour. The non-cleaving embryos were removed and using a fluorescence microscope (Leica M165FC and Leica UV Lamp) any non-injected embryos were removed.

See table 2.1 for antisense morpholino sequences.

2.1.4 In vivo morpholino assay

To test translation blocking MO2 *in vivo*, embryos were injected with 50pg *mixer*-HA mRNA, and then split into four plates. Three plates were injected with different doses of MO2 at 5ng, 10ng and 15ng doses. Embryos were then incubated at 25°C and 40 embryos from each injected condition, and WT embryos were harvested at 7.0 hpf and protein was immediately extracted as described above. Immunoprecipitation and western blot were carried out as described below.

2.1.5 Time course knockdown experiment

Embryos were injected as described above at the one cell stage with either a morpholino or control morpholino and some were left as un-injected WT controls. The first 6 cleavages were observed and rapid or slow cleaving embryos were removed to improve synchronicity of cell cycles within each clutch. The timing of the cleavages was also recorded in order to predict the timing of MBT for a particular clutch. Embryos were then cultured in 1/20X Marc's Modified Ringers (MMR) + gentamycin (100ug/ml) at 25°C. For each time point in each time-series, 5 embryos were harvested for each condition and were homogenised by pipetting up and down in 200ul Trizol (Life Technologies). Samples were then snap frozen in liquid nitrogen.

2.1.6 RNA extraction from *Xenopus* tissue

The trizol homogenate samples were thawed and 80ul of chloroform was added. Samples were vortexed thoroughly and centrifuged at 13000RPM at 4°C for 20 minutes to allow separation of the phases. The top layer was collected, an equal volume of isopropanol and samples were left to precipitate at -20°C for at least 30 minutes. Tubes were then centrifuged at 13000RPM at 4°C for 30 minutes, washed in 70% ethanol and centrifuged for a further 5 minutes at 13000RPM at 4°C. Samples were then re-suspended in 40ul nuclease-free water (Ambion AM0038). Samples to be used for qPCR were DNase treated with 1ul DNase I (Ambion AM2222) for each 1ug DNA present in each sample and incubated at 37°C for 20 minutes. An equal volume of LiCl (Sigma L7026) was added to each sample and precipitated for at least 1 hour at -20°C. Samples were then centrifuged at 13000RPM at 4°C for 30 minutes, washed in 70% ethanol and centrifuged for a further 5 minutes at 13000RPM at 4°C. Samples were air dried for 5 minutes then re-suspended in 40ul nuclease-free water. 1ul of RNA from each sample was quantified using a NanoDrop (Thermo scientific ND1000 Spectrophotometer). Samples to be used in RNA-seq library preparation were sent for QC (see below for details).

2.1.7 Animal Cap Explants

Animal cap experiments were carried out in *X. laevis* as the larger embryos and slower divisions are more practical for dissecting animal caps. Embryos were injected with 250pg *mix1* mRNA into the top of the animal pole at the 1-cell stage. Injected and WT embryos were placed at 14°C overnight in 0.1X MMR. At stage 8.5, the embryos were placed on an agarose

dish in 0.7X MMR and forceps were used to remove the vitelline envelope. The animal cap was then dissected using forceps and placed in an agarose plate with 0.7X MMR. Animal caps and whole embryos were then incubated at 22°C. After two hours, animal caps were harvested at stage 9. 30 animal caps were pooled and homogenised in trizol for each sample, by pipetting up and down, and snap frozen in liquid nitrogen. RNA was extracted and qPCR was carried out as described below to assay for transcriptional induction in animal caps. *Gsc* primers were used to assay for induction by *mix1* mRNA, *vegt* was used as a negative control. Illumina TruSeq libraries were then prepared and sequenced as described below.

2.1.8 Over-expression experiments

To test the effects of over-expressing a gene, *mix1* mRNA was injected into the 1 cell embryo. Concentrations of 25ng/ul or 50ng/ul were prepared with 5% Dextran Texas Red 10,000MW (Life Technologies D-1863). 1nl was injected to give a dose of 25pg or 50pg, respectively. The mRNA was injected into the vegetal pole where endogenous *mix1* is expressed.

2.1.9 Whole Mount *in situ* Hybridisation

Embryos were fixed in MEMFA overnight at 4°C in glass vials, the following day they were gradually dehydrated by washing for 5 minutes in 75% MEMFA/25% methanol, then 50% MEMFA/50% methanol, then 25% MEMFA/75% methanol and finally 100% methanol for 3 washes. Fixed embryos can then be stored at -20°C long term.

For *in situ* hybridisation, fixed embryos were first rehydrated in methanol gradually with 5 minute washes in 100% methanol, 75% methanol, 50% methanol, 25% methanol/75% PBST, and finally 3 washes in PBST. Embryos were permeabilised for 5 minutes in 10ug/ml proteinase K (Sigma) in PBST and washed twice for 5 minutes in 0.1M triethanolamine (TEA Sigma 90279). 12.5ul of acetic anhydride (Sigma 320102) was added to 5ml TEA and added to embryos for two washes. Embryos were then re-fixed in 4% paraformaldehyde (Santa Cruz sc-281692) for 20 minutes, and then washed 5 times in PBST for 5 minutes. Embryos were then submerged in hybridisation buffer and incubated at 60°C for 3 hours to pre-hybridise embryos. Embryos were then separated into baskets which were individually placed in glass vials and 500ul DIG labelled antisense RNA probe was added at 1ug/ml and incubated overnight at 60°C.

The following day embryos were given a 5 minute wash in hybridisation buffer at 60°C and 3x 30 minute washes in 2XSSC at 60°C. Probes were stored at -20°C for re-use. They were then incubated for 30 minutes at 37°C in 20ug/ml RNase A (Sigma) and 10ug/ml RNase T1 (Sigma). This was followed by a 10 minute incubation at room temperature in 2XSSC and 2x 30 minute incubations in 0.2XSSC at 60°C. Embryos were then given 2x 15 minute washes in MAB before blocking for an hour in MAB +2% BMB Blocking reagent (Roche 11096176001). Anti-digoxigenin AP antibody (Roche 11093274910) was then applied in MAB +2%BMB at 1:3000 concentration and incubated for 4 hours at room temperature. Embryos were then given 2x 5 minute MAB washes at room temperature and an overnight MAB wash shaking gently at 4°C.

The following day embryos were washed at room temperature 3x in MAB for 15 minutes, then 2x in alkaline phosphate buffer for 5 minutes. Embryos were then removed from baskets and placed in a 24 well plate and submerged in BM purple (Roche 11442074001) and wrapped in aluminium foil to protect from light and kept at room temperature. The duration of incubation of BM purple depends on the probe so embryos were checked hourly for sufficient staining. If necessary, the plate could be left overnight at 4°C to allow the reaction to develop slowly. To stop the reaction, embryos were washed twice in MAB. Embryos were then fixed in Bouin's solution (VWR 7000.1000) for 1 hour at room temperature. Embryos were washed for 5 minutes in 70% ethanol, then 100% methanol. Bleaching solution was applied (formula) to remove pigment from the embryos by placing on a light box for 2 hours. Embryos were then washed in 100% methanol, and rehydrated with 5 minute washes in 75% methanol/25% PBST, 50% methanol/50% PBST, 25% methanol/75% PBST and 2 washes in PBST.

Embryos were placed on a thick agarose petri dish in PBS for imaging on the microscope (Leica M165FC).

2.1.10 β -Galactosidase Staining

To detect the morpholino injection site, NLS- β -Galactosidase mRNA was co-injected with Mix1 MO into 1 cell of the stage 2 embryo to give unilateral depletion. The reaction for β -Galactosidase was carried out using Salmon gal (Apollo scientific) to give a red nuclear stain. To prepare Salmon gal 1ml was set up as follows- 20ul 50X Ferricyanide, 20ul 50X Ferrocyanide, 2ul 1M MgCl₂, 10ul 100X Salmon gal, 948ul PBS. To stain *X. tropicalis* embryos, they were first fixed in MEMFA for 20 minutes, washed twice in PBS and then submerged in

the Salmon gal stain for 40 minutes. Finally, embryos were washed twice in PBS and fixed overnight in MEMFA.

2.2 Sequencing Techniques

2.2.1 Illumina RNA-seq library preparation

For Illumina library preparation, the TruSeq RNA Sample Prep Kit v2 (Illumina kit RS-122-2001) was used to generate polyA+ RNA libraries.

For library preparation 400ng-1.5ug of mRNA was suspended in 50ul nuclease-free water for each sample. For the animal cap sequencing libraries 400ng of mRNA was used, for the knockdown time-series sequencing libraries 1.5ug of mRNA was used, for the RiboZero and accompanying PolyA+ time-series 1ug of mRNA was used. Starting amounts were decided to allow sufficient mRNA for library repeats in the case of failure during preparation.

1. Purify and fragment mRNA

50ul of RNA Purification Beads were mixed into each mRNA sample and heated in a thermocycler (MJ Research PTC-225) at 65°C for 5 minutes, then incubated at room temperature for 5 minutes to allow the mRNA and beads to bind.

The plate was then placed on a magnetic stand for 5 minutes at room temperature, the supernatant was removed and beads were resuspended in 100ul bead washing buffer by pipetting up and down. The plates were incubated for 5 minutes at room temperature and then on the magnetic stand for 5 minutes before removing the supernatant.

The beads were then suspended in 50ul elution buffer and placed in a thermocycler at 80°C for 2 minutes. 50ul of Bead Binding Buffer was mixed into each well and incubated for 5 minutes at room temperature. The plate was then placed on a magnetic stand for 5 minutes and the supernatant was removed. The beads were washed once more in 200ul bead washing buffer. Finally, the beads were suspended in 19.5ul Elute, Prime, Fragment Mix and the plate was incubated in a thermocycler at 94°C for 8 minutes.

2. Synthesise first strand cDNA

The plate was placed on the magnetic stand for 5 minutes and 17ul of supernatant was placed into a new plate. 50ul of Superscript II polymerase was added to the tube of First Strand Master Mix and mixed. 8ul of the First Strand Master Mix was added to each well of the plate. The plate was then incubated in a thermocycler with the following program:

25°C	10 minutes
42°C	50 minutes
70°C	15 minutes
4°C	hold

3. Synthesise second strand cDNA

25ul of Second Strand Master Mix was then added to each well and incubated in a thermocycler at 16°C for 1 hour.

90ul of well-mixed AMPure XP beads were mixed into each well and incubated at room temperature for 15 minutes. The plate was incubated at room temperature for 5 minutes. 135ul of supernatant was then removed from each well. Remaining on the magnetic stand, 200ul of 80% ethanol was placed in each well without disturbing the beads, and then removed. The ethanol washing process was repeated once. The plate was left at room temperature to dry for 15 minutes.

The beads were then suspended in 62.5ul of Resuspension Buffer, incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes. 60ul of supernatant was removed from each well and placed into a new plate.

4. Perform End Repair

40ul of End Repair Mix was mixed into each well and the plate was incubated in a thermocycler at 30°C for 30 minutes.

160ul of AMPure XP Beads were mixed into each well, incubated at room temperature for 15 minutes and placed on the magnetic stand for 5 minutes. 127.5ul of supernatant was removed twice. 200ul of 80% ethanol was placed in each well

without disturbing the beads, and then removed. The ethanol washing process was repeated once. The plate was left at room temperature to dry for 15 minutes.

The dried bead pellet was suspended in 20ul Resuspension Buffer and the plate was incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes. 17.5ul of supernatant was then removed from each well and placed into a new plate.

5. Adenylate 3' Ends

12.5ul of A-Tailing Mix was mixed into each well and the plate was incubated in a thermocycler at 37°C for 30 minutes.

6. Ligate Adapters

A different adapter was added to each well for every row of 8 samples. For each adapter a mix was made containing 3ul Resuspension Buffer, 3ul Ligation Mix and 3ul RNA Adapter Index. Then 7.5ul of the mix was pipetted into each well of the plate and the whole volume was mixed. The plate was incubated in a thermocycler at 30°C for 10 minutes. 5ul of Stop Ligation Buffer was then added to each well and the entire volume mixed.

42ul of AMPure XP Beads were mixed into each well, incubated at room temperature for 15 minutes and placed on the magnetic stand for 5 minutes. 79.5ul of supernatant was removed from each well. 200ul of 80% ethanol was placed in each well without disturbing the beads, and then removed. The ethanol washing process was repeated once. The plate was left at room temperature to dry for 15 minutes.

The bead pellets were then suspended in 52.5ul Resuspension Buffer and the plate was incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes. 50ul of supernatant was then removed from each well and placed into a new plate.

50ul of AMPure XP Beads were mixed into each well, incubated at room temperature for 15 minutes and placed on the magnetic stand for 5 minutes. 79.5ul of supernatant was removed from each well. 200ul of 80% ethanol was placed in each well without

disturbing the beads, and then removed. The ethanol washing process was repeated once. The plate was left at room temperature to dry for 15 minutes.

The bead pellets were suspended in 22.5ul Resuspension Buffer and the plate was incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes. 20ul of supernatant was then removed from each well and placed into a new plate.

7. Enrich DNA Fragments

A PCR reaction was set up to amplify the fragments. 5ul of PCR Primer Cocktail and 25ul of PCR Master Mix were mixed into each well. The plate was then incubated in a thermocycler with the following program:

98°C	30 minutes	
98°C	10 seconds	15 cycles
60°C	30 seconds	
72°C	30 seconds	
72°C	5 minutes	
10°C	hold	

50ul of AMPure XP Beads were mixed into each well, incubated at room temperature for 15 minutes and placed on the magnetic stand for 5 minutes. 95ul of supernatant was removed from each well. 200ul of 80% ethanol was placed in each well without disturbing the beads, and then removed. The ethanol washing process was repeated once. The plate was left at room temperature to dry for 15 minutes.

The bead pellets were then suspended in 32.5ul Resuspension Buffer and the plate was incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes. 30ul of supernatant was then removed from each well and placed into tubes for sequencing.

The RNA-seq library preparation was carried out over two days, the stopping point was usually after end repair or adapter ligation, and plates were left overnight at -20°C. Libraries were sequenced on the Illumina Hi-Seq 2000.

2.2.2 RiboZero and accompanying PolyA+ library preparation

To generate comparable RiboZero and PolyA+ RNA-seq libraries, 1ug was taken from each time-series sample, and was split to process 500ng as RiboZero and 500ng as PolyA+ libraries. The first steps were separate to purify polyA+ and ribosomal-depleted RNA, and then the samples were processed simultaneously for the remainder of the library preparation using the ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Epicentre kit SSV21106).

1. PolyA+ mRNA purification

For polyA+ mRNA purification, the Illumina TruSeq RNA Sample Prep Kit v2 (Illumina kit RS-122-2001) was used to for the first few steps according to the following protocol:

50ul of RNA Purification Beads were mixed into each mRNA sample and heated in a thermocycler (MJ Research PTC-225) at 65°C for 5 minutes, then incubated at room temperature for 5 minutes to allow the mRNA and beads to bind.

The plate was placed on a magnetic stand for 5 minutes at room temperature, the supernatant was removed and beads were resuspended in 100ul bead washing buffer by pipetting up and down. The plates were incubated for 5 minutes at room temperature and then on the magnetic stand for 5 minutes before removing the supernatant.

The beads were suspended in 11ul elution buffer and placed in a thermocycler at 80°C for 2 minutes. The plate was then placed on a magnetic stand for 5 minutes and 10ul of supernatant was collected from each well.

2. Ribosomal RNA depletion

For ribosomal RNA depletion for RiboZero library preparations the RiboZero ScriptSeq complete gold low input kit was used (Epicentre kit SCL24EP).

90ul of magnetic beads were prepared for each sample by placing beads in a 1.5ml microcentrifuge tube, placing the tube on a magnetic stand for one minute, removing supernatant and re-suspending beads in RNase-free water. Beads were then vortexed briefly and washed twice more in RNase-free water using the magnetic stand. Finally, the RNase-free water was removed and beads were suspended in 35ul Magnetic Bead Resuspension Solution per sample. 35ul of magnetic beads was placed into a new 1.5ml microcentrifuge tube in preparation for the RNA samples.

RNA samples were then treated by adding 2ul RiboZero rRNA Removal Solution to each RNA sample of 500ng, along with 2ul of RiboZero Reaction buffer and RNase-free water up to a total volume of 20ul. Samples were mixed by pipetting and incubated at 68°C for 10 minutes and room temperature for 5 minutes.

RNA samples were placed in the 1.5ml microcentrifuge tubes containing 35ul of washed magnetic beads and mixed thoroughly by pipetting and immediate vortexing. This was then repeated for each sample to ensure thorough mixing. Samples were incubated at room temperature for 5 minutes. Samples were then vortexed for 5 seconds and incubated at 50°C for 5 minutes. Tubes were placed on a magnetic stand for at least 1 minute, 50ul of supernatant was removed from each tube and placed into a new 1.5ml microcentrifuge tube.

The ribosomal RNA depleted samples were then purified using AMPure XP beads. 100ul of mixed AMPure XP beads were added to each sample by pipetting, tubes were incubated at room temperature for 15 minutes and placed on the magnetic stand for 5 minutes. The supernatant was removed from each tube. 200ul of 80% ethanol was placed in each tube without disturbing the beads, and then removed. The ethanol washing process was repeated once. The tubes were left at room temperature to dry for 15 minutes. The dried bead pellets were suspended in 11ul RNase-free water and tubes were incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes. The supernatant was removed from each tube and placed into a new 1.5ml microcentrifuge tube.

3. ScriptSeq library preparation

The maximum volume of 9ul was taken from RiboZero and polyA+ RNA samples for use in the ScriptSeq library preparation.

To fragment RNA and anneal the cDNA synthesis primer 2ul cDNA Synthesis Primer and 1ul RNA Fragmentation Solution were added to each 9ul RNA sample in a 0.2ml PCR tube. Tubes were incubated at 85°C for 5 minutes and then placed on ice.

To synthesise cDNA, for each sample 3ul cDNA Synthesis PreMix, 0.5ul 100mM DTT and 0.5ul StarScript Reverse Transcriptase were mixed by pipetting. The 4ul of cDNA Synthesis Master Mix was then added to each RNA sample on ice and mixed by pipetting. Samples were incubated at 25°C for 5 minutes and 42°C for 20 minutes. Once reactions had cooled to 37°C, the thermocycler was paused and 1ul of Finishing Solution was mixed into each sample by

pipetting. Samples were incubated for a further 10 minutes at 37°C on the thermocycler. Reactions were then incubated at 95°C for 3 minutes. During the incubation the Terminal Tagging Master Mix was prepared by mixing 7.5ul of Terminal Tagging Premix with 0.5ul DNA Polymerase and placing on ice. Reactions on the thermocycler were then cooled to 37°C, 8ul of the Terminal Tagging Master Mix was added individually to each reaction and mixed by pipetting. Samples were then incubated at 25°C for 15 minutes, 95°C for 3 minutes and then placed on ice.

To purify the cDNA, samples were placed in a 96 well plate and 45ul of mixed AMPure beads were added to each sample by pipetting. The plate was incubated at room temperature for 15 minutes and placed on the magnetic stand for 5 minutes. The supernatant was removed from each well. 200ul of 80% ethanol was placed in each well without disturbing the beads, and then removed. The ethanol washing process was repeated once. The plate was left at room temperature to dry for 15 minutes. The dried bead pellets were suspended in 24.5ul RNase-free water and the plate was incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes. 22.5ul of supernatant was then removed from each well and placed into a new 0.2ml PCR tube.

To amplify the library and index samples, each 22.5ul sample of cDNA was mixed with 25ul FailSafe PCR PreMix E, 1ul Forward PCR Primer, 1ul ScriptSeq Index PCR Primer and 0.5ul FailSafe PCR Enzyme.

Samples were then placed on a thermocycler with the following program:

95°C	1 minutes	
95°C	30 seconds	15 cycles
55°C	30 seconds	
68°C	3 minutes	
68°C	7 minutes	

To purify the RNA-seq libraries, samples were placed in a 96 well plate and 50ul of mixed AMPure XP beads were added to each sample by pipetting. The plate was incubated at room temperature for 15 minutes and placed on the magnetic stand for 5 minutes. The supernatant was removed from each well. 200ul of 80% ethanol was placed in each well without disturbing the beads, and then removed. The ethanol washing process was repeated

once. The plate was left at room temperature to dry for 15 minutes. The dried bead pellets were suspended in 20ul RNase-free water and the plate was incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes. The clear supernatant was then removed from each well and placed into a new 1.5ml microcentrifuge tube.

2.2.3 Quality control of RNA and libraries

For quality control of RNA samples prior to library preparation, samples were analysed for quality using the Agilent 2100 bioanalyzer (G2943CA). For quality control of libraries prior to sequencing, libraries were quantified using with the Qubit 2.0 flourometer (Q32851) with the Invitrogen high sensitivity assay kit (5067-4626) and analysed for quality using the Agilent 2100 bioanalyzer (G2943CA). All QC and sequencing of libraries was carried out by Leena Bhaw-Rosun, Abdul Sesay and Debbie Jackson. Following successful quality control, libraries were sequenced on the Illumina HiSeq 2000.

2.3 Molecular biology techniques

2.3.1 Synthesis of cDNA

RNA was extracted as described above and the Transcriptor First Strand cDNA synthesis Kit (Roche 04897030001) was used for reverse transcription. 1ug total RNA was diluted to 11.4ul in nuclease-free water and mixed gently with 2ul Random Hexamer primer. Samples were heated at 65°C for 10 minutes to denature RNA, then immediately placed on ice.

The following reaction master mix was set up for all samples-

4ul Buffer

0.5ul RNase Inhibitor

2ul Deoxynucleotide Mix

1ul DTT

1.1ul Reverse Transcriptase

8.6ul was added to each tube and mixed gently. A control sample containing all components apart from the reverse transcriptase enzyme was also set up. All samples were then placed on a thermocycler with the following settings-

Temperature	Time
25°C	10 minutes
55°C	30 minutes
85°C	5 minutes
4°C	hold

2.3.2 Quantitative PCR

QPCR reactions were set up on ice using the KAPA SYBR FAST qPCR Master Mix 2x (Kapa Biosystems KK4602). KAPA SYBR Master mix was kept out of direct light when possible, and allowed to reach room temperature before use. For each primer pair to be used, a master mix was prepared for all reactions as follows-

5ul KAPA Master Mix 2x

3.6ul Nuclease-free water

0.2ul 10uM Forward Primer

0.2ul 10uM Reverse Primer

The master mix was pipetted manually into wells of a 384 well clear qPCR plate, then 1ul of the appropriate cDNA was added to each well. CDNA was usually diluted 1:2 or 1:3 depending on concentration. 3 technical replicates were run for each cDNA sample and values were averaged during analysis. For each primer set, a standard curve of increasingly diluted cDNA was run. A 1:2 dilution series was generated using a cDNA sample to give a series of 6 samples. 1ul triplicates of each dilution were included on the PCR plate. Plates were sealed, centrifuged at 1000RPM for 1 minute and placed in the Light Cycler 480 II (Roche). The thermocycler conditions were as follows-

Program	Temperature	Time	
Pre incubation	95°C	3 minutes	
Amplification	95°C	3 seconds	40 cycles
	60°C	20 seconds	
	72°C	1 second	
Melting curve	95°C	3 seconds	

	60°C	20 seconds	
	72°C	1 second	
	95°C	continuous	
Program	40°C	30 seconds	

To quantify the amplified products, the LightCycler 480 software calculated concentrations based on the cp values and the standard curves run alongside samples. Using Excel (Microsoft), the concentrations were normalised against a control gene by averaging replicate samples, and dividing the average of the housekeeping gene *odc1*, by the average of each cDNA sample. The Student's t-test was applied to the concentration values using Excel to compare control and experimental samples.

See tables 2.2 and 2.3 for list of qPCR primers.

2.3.3 PCR and Agarose Gel Electrophoresis

PCR reactions were set up using Phusion High Fidelity PCR Master Mix with HF Buffer (NEB M0531S). The 20ul reaction was set up on ice:

1ul 10uM Forward Primer
 1ul 10uM Reverse Primer
 1ul template cDNA
 10ul 2x Phusion Master Mix
 7ul Nuclease-free water

When running several reactions, a master mix was made and aliquoted, before adding 1ul cDNA to each PCR tube.

Samples were placed in a thermocycler (MJ Research PTC-225) with the following settings-

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
25 Cycles	98°C 55-65°C 72°C	10 seconds 30 seconds 1 minute – repeat x24 cycles
Final Extension	72°C	10 minutes
Hold	4°C	hold

The annealing temperature is usually 55-65°C and depends on the specific annealing properties of each primer pair. New primers were tested with a gradient from 55-65°C across the thermocycler plate to test for optimal conditions.

A 1.5% agarose gel was prepared by dissolving 1.5g agarose in 100ml 1X TAE and heating in a microwave for 2 minutes. After cooling, 1ul SybrSafe (Life Technologies S33102) was added and agarose was poured into a mould with comb for wells and left for at least 30 minutes to set. PCR products were mixed with 5X DNA loading buffer blue (Bioline BIO-37045) to a final concentration of 1X, and 20ul was loaded onto the gel. Hyperladder 1kb (Bioline BIO-33053) or Hyperladder 100bp (Bioline BIO-33056) was loaded alongside samples as a marker. The gels were run at 100V for the appropriate time to give good separation of bands. Gels were then viewed on the gel dock (BioRad ChemiDoc XRS+ 170-8265) using the Quantity One software.

2.3.4 RT-PCR to test activity of Mix1 splice blocking morpholino

Primers targeted to exon 1 and 3 were forward (TGGACTCATTCAGCCAACAA) and reverse (ACTGGCATCTGCTCAGGTCT) to give a PCR product of 827bp in control cDNA and 1332bp in Mix1 morpholino injected cDNA due to inclusion of the first intron. Primers targeted to exon 2 and exon 3 were forward: (AGCCCGGTACAGAACATCAG) and reverse: (GCAGAACATTGCCAAACTCA) to give a PCR product of 947bp in both conditions.

2.3.5 *In vitro* transcription of mRNA

The mMESSAGE mMACHINE® SP6 Transcription Kit (Life technologies AM1340) was used to generate capped mRNA.

The following reaction was set up-

10ul NTP/CAP

2ul 10X Reaction Buffer

1ug linear template DNA

2ul Enzyme Mix

Up to 20ul Nuclease-free Water

Reactions were mixed gently and incubated at 37°C for 2-4 hours. Then 1ul TURBO DNase was added, mixed and incubated at 37°C for 15 minutes. 30ul Nuclease-free Water and 30ul LiCl were added to precipitate RNA for 1 hour at -20°C. Samples were then centrifuged at 13000RPM for 30 minutes, washed in 500ul 70% ethanol, and centrifuged at 13000RPM for 10 minutes. Samples were then resuspended in 40ul Nuclease-free water. 1ul of RNA from each sample was quantified using a NanoDrop. 2ul was run on a 1.5% agarose gel to check for a distinct band which indicates good quality RNA.

2.3.6 Molecular Cloning

CDNA clones were generated using Zero Blunt® TOPO® PCR Cloning Kit (Life technologies K2800-02) or pENTR™/D-TOPO® Cloning Kit (Life Technologies K2400-20). Both kits are useful for efficient cloning, the Zero Blunt TOPO kit is used generate clones used for transcription of antisense RNA probes for *in situ* hybridisation. The pENTR™/D-TOPO kit is designed for directional cloning, and can be used for flipping the vector insert into destination vectors using the gateway system.

To clone into the pENTR™/D-TOPO vector, forward primers had an additional 5'CACC overhang sequence which is used to insert the PCR fragment into the vector in the correct orientation. To clone into the Zero Blunt TOPO kit, primers do not require any additional sequence and blunt PCR products can be cloned. Phusion High Fidelity PCR Master Mix with HF Buffer (NEB M0531S) was used to generate blunt ended PCR fragments. The PCR product was isolated by visualising the PCR band on the ChemiDoc and cutting out the band. DNA was

extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen 28706) according to manufacturer's guidelines.

The following TOPO cloning reaction was set up and incubated at room temperature for 30 minutes-

0.5-2ul PCR product

1ul salt solution

0.5ul TOPO vector

Up to 6ul sterile water

2.3.7 Gateway Cloning

The Gateway cloning system utilises the attL sites in TOPO entry vectors for efficient recombination of insert sequence into a destination vector containing attR sites. The pENTR™/D-TOPO vectors produced were used for insert recombination into Gateway pCS2+ vectors. This was useful for generating tagged constructs, a C-terminal HA tagged pCS2+ clone was used for making HA tagged constructs. pENTR™/D-TOPO vector inserts contained a coding sequence with the stop codon removed to allow expression of C-terminal tags. The Gateway® LR Clonase® II Enzyme mix (Life Technologies 11791-020) was used to catalyse the reaction which was set up as follows-

1ul entry vector 50ng/ul

1ul destination vector 150ng/ul

6ul TE buffer pH8

2ul LR Clonase Enzyme Mix

Reactions were vortexed briefly and incubated for 1 hour minutes at 25°C then 1ul of Proteinase K was added to stop the reaction and 2ul was transformed into bacterial cells. See tables 2.4 and 2.5 for cloning primer sequences.

2.3.8 Bacterial Transformation and plasmid DNA isolation

2ul of each cloning reaction was gently mixed with 25ul of One Shot® Chemically Competent E. coli (Life Technologies) and incubated on ice for 15 minutes. The cells were then heat shocked at 42°C for 30 seconds, placed on ice, and mixed with 250ul SOC (Invitrogen 15544-034). Transformations were shaken at 37°C for 1 hour, then 100-250ul was plated onto a LB

agar plate containing 50ug/ml kanamycin. The following day, colonies were picked from each plate using a sterile cocktail stick which was placed into a pop-cap tube containing 3-5ml of Lysogeny broth (LB) containing 100ug/ml kanamycin or ampicillin. Tubes were shaken at 37°C overnight and the following day, were centrifuged at 4000RPM for 5 minutes. The plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen 27106) according to manufacturer's guidelines and quantified using a NanoDrop. Each clone was then sequence verified using an overnight sequencing service (Source Bioscience).

To obtain a large amount of plasmid DNA a midi-prep was sometimes carried out. This was done by taking 100ul from a 3ml pop-cap tube of LB inoculate which had been shaken overnight at 37°C. The inoculate was then transferred into a 250ul flask containing 50ul LB with 100ug/ml kanamycin or ampicillin and shaken overnight at 37°C. Plasmid DNA was isolated using a HiSpeed Plasmid Midi Kit (Qiagen 12643) according to manufacturer's guidelines and quantified using a NanoDrop.

2.3.9 Vector Linearisation

Plasmid DNA was linearised using restriction enzymes (NEB). I searched for restriction sites within the insert sequence using the NEBcutter web tool (nc2.neb.com/NEBcutter2/). The restriction enzymes used are outlined in the table of clones (tables 2.4 and 2.5). Restriction digest reactions were set up and incubated for a minimum of 2 hours at 37°C as follows-

20ug DNA

5ul 10X NEBuffer

3ul Restriction enzyme

Up to 50ul sterile water

Linearised DNA was then purified by adding 200ul TE buffer pH8, 200ul phenol (Life tech) and 200ul chloroform, vortexing thoroughly, and centrifuging at 13000RPM for 1 minute. The upper phase was transferred to an RNase-free tube with 30ul ammonium acetate (AppliChem A4716,0250), 3ul MgCl₂ and 800ul 100% ethanol. Tubes were placed at -80°C for at least 30 minutes to precipitate DNA. Samples were then centrifuged at 13000RPM for 15 minutes, washed in 500ul 70% ethanol, and centrifuged at 13000RPM for 5 minutes. Samples were air dried for 5 minutes then resuspended in 40ul nuclease-free water. 1ul of RNA from each sample was quantified using a NanoDrop. 1ul was run on a 1.5% agarose gel to check for linearisation.

2.3.10 Antisense RNA probe synthesis

Digoxigenin-labelled antisense RNA probes were synthesised using DIG-RNA Labeling Mix (Roche 11277073910) and T7 Polymerase (Thermo EP0111) or SP6 Polymerase (Thermo EP0131).

The reaction was set up as follows-

10ul 5X Transcription buffer

2ul DIG-RNA labelling Mix

0.5ul RNase Inhibitor, Murine (NEB MO315S)

2ug linearised template DNA

4ul T7/SP6 RNA polymerase

Up to 50ul Nuclease-free water

The reaction was incubated at 37°C for 2-5 hours, then DNase treated with 1ul DNase I (Ambion AM2222) at 37°C for 20 minutes.

RNA was purified by adding 30ul LiCl (Sigma L7026) and precipitating overnight at -20°C. Samples were then centrifuged at 13000RPM at 4°C for 30 minutes, washed in 70% ethanol and centrifuged for 5 minutes at 13000RPM at 4°C. Samples were air dried for 5 minutes then resuspended in 40ul nuclease-free water. 1ul of RNA from each sample was quantified using a NanoDrop. 2ul was run on a 1.5% agarose gel to check for a distinct band which indicates good quality RNA.

2.3.11 NanoString transcript quantification

Probes to target genes were produced and designed by NanoString. For the time-course experiment, 5 embryos were harvested every 15 minutes and RNA was extracted from each sample and diluted to 50ul. Samples were processed by NanoString using the nCounter fluorescence-based probe digital counting system.

2.4 Western blot techniques

2.4.1 Extraction of protein from embryos

Embryos were homogenised by pipetting up and down in PhosphoSafe lysis buffer (Millipore 71296) supplemented with 1 Complete Mini protease inhibitor tablet (Roche 11836153001) per 10ml buffer. Typically for a western blot 40x embryos were homogenised in 250ul buffer and incubated on ice for 10 minutes. 250ul FREON (TCTFE, 1,1,2-trichloro-trifluoroethane) was then added to each tube and vortexed thoroughly. Samples were then centrifuged at 4°C at 13000RPM for 15 minutes. Supernatant was then transferred to a new Eppendorf tube and was either used immediately for western blot or immunoprecipitation or snap frozen in liquid nitrogen and stored at -80°C.

2.4.2 Immunoprecipitation of proteins

Immunoprecipitation (IP) buffer was supplemented with 1 Complete Mini protease tablet per 10ml. Approximately 40 embryos were homogenised in a 2ml tube with 250ul ice cold IP buffer by pipetting up and down. Samples were incubated on ice for 5 minutes, then FREON was added and tubes were vortexed thoroughly. Samples were then centrifuged at 4°C at 13000RPM for 5 minutes. Supernatant was then transferred to a new tube. An input sample of 30ul was taken for western blot and stored at 4°C. 1ug of antibody was added to each tube of protein extract and kept overnight at 4°C on a rotator. For all HA tag immunoprecipitation, HA antibody produced in rabbit was used (Abcam AB9110).

The following day Dynabeads Protein A (Life Technologies 10001D) were washed 3 times in 1ml IP buffer for 15 minutes on a rotator at 4°C and using a magnetic stand to bind beads. 10ul of Dynabeads was added to each protein sample to bind to the antibody. Samples were incubated overnight at 4°C on a rotator.

The following day samples were washed 5 times in 1ml IP buffer for at least 15 minutes on a rotator at 4°C and using a magnetic stand to bind beads. Samples were eluted in 40ul 1X Licor protein sample loading buffer (Licor 928-4004). Samples were denatured by heating to 95°C for 5 minutes and were then immediately loaded onto a western blot gel.

2.4.3 Western Blot

12% western gels were prepared:

Casting gel-

Distilled water- 8.3ml

30% Protogel (National diagnostics EC-890) polyacrylamide 10ml

1.5M Tris pH 8.8 6.3ml

10% SDS 250ul

10% ammonium persulphate 250ul

TEMED 10ul

Stacking gel-

Distilled water- 6.8ml

30% Protogel (National diagnostics EC-890) polyacrylamide 1.7ml

1.5M Tris pH 8.8 1.25ml

10% SDS 100ul

10% ammonium persulphate 100ul

TEMED 10ul

Gels were made by pouring the casting gel mix into a mould consisting of 2 glass plates in a plastic holder. A space of 1cm was left at the top, and water poured gently over gels to remove bubbles. When gels had set the stacking gel was poured on top and the comb inserted to make the wells. Once set, gels could be stored at 4°C for up to two weeks.

Gels were submerged in 1X SDS running buffer in a BioRad tank (Mini-PROTEAN Tetra Cell). Proteins were diluted in 1X Licor protein sample loading buffer (Licor 928-4004) and denatured for 5 minutes at 95°C. 20ul of each protein sample was loaded into a well of the gel, alongside any input controls and 5ul of Novex Sharp Pre-stained Protein Standard ladder (Life technologies LC5800). Gels were run at 100V initially for running through the stacking gel at the top, then the voltage was increased to 140V for the remaining time. Gels were run until the protein stain had visibly run off the bottom of the gel.

Gels were then transferred to a membrane. The Immobilon-FL (Millipore IPFL00010) membrane was immersed in Methanol and the top right corner snipped to orient the membrane in later steps. Two sheets of filter paper and sponges were soaked in ice cold

transfer buffer. The gel and membrane were placed between two sheets of filter paper and two sponges, inside a plastic holder. Bubbles were removed from between the layers by rolling a falcon tube over the filter paper. The holder was then placed inside a BioRad tank filled with ice cold transfer buffer with the gel closest to the negative pole and the membrane towards the positive pole. The blot was transferred at 200mA for 1 hour.

Membranes were then removed and placed in a tray with 10ml blocking buffer (5ml Odyssey Blocking Buffer (PBS) 927-40000 in 5ml PBS) for 1 hour at room temperature on a shaker. The Monoclonal Anti-HA antibody produced in mouse (Sigma H3663) was diluted to 1:1000 in TBST. The blocking buffer was removed from the membrane and replaced with antibody, which was then incubated overnight at 4°C with gentle shaking.

The following day membranes were given 5x 15 minute washes in TBST at room temperature on a shaker. Licor secondary antibodies are light sensitive and were added at a concentration of 1:15000 (Licor IRDye 800CW secondary antibody, goat anti-mouse IgG 926-32210), then incubated for 1 hour in the dark at room temperature. Membranes were given 5x 15 minute washes in TBST at room temperature on a shaker. Membranes were then visualised on the Licor Odyssey gel detection system using the Image Studio v2.1 software.

2.4.4 In vitro protein translation

The TnT[®] SP6 Quick Coupled Transcription/Translation kit (Promega L2080) was used to synthesise proteins directly from plasmid DNA *in vitro*. This technique was used to test the translation blocking activity of morpholinos. Reactions were set up as follows-

10ul TNT Sp6 Quick Master Mix
0.25ul Methionine
200ng Plasmid DNA
1ul Nuclease-free water/ morpholino.

Reactions were incubated at 30°C for 30 minutes. The morpholinos were added at increasing concentrations, translation blocking morpholino 1 was added at a concentration of 40uM, 20uM and 10uM. Translation blocking morpholino 2 was added at a concentration of 40uM, 20uM, 10uM, 5uM, 2.5uM and 1.25uM. These increasingly lowered concentrations of morpholino allowed higher protein production of HA-tagged Mix1 or Mixer protein which was visualised by western blot.

2.5 Data Analysis

2.5.1 Analysis of time-course RNA-seq data

All RNA-seq data was analysed by Nick Owens. To evaluate gene expression abundances, time-series reads were aligned to the *X. tropicalis* v7.2 transcriptome and known off-genome sequences derived from EST assemblies using bowtie2 v2.1.0. Reads from *X. laevis* animal cap sequencing were aligned to the *X. laevis* v1.6 transcriptome using bowtie2 v2.1.0 (Karpinka et al. 2015). Time-series data was normalised for each gene against the total reads in each library to a standard library size of 25 million reads. For cases in *X. laevis* where a homeolog pair were present they were counted together.

To detect differential expression between different time-series conditions, for each gene, data-points were modelled with a Gaussian process which produced median lines of best fit and 95% confidence intervals (Owens 2015). Using the Gaussian process framework, differential expression was determined by hypothesis testing (Owens 2015). We compared the null hypothesis that knockdown and control conditions had equal temporal expression, against the alternative hypothesis that knockdown and control conditions had different temporal expression. Genes for which we prefer the alternative hypothesis were considered for morpholino-induced differential expression.

Different models were tested for each gene in the knockdown and control conditions; a single model which places the knockdown and control data points together and two separate models for each condition. Genes which fit the two model condition were considered for differential expression. Only genes which diverge after the activation time of *mix1* were considered for differential expression for the splice MO and translation MO2 time-series. Genes were considered to be differentially expressed at the “divergence time” at which the Gaussian process models overlap by only 10%. The 30 minute time-point following this divergence was assigned the divergence time for a particular gene. A further condition was that the models must overlap by at least 15% before Mix1 was activated in the longer Mix1 splice blocking MO time-series and translation blocking MO2 time-series. The translation MO1 time-series started after *mix1*-activation so genes were included in analysis if the models of control MO and MO1 overlap by at least 7.5% at the first time point. For each gene, time-series expression profiles were plotted using MATLAB.

Fold changes at 9.0hpf were calculated by taking number of reads from the median line and dividing control MO by Mix1 splice MO. The error bars were then derived from Gaussian process models.

2.5.2 Analysing splicing of RNA-seq reads

This analysis was done by Nick Owens. The sequencing reads from the six time-points from 5.0-7.5 hpf were used in the analysis because *mix1* transcripts are generated at these times and there are nascent un-spliced transcripts present. The reads were aligned against *X. tropicalis* transcript sequences from the v7.1 genome assembly using tophat2 v 2.0.10 (Hellsten et al. 2010). Reads with a section in exon 1 and a section in intron 1 or exon 2 were counted to determine whether they had been spliced. Percentages of spliced transcripts were calculated for each condition and the average percentage of splicing over the 6 time-points was reported.

2.5.3 Detecting gene activations

This analysis was done by Nick Owens. When normalised RNA-seq reads for a particular gene increase by a certain threshold over successive time-points, a gene is classified as “activated”. To indicate gene activation, the gene expression increase must be measured over at least three successive time-points and the fold change from the start of the increase to the end must exceed a threshold (e.g. 5x, 10x, 20x fold change, see text for details) (Collart et al. 2014). An activation point is assigned to the time-point at the start of the rapidly increasing run.

2.5.4 NanoString data analysis

Counts for each time point were normalised to *odc1*. Gene profiles plotted alongside RNA-seq data using MATLAB, this analysis was done by Nick Owens.

2.5.5 Statistics for detecting differential expression in animal cap sequencing data

This analysis was done by Nick Owens. For animal cap data differential expression between WT and *mix1*-expressing animal caps for each gene was established using a likelihood ratio test on read counts to generate p values which were false discovery rate (FDR) corrected (De Domenico et al. 2015). All genes with FDR>0.1 were called as differentially expressed.

2.5.6 Mapping *X. tropicalis* and *X. laevis* genes

This analysis was done by Nick Owens. The genes differentially expressed in *X. laevis mix1*-expressing animal caps were compared to *X. tropicalis* genes by gene name matching and reciprocal blast of sequences. Each blast match with at least 85% identify over 25% of transcript was retained and given an alignment score. Up to four *X. laevis* genes could map to a single *X. tropicalis* gene, but all were required to have an alignment score within 75% of the top alignment. Genes were excluded where an *X. laevis* gene mapped to multiple *X. tropicalis* genes. 23% of relationships assign a single *X. laevis* transcript to a single *X. tropicalis* transcript and 70% of relationships assign two *X. laevis* transcripts to a single *X. tropicalis* transcript.

2.5.7 Gene Ontology Analysis

This analysis was done by Nick Owens. The genes detected in each time-series were assigned GO terms from an annotated gene list generated using blast2go. Fisher exact test was used to calculate p values. I calculated the enrichment of GO terms within the Group 1, 2 or 3 was calculated by dividing the observed proportion of genes annotated to a particular GO term within Group with the proportion of genes annotated to that GO term within the background of all genes detected in the time-series.

2.5.8 Comparing differentially expressed genes from different experimental conditions

Gene lists from different experimental conditions were compared using a web tool for list comparison (Whitehead-Institute 2015). To calculate p values using Fisher's exact test and Chi-square test (both two-tailed) a web tool was used (Graphpad 2015). Chi-square test for comparison of three gene sets was calculated with a web tool (Lowry 2015).

The expected overlap between three lists of genes A, B, C of respective size n_A , n_B , n_C from a total list of genes size n_T was calculated. To calculate the overlap (E) that would be expected if three sets of genes were chosen at random from the total set of genes n_T , e.g. for list A probability $p_A = n_A/n_T$ $E = n_T p_A p_B p_C$.

2.5.9 Transcription factor identification

In order to identify which genes are transcription factors within the lists of differentially expressed genes, a set of transcription factors which had been compiled by Mike Gilchrist was used for list comparison. This set was generated using a program HMMER3 to look for Pfam DNA-binding motifs within *X. tropicalis* genes (Collart et al. 2014).

2.6 Solutions and Reagents

All solutions are dissolved in distilled water.

10X Marc's Modified Ringers (MMR)

NaCl	1M
KCL	20mM
CaCl ₂	20mM
MgSO ₄	10mM
Hepes pH7.8	50mM
EDTA	1mM

Tris-Buffered Saline (TBS)

Tris-HCl pH7.6	50mM
NaCl	150mM

TBST

TBS + Tween 20 (0.2%).

TAE

Tris base	40mM
Glacial acetic acid	20mM
EDTA pH8	1mM

MEMFA

MOPS	1M
EGTA	20mM
MgSO ₄	10mM
Formaldehyde	3.7%
pH7.5	

Phosphate Buffered Saline (PBS)

NaCl	137mM
KCl	2.7mM
Na ₂ HPO ₄	10mM
KH ₂ PO ₄	1.8mM
pH7.4	

PBST

PBS + Tween 20 (0.2%).

Bead Blocking Buffer

PBS +10% BSA

1X Hybridisation buffer

Formamide	50%
SSC	5X
Torula yeast RNA	1mg/ml
Heparin	100ug/ml
Denhart's solution	1X
Tween-20	0.1%
CHAPS	0.1%
EDTA	10mM
Store -20°C	

20X Saline-Sodium Citrate buffer (SSC)

NaCl	17.5%
Sodium citrate	8.8%
pH7	

2X Maleic Acid Buffer (MAB)

Maleic Acid	200mM
NaCl	300mM

1X Alkaline Phosphatase Buffer

Tris	100uM
NaCl	100uM
Tween-20	0.1%
MgCl ₂	40mM

SDS-PAGE Running Buffer

Tris base	25mM
Glycine	192mM
SDS	0.1%

Transfer Buffer

Tris Base	25mM
Glycine	190mM
Methanol	20%

IP Buffer

Tris, pH7.5	50mM
NaCl	150mM
EGTA	1mM
Igepal CA-630	1%
Sodium deoxycholate	0.25%

2.7 Tables

Table 2.1 : Morpholino sequences

Gene	Morpholino	Sequence
<i>Bix1.1</i>	Splice blocking	5'-GATGACATATAAGCACCTACCTGAA-3'
<i>Eomes</i>	Splice blocking	5'-AGTTGGGCTAGTAACCTTACCTGCC-3'
<i>Foxi4.2</i>	Splice blocking	5'-GGGCATTACAGACGCTTACCTGGAT-3'
<i>Grh3</i>	Splice blocking	5'-AATTGCATTATATGCATTACCTGT-3'
<i>Gsc</i>	Splice blocking	5'-TGCGATTTGGTAGCGCTTACCTGTA-3'
<i>Klf17</i>	Splice blocking	5'-TGTAGAAAAGCTGTACTTACCGACA-3'
	Translation blocking	5'-GGGTTGAGAAAGCCACTCATCCT-3'
<i>Lhx5</i>	Splice blocking	5'-AGGCCATGCCAACTCCTTACCTGAA-3'
<i>Mix1</i>	Splice blocking	5'-TCCATGCCAATGTCCCTTACCTCTC-3'
	Splice blocking 5 base mismatch control	5'-TCgATcGgAATcTCCCTTAgCTCTC-3'-3'
<i>Mixer</i>	Splice blocking	5'-GAAATGACTGACTTACCCTTTAAGC-3'
<i>Not</i>	Splice blocking	5'-TACAATTAGGTGCCACTCACCTCTC-3'
	Translation blocking	5'-CAGGAGAGTAGTCCAAGGTCTCACT-3'
<i>Tfap2a</i>	Splice blocking	5'-GATTGAAGGGATTATGCTTACCGAT-3'
<i>Ventx1.2</i>	Splice blocking	5'-AAGAAAGACACATACCTGGTAAGGA-3'
Standard control	N/A	5'-CCTCTTACCTCAGTTACAATTTATA-3'
<i>Mix1</i> and <i>Mixer</i>	Translation blocking MO1	5'-GACTGGCTGCTTCTGGGTTCCCAA-3'
<i>Mix1</i> and <i>Mixer</i>	Translation blocking MO2	5'-GGTCCCAAGGCTTGTGGCTCCCAA-3'

Table 2.2: *X. tropicalis* qPCR Primers

Odc1 Forward	GTTGACCTGCCAGAGCTAC	(Collart et al. 2014)
Odc1 Reverse	CAGGGAGAATGCCATGTTCT	(Collart et al. 2014)
Crx Forward	TACCCAACAACCCACGGAAA	Designed myself
Crx Reverse	CACTCGGGACTCTGGTAGAT	Designed myself
Cer1 Forward	CCCACGCAAAACAAAAGTTCAA	(Collart et al. 2014)
Cer1 Reverse	TGGCACCAGGCTTTTCAGTA	(Collart et al. 2014)

Tnfrsf21 Forward	GACTAACTATGAACCGACAGATG	Designed myself
Tnfrsf21 Reverse	GGGCGTCCTCTTGATGGT	Designed myself
Ppp1r10 Forward	TTCTGTCCAAGTTTATTCGTGTTG	Designed myself
Ppp1r10 Reverse	CCATCCAGTCATTGACCAAA	Designed myself
Ventx3.2 Forward	CTCTCTGAGAATCAGATCAAAACCT	Designed myself
Ventx3.2 Reverse	GACAGGAACTCCGACTGGA	Designed myself
Szl Forward	GCAATGACATCGGCTACTCG	Designed myself
Szl Reverse	CAGGGCTGGATGAACGTATC	Designed myself
Zic3 Forward	TCACATATGCTACTGGGAGGA	Designed myself
Zic3 Reverse	GAATGGCTTCTCACCTGTATGA	Designed myself
Foxc1 Forward	CATCGTTGTGCCATCTGAG	Designed myself
Foxc1 Reverse	CGCCCTGTAGTAACTCTGCT	Designed myself

Table 2.3: *X. laevis* qPCR Primers for animal caps

Gsc-a Forward	GCTGGCAAGGAGAGTTCATC	George Gentsch
Gsc-a Reverse	TGGGCATTTTCTGATTCCTC	George Gentsch
VegT Forward	TTTAGGAACATGCATTCTCTGC	(Gentsch et al. 2013)
VegT Reverse	CAGTGTTGGGCAGGTAGAGG	(Gentsch et al. 2013)

Table 2.4: Cloning primers for Mix1 targets for use making *in situ* hybridisation probes

Gene	Vector	Forward primer	Reverse primer	Linearisation site	Antisense promoter
crx	Zero blunt TOPO	TGGCGTGGATCT CAGACTTT	TCCAGGAAGCAG TTTGGTCT	XhoI	SP6
cer1	Zero blunt TOPO	ATGTTACTCTGT GTACTTA	AACTGTAATTTTA GTCAA	NotI	SP6
foxc1	pENTR D-TOPO/ pCS2+	CACCAACGCCCC TGAAAAGAAGA T	GCAGAGGTGGTG GTGGTAGT	Apal	T7

Table 2.5: Cloning primers for expression of genes of interest

Gene	Vector	Forward primer	Reverse primer	Linearisation site	Sense promoter
Mix1 with stop codon	pENTR D-TOPO/ pCS2+	CACCATGGAC TCATTCAGCCA ACAAC	TCAAAGGTGG AGGAGCA	Apal	SP6
Mix1 5'UTR no stop codon	pENTR D-TOPO/ pCS2+	CACCTTGGGA ACCCAGGAAG CAGC	AAGGTGGAGG AGCACACAGA	Apal	SP6
Mixer 5'UTR no stop codon	pENTR D-TOPO/ pCS2+	CACCTCACTT TGGGAGCCA CAAGC	GGCAGAATA GATGGACTCT GGACA	Apal	SP6

Chapter 3: Mechanisms of gene activation

3.1 Introduction

3.1.1 Background and aims

Two distinct early waves of gene activation are found in the early *Xenopus tropicalis* embryo, the first wave begins immediately after fertilisation and the second begins shortly before MBT. The MBT occurs after the first 12 rapid cell divisions. These waves of activation were identified using polyA+ RNA-seq to measure gene expression in clutches of embryos sampled over a time-series of the first 9.5 hours of development (Collart et al. 2014). Here I set out to distinguish between polyadenylation of maternal mRNAs and *de novo* zygotic transcription. My work is based on polyA+ time-series data and genes activated in this dataset; I will refer to this polyA+ time-series as the original time-series (Collart et al. 2014). Using the original time series data alone, it is not possible to differentiate between transcript accumulation through *de novo* zygotic transcription and polyadenylation of existing maternally deposited transcripts. Both mechanisms of activation appear the same as increases in the polyA+ RNA-seq data.

The aim of this work was to measure mRNA without polyA+ selection to measure time-course gene expression in order to differentiate between these two classes of gene activation behaviour. Two techniques were applied; NanoString which uses hybridisation probes to quantify specific transcript numbers and Ribo-Zero RNA-seq which is used to measure relative global transcript abundance, independent of polyadenylation status. In this chapter I report two mechanisms of activation: polyadenylation of maternal transcripts dominant up to the 64-cell stage and transcription dominant from the 128-256 cell stage.

3.1.2 Gene activation and gene expression profiles

Here I use the term “gene activation” when expression levels of a particular transcript increase significantly over at least three consecutive time points at 30 minute intervals (see Materials and Methods for details)(Collart et al. 2014). Gene activations are either the rapid accumulation of new transcripts or rapid polyadenylation of existing maternal mRNA within embryos. In some cases, genes are activated by both of these mechanisms. The overall number of gene activation events for each time point was plotted to show the two distinct early waves of activation in the original time-series. In the first wave of activation 551 genes

were activated and in the second wave of activation 409 genes were activated (Collart et al. 2014).

3.1.3 Polyadenylation-independent transcript quantification

Here I use two different techniques to differentiate between transcription and polyadenylation based gene activation. The NanoString nCounter system is a highly sensitive technique to quantify specific transcript numbers, independently of polyadenylation state, within cell lysate or extracted RNA samples. The technique uses two probes designed to bind to a specific mRNA; a capture probe comprised of a complementary binding sequence alongside a biotin affinity tag, and a reporter probe containing a complementary binding sequence alongside a series of sequences encoding fluorophores. The differently coloured fluorophores are arranged in a unique order in each mRNA reporter probe to allow identification of transcripts. RNA or cell lysates are mixed with the capture probes and reporter probes and pulled down using the biotin affinity tag onto an electrically charged streptavidin surface to elongate the mRNA strands. The unique fluorescent tags are then used to count transcripts in the sample (Geiss et al. 2008). The technique does not require any amplification steps and is therefore not subject to 3' bias and can be carried out on low quantities of mRNA (Geiss et al. 2008).

Alternatively, one can use Ribo-Zero RNA-seq to quantify overall transcript numbers independently of polyadenylation state. Ribosomal RNA (rRNA) makes up a large component of the RNA within a cell so must be removed to allow efficient sequencing of ribosomal-depleted RNA. For Ribo-Zero library preparation, Ribo-Zero beads bind to rRNA and facilitate its removal. This allows RNA-seq of ribosomal-depleted RNA rather than selecting for polyA+ transcripts with oligo(dT)beads (Sooknanan, Pease, and Doyle 2010).

3.2 Identifying mechanisms of gene activation using NanoString technology

In order to investigate the mechanism of activation for a set of genes, I used the nCounter NanoString system to quantify transcript abundance of 43 genes. These 43 genes were selected with the aim to investigate the mechanisms of activation of genes which were some of the latest maternal transcripts to be polyadenylated and the earliest mRNAs to be transcribed. Therefore, genes were selected from within a few hours of when the earliest transcription has been detected previously (Skirkanich et al. 2011). These genes were distributed into two subsets based on their mechanism of activation; 22 genes were activated

before 3.0 hpf and are in the early onset category and 21 genes were activated at 3.0 hpf or later in the original time-series and are in the late onset category (Fig. 3.1).

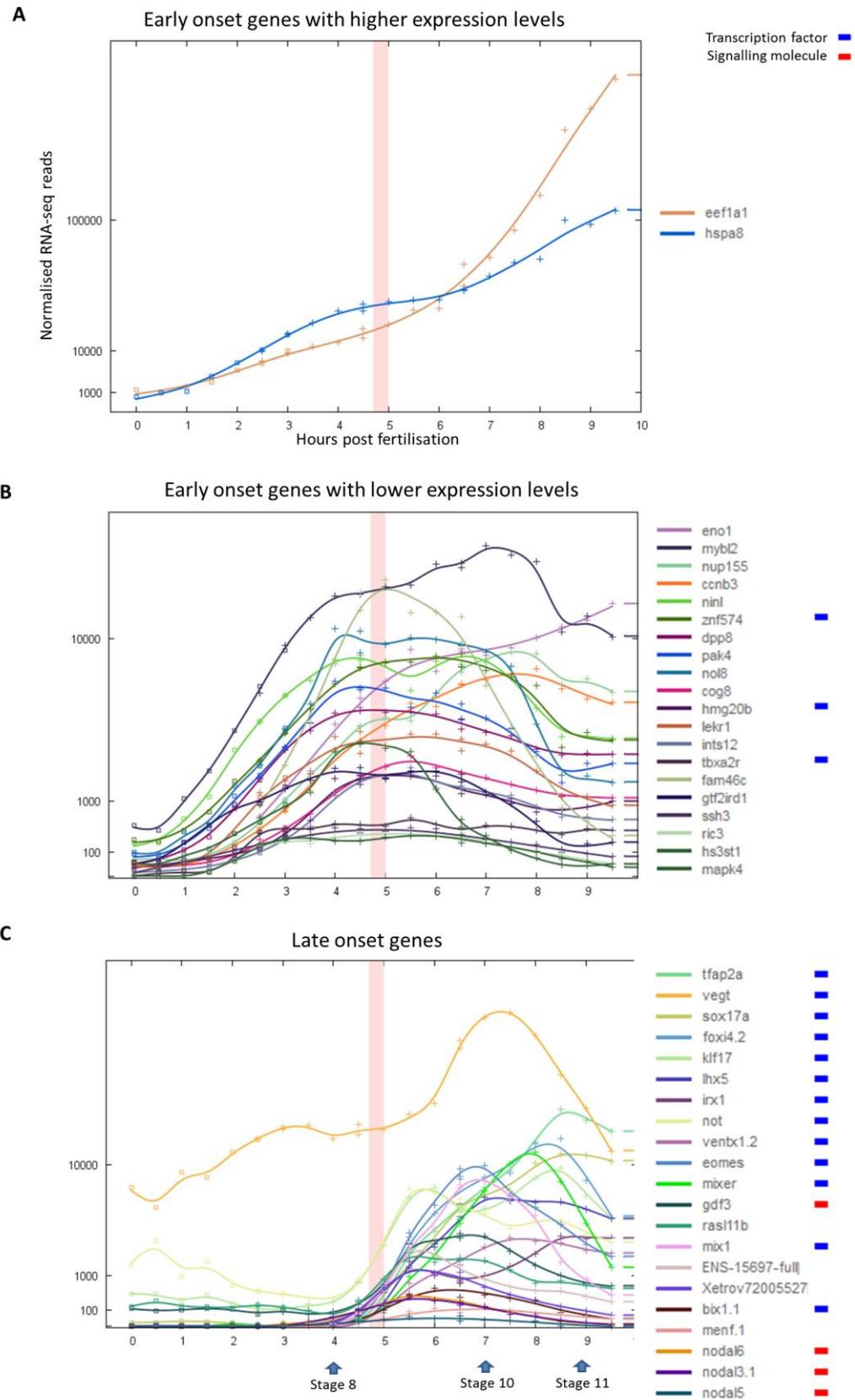


Figure 3.1. Early and late onset gene expression profiles. Time-series gene expression profiles from the original RNA-seq time-series. Transcription factors have a blue tag and signalling molecules have a red tag. (A) Early onset genes with higher expression levels

targeted in the NanoString experiment. (B) Early onset genes with lower expression levels targeted in the NanoString experiment. (C) Late onset genes targeted in the NanoString experiment. (Collart et al. 2014). NF stages are marked below with blue arrows.

Batches of five embryos were sampled every 15 minutes over the first 8 hours of development and extracted total RNA was assayed for expression of the 43 selected genes. Using NanoString, the normalised transcript counts for each gene were plotted and compared to the original RNA-seq time-series data to identify whether each gene is polyadenylated or transcribed. The timing of the cleavage cycles in the clutch of embryos used for collection was 20 minutes for the original time-series and 21 minutes for the NanoString time-series. The horizontal timescale on the original time-series data was therefore corrected by 20/21 to align the two data-sets. This procedure produced excellent agreement between the NanoString and RNA-seq data (Fig. 3.2). The relative abundance of transcripts is on an arbitrary vertical scale and was adjusted for the purposes of visualisation of the NanoString and RNA-seq data together (data was corrected and profiles plotted by Nick Owens, see Materials and Methods for detail).

Transcripts which display an onset in the original polyA+ RNA-seq profile but have no onset in the NanoString data are maternal transcripts undergoing rapid polyadenylation (Fig. 3.2A). Transcripts which display an onset in both the original time-series and in the NanoString data are zygotically transcribed in the early embryo (Fig. 3.2B). RNA-seq read counts from the original time-series and NanoString expression counts are shown for four polyadenylated genes (Fig. 3.2A) and four transcribed genes (Fig. 3.2B).

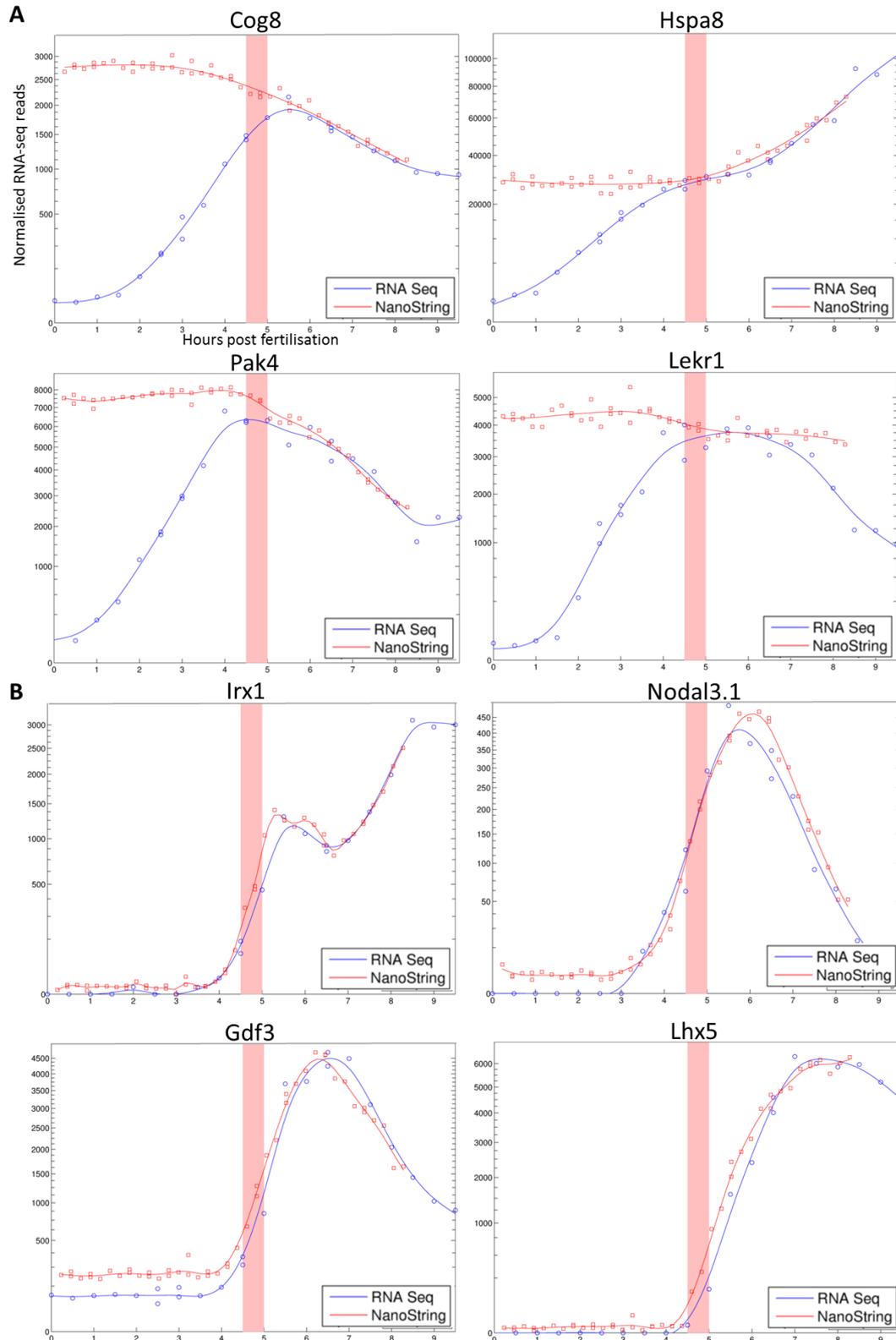


Figure 3.2. NanoString time-series data. Time-series NanoString data (red points) is overlaid on RNA-seq expression profiles from the original time-series (blue points). NanoString horizontal axis is adjusted 20/21 to align time-series, NanoString vertical axis is manually adjusted to give the best fit for each gene. (A) Examples of genes activated by

polyadenylation of maternal mRNA indicated by indicated by a relatively high, constant level of transcripts detected in the early NanoString data points and an early increase in original time-series polyA+ transcripts. *Cog8, hspa8, pak4* and *lekr1* are shown as examples. (B) Examples of genes activated by zygotic transcription shortly before MBT, indicated by an increase in both Nanostring and original time-series polyA+ transcripts at or shortly before MBT. *Irx1, nodal3.1, gdf3* and *lhx5* are shown as examples.

I found that the behaviour of the 22 early onset genes was consistent with polyadenylation and the behaviour of the 21 later onset genes was consistent with zygotic transcription. This demonstrates the existence of 2 classes of activation behaviour in early development. These two gene behaviours are temporally distinct, at least for the 41 genes tested, with maternal polyadenylation dominant in the post-fertilisation embryo up to and including 2.5 hpf and zygotic transcription dominant from 3.0 hpf which coincides with the 128-cell stage. This finding is supported by the gap found between the two separate waves of gene activation in the early embryo (Collart et al. 2014).

3.3 Identifying global activation using Ribo-Zero RNA-seq

3.3.1 Polyadenylation is the initial mechanism of gene activation and transcription is dominant shortly before MBT

Ribo-Zero RNA-seq can be used to measure global expression in ribosomal-RNA depleted RNA samples. I used Ribo-Zero RNA-seq to build on the preliminary results obtained using NanoString, to investigate the mechanisms of activation for all genes detected in the early embryo. By measuring gene expression over a time-course I was able to differentiate between transcription and polyadenylation; the two mechanisms of gene activation found in the early embryo.

Embryos were collected over a time-series to measure the dynamic gene expression changes during early development. Pools of five embryos were sampled at 30-minute intervals and total RNA was extracted. The samples were then split into two technical replicates to be used for polyA+ and Ribo-Zero RNA-seq. Ribo-Zero RNA-seq libraries were sequenced for every 1 hour time point up to 8.0 hpf (stage 10), and polyA+ RNA-seq libraries were sequenced for every 30 minute time point up to 9.0 hpf (stage 11). PolyA+ and ribosomal-depleted RNA profiles were generated from normalised read counts and used to compare readouts between the two conditions (RNA-seq data analysed by Nick Owens, see Materials and

Methods for details). Full lists of activated genes from RiboZero and polyA+ RNA-seq can be found in the appendix on the attached CD (tab 1).

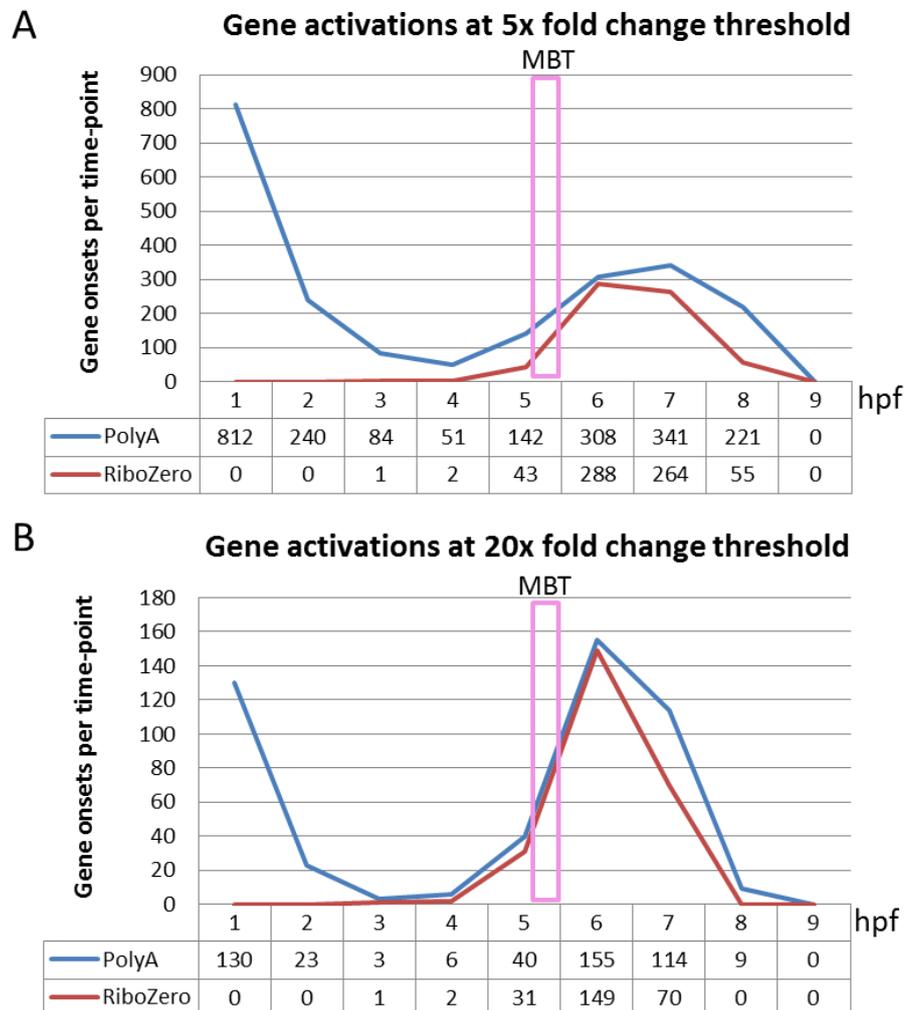


Figure 3.3. All activated genes in the polyA+ and Ribo-Zero RNA-seq time-series. Total genes that display an onset (activation) at each one hour time point in the polyA+ time-series (blue) and the RiboZero time-series (red). The mid-blastula transition (MBT) is marked as a vertical pink bar. (A) Total activations at a 5x fold change threshold. (B) Total activations at a 20x fold change threshold. Tables show total numbers of activated genes at each time point.

I took the activated genes in the RNA-seq polyA+ time-series and compared their expression profiles to those in the ribosomal-depleted RNA time-series (Fig. 3.3). There is a clear early wave of activations immediately from fertilisation in which several hundred genes are activated by polyadenylation. After 4.0 hpf there is an increase in the numbers of activated genes in the ribosomal depleted RNA, indicative of transcriptional activation. Applying different fold change thresholds alters the number of gene activations measured. At a 5x fold

change threshold there are continual gene activations throughout the time series, but still I find two distinct activation waves (Fig. 3.3A). At a 20x fold change threshold, there is a clearer separation between the two peaks of activation (Fig. 3.3B), which suggests there is a distinct early wave of polyadenylation and a second wave where transcriptional activation is dominant.

The comparison of the polyA+ and ribosomal-depleted RNA datasets revealed that genes were activated by polyadenylation after fertilisation, because the vast majority of activations are detected only in the polyA+ data at the earliest time points. Transcription of very few genes was detected from 3.0 hpf, approximately 2.0 hours before MBT (Fig. 3.3). From 4.0 hpf I found that the number of activations in the RiboZero and polyA+ time-series matched more closely. This indicates that from 4.0 hpf which is approximately one hour before MBT, genes were predominantly activated by zygotic transcription. In the clutch of embryos used to generate the RiboZero and accompanying polyA+ sequencing data, the 256-cell stage was at 4.0hpf. Overall low levels of transcription were detected here from the 64-cell stage, and widespread transcription was found from the 256-cell stage. This demonstrates a mechanistic shift where before 4.0 hpf, polyadenylation is the predominant mechanism of activation and from the 256-cell stage at 4.0 hpf transcription is the predominant mechanism of gene activation. PolyA+ and Ribo-Zero normalised read counts are shown for four polyadenylated genes and four transcribed genes (Fig. 3.4). The data points for the RiboZero and polyA+ sequencing do not align well due to differences in relative normalisation and biases of the two different sequencing techniques used.

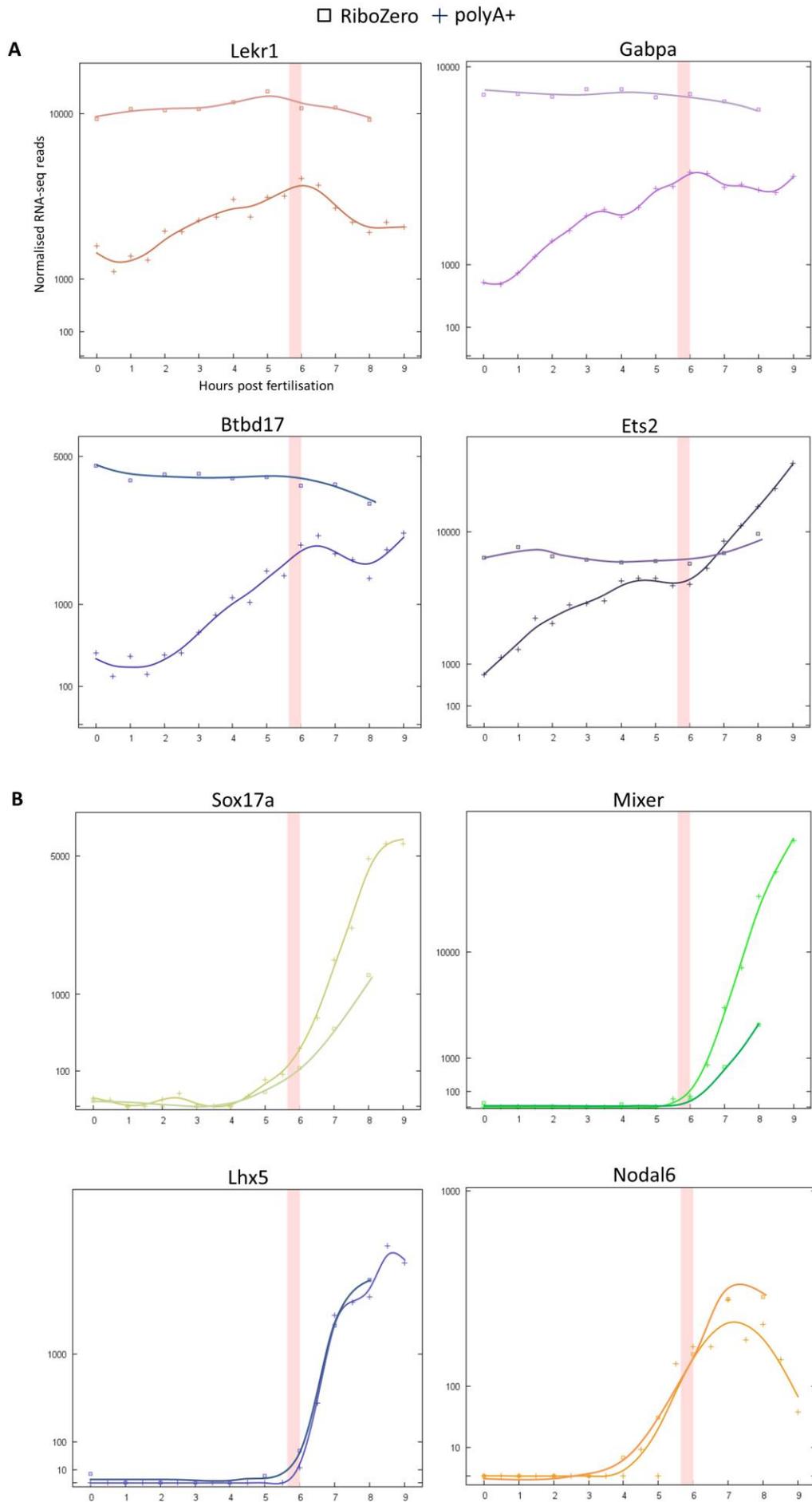


Figure 3.4. Ribo-Zero and polyA+ RNA-seq time-series data. (A) Genes activated by polyadenylation of maternal mRNA indicated by a relatively high, constant level of transcripts detected in the early RiboZero data points (squares) and an early increase in polyA+ transcripts (crosses), *lekr1*, *gabpa*, *btbd17* and *ets2* are shown as examples. (B) Genes activated by zygotic transcription indicated by an increase in both RiboZero (squares) and polyA+ transcripts (crosses) at or shortly before MBT, *sox17a*, *mixer*, *lhx5* and *nodal6* are shown as examples. Gene profiles generated by Nick Owens.

3.3.2 Timing of gene activations are later than in the original time-series

The temperature in the room used to collect embryos was lower than usual, causing a slower developmental rate in the clutch of embryos used for the Ribo-Zero time-series than in the original time-series and the NanoString time-series. I compared the activation times of several genes between my polyA+ time-series and the original polyA+ time-series. I found that *nodal3.1* and *sox2* were activated 1 hour later in my time-series (Fig. 3.5A, B), and *nodal6* and *foxb1* were activated 1.5 hours later in my time-series (Fig. 3.5C, D). As a result of this, the timing of the switch between polyadenylation and transcriptional activation of genes is later in the Ribo-Zero time-series compared to the original time-series and the NanoString time-series.

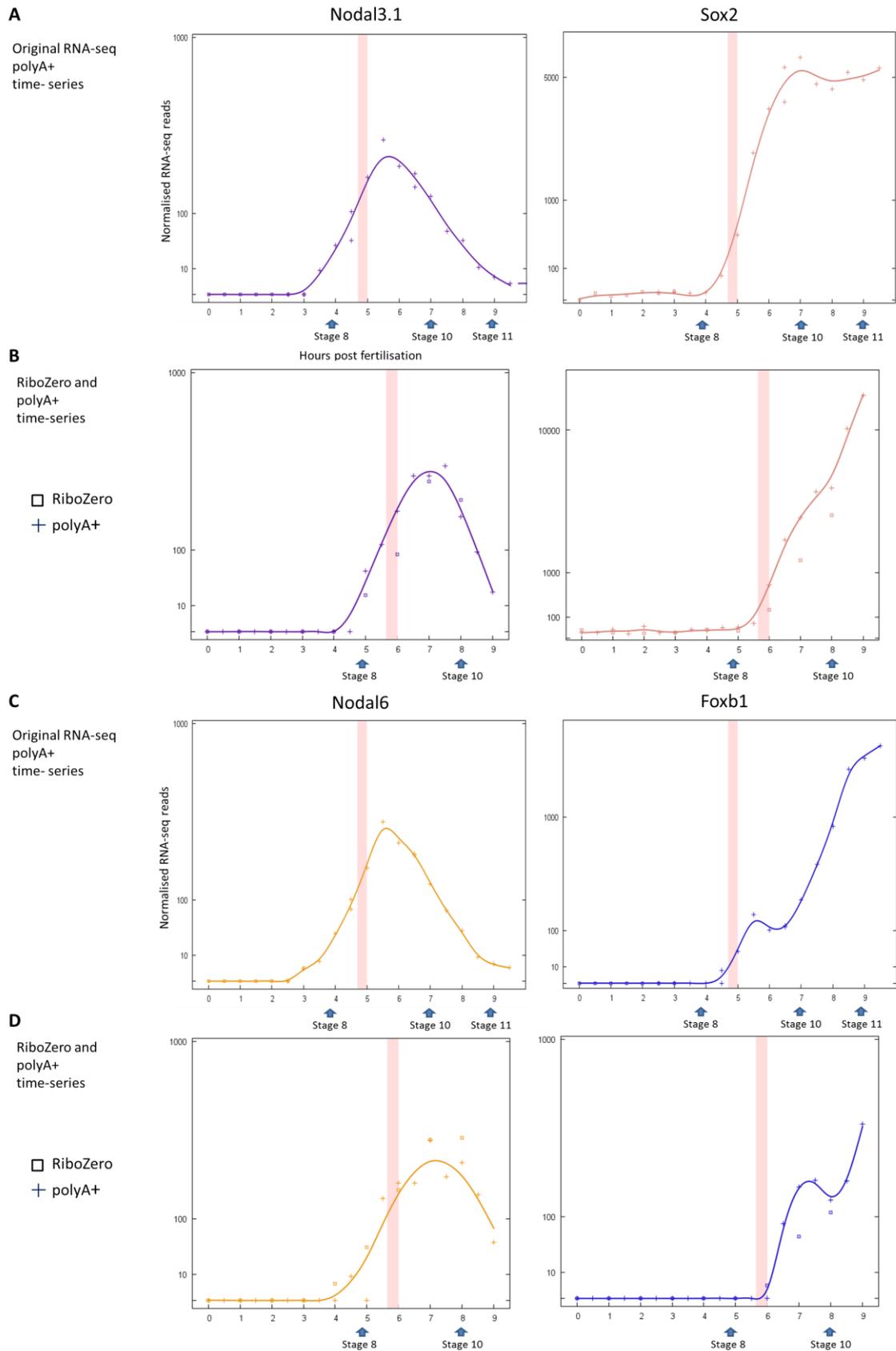


Figure 3.5. Time shift of gene activations between the original and the Ribo-Zero time-series data (Collart et al. 2014) (A) Gene expression profiles showing activation in the

original time-series, *nodal3.1* and *sox2* are shown as examples. (B) Gene expression profiles showing activation in the Ribo-Zero and polyA+ time-series with a 1 hour developmental delay compared to the original time-series, *nodal3.1* and *sox2* are shown as examples. (C) Gene expression profiles showing activation in the original time-series, *nodal6* and *foxb1* are shown as examples. (D) Gene expression profiles showing activation in the Ribo-Zero and polyA+ time-series with a 1.5 hour developmental delay compared to the original time-series, *nodal6* and *foxb1* are shown as examples. Gene profiles generated by Nick Owens.

3.4. Discussion

3.4.1 Transcription becomes the dominant mechanism of gene activation shortly before MBT

I differentiate between 2 mechanisms of gene activation behaviour through comparison of polyA+ and ribosomal-depleted RNA-seq time-series data. The two mechanisms of gene activation in the early embryo; polyadenylation and transcription are easily delineated through comparison of the polyA+ and ribosomal-depleted RNA data. Ribosomal-depleted RNA data reflects changes in total transcript levels and polyA+ data reflects changes in numbers of polyadenylated transcripts.

The degradation of transcripts is established at MBT, before these stable un-adenylated transcripts can persist in the embryo and polyadenylation is the mechanism which regulates translation (Paris and Philippe 1990; Voeltz and Steitz 1998). Therefore, it is unlikely that polyadenylation operates as a mechanism of activation after MBT as polyadenylation is then necessary to protect mRNAs from degradation.

In the NanoString data which tested a small set of genes, I found that polyadenylation was the early mechanism of activation, and transcription became the predominant mechanism of activation shortly before MBT. The Ribo-Zero data gave the same conclusion for genome wide expression within the time-series. I observed a general mechanistic shift from polyadenylation to transcription as the predominant mechanism of gene activation over the first several hours of development. Appreciable numbers of genes were activated by transcription from the 256-cell stage in the Ribo-Zero time-series. Comparison of genes activated in the original time-series and in the RiboZero time-series described here, revealed that transcription was widespread in the second wave of activation found from the 128-256 cell cycles (Collart et al. 2014).

It is difficult to determine whether the two mechanisms of activation are temporally distinct. A period of low activation activity was found at the 3.0 and 4.0 hpf time-points which could be caused by variation in the developmental rate of pooled embryos, or reflects that both mechanisms occur simultaneously in the pre-MBT embryo. It is difficult to tell how much overlap there is between the two mechanisms of activation once transcription has begun, but as mentioned it is unlikely that polyadenylation is a widespread activatory mechanism post-MBT. At all time-points, there were higher numbers of activated genes in the polyA+ time-series, this could reflect late polyadenylation for a number of transcripts or this could be due to differing detection rates in the two sequencing techniques. The fold change criteria are subjective, so there is no way to tell whether a gene is “activated” or not in an unbiased objective manner. For this reason, I cannot determine whether the mechanisms of activation are separate, but I find a general mechanistic switch from polyadenylation to transcription that occurs before the MBT.

A limitation of the Ribo-Zero time-series approach to identify mechanisms of gene activation is that maternal-zygotic genes which are replaced by zygotic transcription as the maternal mRNAs are degraded, cannot be detected. This behaviour would not appear as gene activation in my data if the maternal transcripts are gradually replaced. Tackling this problem, zygotic expression was identified in zebrafish embryos by crossing two strains and differentiating between maternal and paternal SNPs to identify maternal and zygotic mRNAs in the sequencing reads. The first zygotic transcription was identified at the 128-cell stage (Harvey et al. 2013).

3.4.2 Early polyadenylation is necessary for progression through MBT

In zebrafish, by blocking polyadenylation using cordycepin, maternal polyadenylation was found to be essential for normal progression through MBT. The cordycepin-induced delay in MBT was measured by the delayed epiboly and death before 10 hpf in treated embryos. However to determine whether cordycepin treatment blocks zygotic transcription, measurement of the transcriptional output of MBT-transcribed genes is necessary (Aanes et al. 2011). This was addressed in cordycepin-treated *X. tropicalis* embryos using NanoString to demonstrate reduced transcription in the majority of genes tested. This indicates that polyadenylation of maternal transcripts is necessary for adequate transcription at MBT (Collart et al. 2014).

3.4.3 Detection of early transcription

Work in *X. laevis* measured incorporation of a radioactive label to detect transcription from as early as the 128-cell stage (Kimelman, Kirschner, and Scherson 1987). This finding is consistent with our finding of genes which are transcriptionally activated from 128-256 cell cycles in the original time-series. Consistent with this, the earliest zygotic transcription in zebrafish was found at the 128-cell stage. Gene expression in zebrafish embryos was measured using RNA-seq and qPCR with random hexamer or oligo-dT primers at several developmental stages to differentiate between maternal and zygotic transcripts (Aanes et al. 2011). This may reflect that radioactive labelling or qPCR are more sensitive than RNA-seq for identifying early transcription where relative transcript abundance may be very low.

In *X. laevis*, *nodal5* and *nodal6* were found to be transcribed from the 256-cell stage using reverse transcription of RNA with either oligo-dT or random hexamer primers followed by RT-PCR (Yang et al. 2002). I find that these two genes were early transcribed genes in the NanoString and polyA+ time-series. In the original time-series, *nodal5* has too weak an onset to be considered as activation by our criteria, but transcripts were present from 4.0 hpf, so 3.5 hpf would be the activation time if the onset of *nodal5* was more rapid. *Nodal6* was activated at 3.5 hpf which coincides with the 256-stage (Collart et al. 2014). A later study detected expression of *nodal5* and *nodal6* in *X. laevis* from the 128-cell stage and pre-MBT expression of *bix4*, *gdf3*, *sox17a* and *mixer* emerging between the 4-cell and 256-cell stages. All six of these early-transcribed genes were regulated by *VegT* (Skirkanich et al. 2011). *Bix4* is not present in *X. tropicalis* but the related gene *bix1.1* is found. *Bix1.1*, *mixer*, *sox17a*, *gdf3*, *nodal5* and *nodal6* were all found to be transcribed pre-MBT in the Ribo-Zero time-series.

A study in *X. tropicalis* identified potential pre-MBT transcribed genes which displayed a two-fold increase in sequencing reads from the 2-cell to the 32-cell stage (Tan, Au, Yablonovitch, et al. 2013). They report 36 pre-MBT rapidly transcribed genes, 13 of which were validated by qPCR with random hexamer primers for reverse transcription to detect total RNA. Less stringent criteria were then applied for gene expression levels and 303 genes were reported as likely to be early transcribed, and 6 of these were validated by qPCR (Tan, Au, Yablonovitch, et al. 2013). The Ribo-Zero data here clearly shows that all 19 validated genes described as pre-MBT transcribed genes were maternal transcripts activated by polyadenylation in the early embryo (*cdc14b*, *dpp8*, *btbd17*, *cdk5r2*, *arih2*, *slc9a3r1*, *Xetrov72024821*, *mtx1*, *slain2*, *lrrcc1*, *gart*, *oct25*, *ets2*, *akt2*, *c20orf72*, *nkrf*, *rhou*, *spats2* and *ier5*). Two of these genes, *ets2*

and *oct25* were activated first by polyadenylation and then were clearly transcriptionally activated after the MBT. It is unclear how Tan et al. find increases in total RNA levels where I find unchanging transcript levels. Generally, the pre-MBT fold changes in the qPCR data are low; between 2- and 3-fold, which is much lower than the fold changes found for the confirmed transcriptional targets *Nodal5* and *Brachyury*, so there is some doubt that these low fold changes reflect genuine changes in transcript levels (Tan, Au, Yablonovitch, et al. 2013).

Time-course Ribo-Zero RNA-seq in *X. tropicalis* used RNA spikes for the absolute quantification of transcripts within the embryo. The earliest mRNAs detected were *Nodal3.1*, *Nodal5*, *Nodal6*, *Sia1* and *Sia2* from the 32-256 cell stages. The first of these to be detected was *Nodal3.1* which appears at the 32-64 cell stage at 2.5 hpf. These genes were found to be transcribed later than this in our datasets, which could mean (i) variation in timing of transcription of specific genes exists between different clutches of embryos, or (ii) the Ribo-Zero RNA-seq recovered more non-ribosomal reads than my RiboZero sequencing resulting in higher sensitivity (Owens 2015), or (iii) the intrinsic technical variation in each time-point sample means that a gene may be detectable at earlier or later time points.

The original time-series, the Ribo-Zero time-series presented here and the Ribo-Zero time-series sequenced by Owens et al. all have different genes emerging as the first transcribed. The order of early transcriptional events has some variation, which could have a technical or biological origin. Although the exact ordering of the activation time differs, genes are activated within a relatively close period of time in different experiments.

The earliest transcripts in zebrafish, *Drosophila* and mouse are short and have few or no introns (Heyn et al. 2014). In *X. tropicalis*, transcripts of the early expressed genes *nodal5* and *nodal6* have relatively short transcripts of less than 2kb and two introns indicating that early transcribed genes in *X. tropicalis* may also be short. The first zygotic gene expression in fish, fly and mouse is not evolutionarily conserved indicating the existence of adaptive potential (Heyn et al. 2014). If the first transcription can vary so much between species, then perhaps early gene expression is not crucial in establishing gene pathways which direct patterning and embryogenesis. The maternally deposited proteins and transcripts may be sufficient to set up early networks. This poses the question of whether the specific genes which are the first to be transcribed are important. *Xenopus* may have evolved to have highly expanded

and diverged gene families responsible for patterning, giving rise to high levels of redundancy within the system. This ensures a highly robust organism which can develop even if something goes wrong and this may have driven the rapid evolution of *Xenopus* (Heyn et al. 2014).

3.4.4 Conclusion

In summary, I report the same finding for the Ribo-Zero and NanoString time-series data, which is that there are two mechanisms of gene activation in the early embryo, genes are activated by polyadenylation in the post-fertilisation embryo and shortly before MBT, transcription becomes the dominant mechanism. There is a rapid increase in the amount of transcription detected as the control of the embryo shifts from maternal to zygotic.

Chapter 4: Morpholino screen of early-activated genes

4.1 Introduction

As a starting point for the investigation of the gene regulatory networks directing early embryonic development, I needed to identify a small number of developmentally important transcription factors. These would be used later in the project to determine their downstream targets through perturbation analysis. In this chapter I describe the use of a preliminary morpholino oligonucleotide (MO) knockdown screen on a larger set of candidate target genes derived from data generated in the lab (Collart et al. 2014) to achieve this. I reasoned that genes whose depletion caused early and obvious developmental defects would likely be important regulators of early development. After surveying twelve zygotically activated and ten rapidly polyadenylated maternal transcription factors, I selected *mix1* for further investigation.

4.2 Screening the candidate zygotic transcription factors

4.2.1 Choice of genes

To select the candidate gene list for the initial morpholino oligonucleotide (MO) knockdown screen of zygotically activated transcription factors, I decided on a set of attributes which would suggest genes likely to be important regulators of early developmental networks. These were; clear zygotic activation shortly before MBT, magnitude of activation and transient behaviour showing decline shortly after activation. I was particularly interested in genes showing transient behaviour, as these were suggestive of a network controlling genes with an on-off, switch-like behaviour, whose functional importance may be limited to an interestingly specific period of development.

I used two sets of existing data to help make my selection: a set of high resolution time-series gene expression data covering the period from fertilisation through to mid-gastrulation (Collart et al. 2014), and a database of assembled expressed sequence tag (EST) data (Gilchrist et al. 2004). I used EST database to check the quality of each gene model by ensuring there was alignment of multiple clone sequences and orthology data along the entire open reading frame of each gene (Gilchrist et al. 2004). Splice junctions were identified by aligning cDNA sequences to genomic sequence using the UCSC genome browser (UCSC 2015). I used the activation times derived from the time-series data to find genes strongly,

and preferably transiently, activated at or around the MBT and then checked the gene models against the assembled EST data, before settling on a list of 12 candidate transcription factors: *bix1.1*, *eomes*, *foxi4.2*, *grhl3*, *gsc*, *klf17*, *lhx5*, *mix1*, *mixer*, *not*, *tfap2a* and *ventx1.2* for the MO screen (Fig. 4.1). The time-series data also showed which genes were present as maternal transcripts before zygotic gene activation, and which were not.

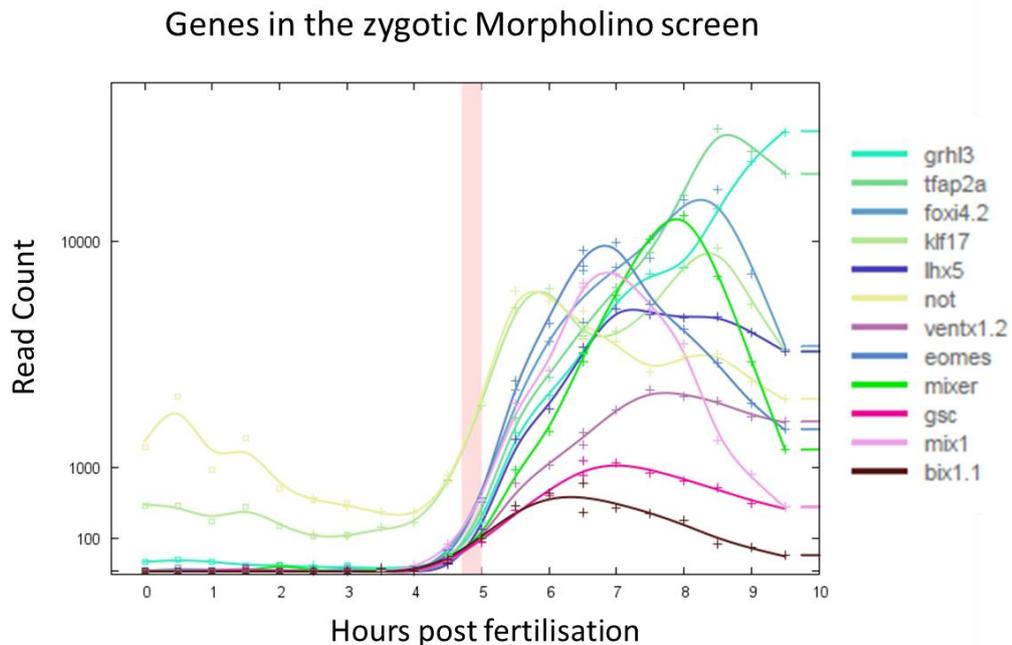


Figure 4.1 Zygotic genes for MO screen. Gene expression profiles from the original RNA-seq time-series. Normalised read counts are shown for the twelve zygotic transcription factors included in the first MO screen for phenotypes. All twelve genes are rapidly and transiently expressed from shortly before the MBT which is marked with a pink bar (Collart et al. 2014).

I used different MO depletion strategies, depending on whether genes had detectable maternal mRNAs prior to zygotic activation. For genes with a significant maternal mRNA component (*klf17* and *not*) I used both a splice blocking MO and a translation blocking MO in separate experiments. For genes with little or no detectable maternal mRNA (<1% of post-MBT levels of zygotic activation) I used only a splice blocking MO. See Table 4.1 for the full list of candidate genes and MO types used. The main reason for using splice blocking MOs was because they would not block injected mRNA used for rescue of the MO induced phenotype, with future validations for selected target genes in mind. In all cases I injected 10ng of MO; and as a negative control I injected the standard control MO from Gene Tools

at 20ng. The control was injected at a higher dose as this was thought to be a robust strategy, however since the Gene tools standard control MO is known to cause no defects, the higher dose would not make a difference as a 10ng would have been sufficient.

4.2.2 Phenotypes arising from the screen

MO knockdown of seven out of twelve genes induced a defective phenotype in injected embryos. The phenotypes ranged from minor defects such as a bent tail, to major defects such as blastopore closure failure (Fig. 4.2). Phenotypes are listed in Table 4.1.

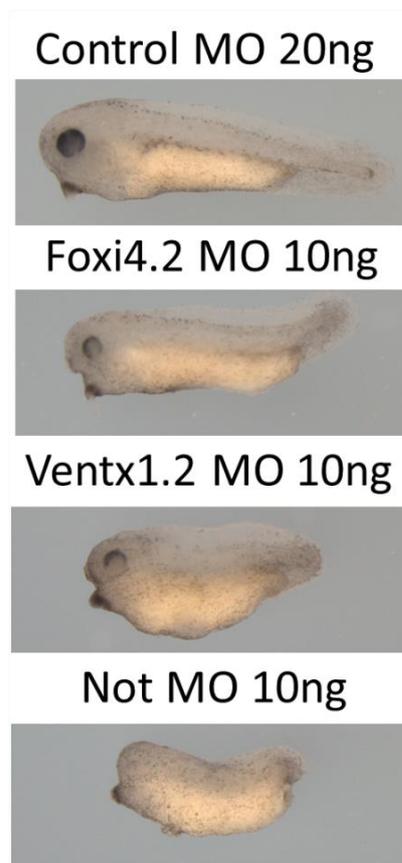


Figure 4.2. Phenotypes from the zygotic MO screen. Embryos injected with 20ng standard control MO exhibit a normal phenotype. Embryos injected with 10ng of Foxi4.2 splice blocking MO have a bent tail (31/46), embryos injected with 10ng Ventx1.2 splice blocking MO are ventralised (36/47) and embryos injected with 10ng Not translation blocking MO have a truncated axis and head and tail defects (24/53). Embryos were imaged at stage 33, lateral view.

Gene	Morpholino	Knockdown phenotype	Penetrance
<i>bix1.1</i>	Splice blocking	No defect	N/A
<i>eomes</i>	Splice blocking	No defect	N/A
<i>foxi4.2</i>	Splice blocking	Bent tail	67% (31/46)
<i>grhl3</i>	Splice blocking	No defect	N/A
<i>gsc</i>	Splice blocking	No defect	N/A
<i>klf17</i>	Splice blocking	Truncated	40% (16/40)
		Cranial or ventral oedema	20% (8/40)
	Translation blocking	Cranial oedema	80% (40/50)
<i>lhx5</i>	Splice blocking	Lethal at early tailbud stage head and eye defect	50% (23/46)
<i>mix1</i>	Splice blocking	Blastopore closure defect	58% (31/53)
		Ventralised	100% (53/53)
		Lethal at late tailbud stage	100% (53/53)
<i>mixer</i>	Splice blocking	Blastopore closure defect	63% (30/48)
		Truncated with head and tail defects	96% (46/48)
		Lethal at late tailbud stage	77% (37/48)
<i>not</i>	Splice blocking	No defect	N/A
	Translation blocking	Truncated axis and head and tail defects	45% (24/53)
<i>tfap2a</i>	Splice blocking	No defect	N/A
<i>ventx1.2</i>	Splice blocking	Ventralised	77% (36/47)

Table 4.1. Zygotic MO screen phenotypes. The observed phenotypes and penetrance of each MO found in injected embryos. Each MO was tested in two clutches of embryos to ensure phenotypes were replicable. The results of the two separate experiments are pooled to give the penetrance of each MO.

4.2.3 Genes with strong and early defective phenotypes are interesting candidates for further study

From the screen I selected *mix1* as a candidate for further study. Mix1 knockdown embryos displayed an early, strong phenotype with high penetrance. Mix1 splice blocking MO-injected knockdown embryos appear ventralised at the tailbud stage and have head and tail defects. Around half of these embryos undergo incomplete blastopore closure during gastrulation. Mix1 MOs induce a phenotype which is lethal at the late tailbud stage (Fig. 4.3).

Mixer was selected as a candidate gene for further study, however problems discussed in Chapter 6 and time limitations meant that this was not followed up. Mix1 and Mixer are related Mix family homeobox transcription factors and their proteins share 40% identity. Mixer splice blocking MO-injected knockdown embryos display incomplete blastopore closure and defective head and tail development and are truncated (Fig. 4.3). Similar phenotypes were found in the Mix1 and Mixer MO knockdowns. Mixer MO-injected embryos were not ventralised, but did have a blastopore closure defect and were truncated with head and tail defects (Fig. 4.3).

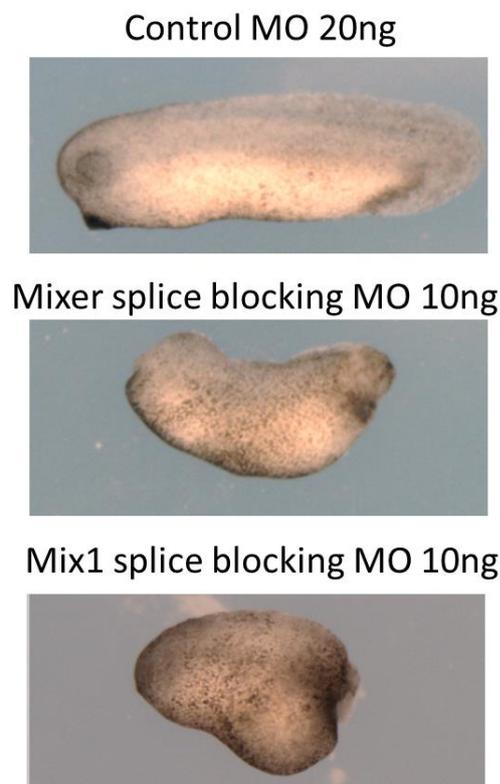


Figure 4.3. Mix1 and Mixer MO phenotypes. Embryos injected with 20ng standard control MO exhibit a normal phenotype. Embryos injected with 10ng Mixer splice blocking MO

displayed a blastopore closure defect (30/48), a truncated axis with head and tail defects (46/48) and were lethal at the late tailbud stage (37/48). Embryos injected with 10ng Mix1 splice blocking MO displayed a blastopore closure defect (31/53), were ventralised (53/53) and were lethal at the late tailbud stage (53/53). Lateral view of Mix1 knockdown embryos, Mixer knockdown embryos and control embryos at stage 29.

Further investigation of Mix1 knockdown embryos was carried out to identify the targets of this transcription factor, see Chapters 5 and 6. I note that Mixer, Foxi4.2, Ventx1.2 and Not knockdown embryos also generated strong phenotypes, including bent tail and head and tail defects and would make good candidates for further investigation (Fig. 4.2).

4.3 Screening the candidate maternal transcription factors

4.3.1 Choice of genes

In addition to testing a set of zygotically transcribed genes, I screened genes which exist as maternal transcripts in the early embryo and are rapidly polyadenylated shortly after fertilisation (presumably to make them translationally available). Knockdown embryos were screened for phenotypes as a means of identifying genes which potentially have an important role in development. I selected ten genes based on their expression, which had to be strongly activated shortly after fertilisation and closely followed by some level of deactivation (Fig. 4.4). As mentioned earlier, transiently activated genes may function to briefly regulate downstream events.

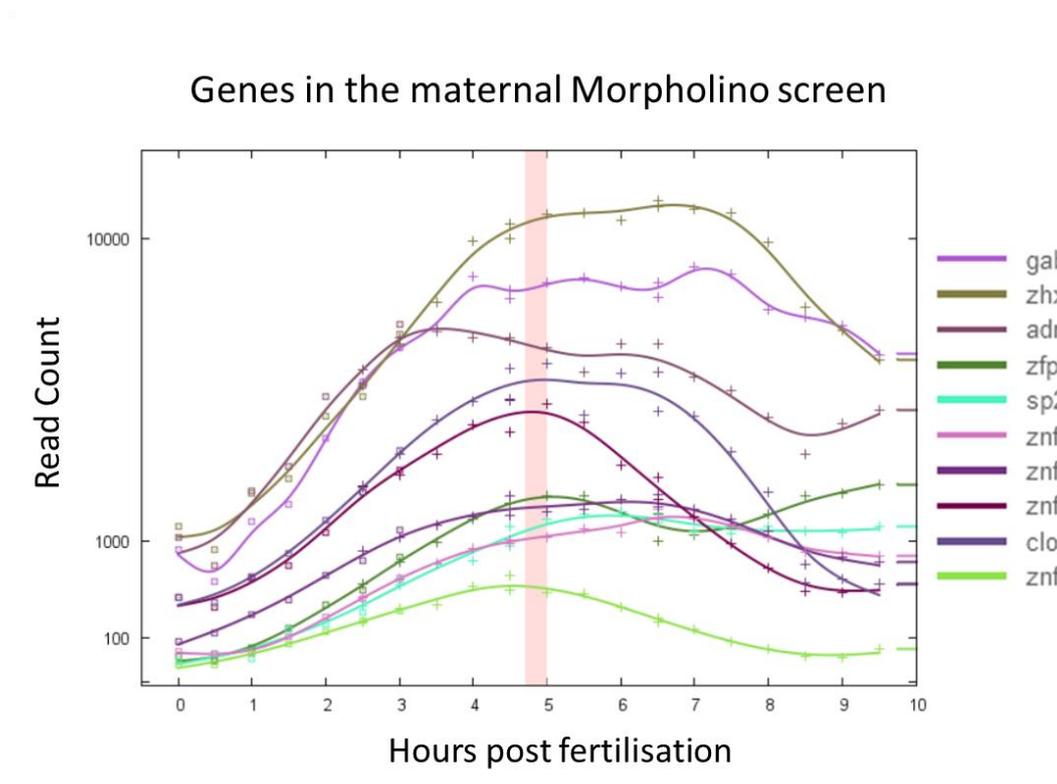


Figure 4.4. Maternal genes for MO screen. Gene expression profiles from the original RNA-seq time-series. Normalised read counts are shown for the ten maternal transcription factors included in the second MO screen for phenotypes. All ten genes are rapidly and transiently expressed from shortly after fertilisation. The MBT is marked with a pink bar (Collart et al. 2014).

As before, I used the EST database to check the quality of each gene model. 10ng of a translation blocking MO was injected in each knockdown experiment and 20ng of the standard control MO from Gene Tools was used for control injections.

4.3.2 Phenotypes from the MO screen

Genes tested and phenotypes are listed in Table 4.2. Two out of ten maternal gene knockdowns generated a defective phenotype, these were Znf821 and Clock. Both gene knockdowns generate a phenotypic defect at gastrulation; the incomplete closure of the blastopore (Fig. 4.5). The severity of the phenotype implies that these two genes may have important roles in early development. Given more time, Znf821 and Clock would make interesting targets for downstream analysis.

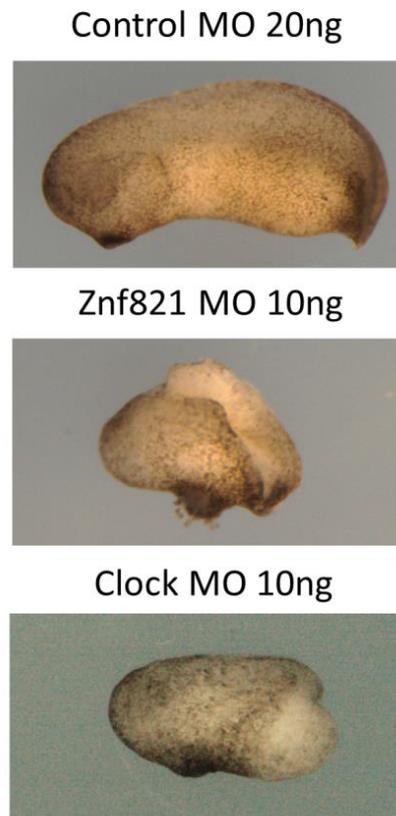


Figure 4.5. Phenotypes from the maternal MO screen. Embryos injected with 20ng standard control MO exhibit a normal phenotype. Embryos injected with 10ng of Znf821 MO or Clock MO display defective blastopore closure phenotypes (31/47) for Znf821 MO and (41/41) for Clock MO. Embryos were imaged at stage 26, lateral view.

Gene	Morpholino	Knockdown phenotype	Penetrance
<i>gabpa</i>	Translation blocking	No defect	N/A
<i>zhx3</i>	Translation blocking	No defect	N/A
<i>adnp</i>	Translation blocking	No defect	N/A
<i>zfp64</i>	Translation blocking	No defect	N/A
<i>sp2</i>	Translation blocking	No defect	N/A
<i>znf526</i>	Translation blocking	No defect	N/A
<i>znf821</i>	Translation blocking	Blastopore closure defect	84% (31/37)
<i>znf639</i>	Translation blocking	No defect	N/A
<i>clock</i>	Translation blocking	Blastopore closure defect	100% (41/41)
<i>znf827</i>	Translation blocking	No defect	N/A

Table 4.2. Maternal MO screen phenotypes. The penetrance of observed phenotypes of MO-injected embryos. Each MO was tested in two clutches of embryos to ensure phenotypes were replicable. The results of the two separate experiments are pooled to give the penetrance of each MO.

4.4 Discussion

4.4.1 MOs as a knockdown tool

The MO screen for phenotypes was an efficient way to identify genes which are likely to have an important early developmental function. A problem with MO knockdown is the potential for off-target effects where the MO erroneously targets incorrect mRNAs (Eisen and Smith 2008) Ideally, rescue or antibodies would be used to validate the knockdowns but time restrictions and a general lack of antibodies for *Xenopus* proteins meant this was not possible.

4.4.2 Mix1 and Mixer are interesting candidates for further investigation

The strong phenotypes displayed by Mix1 and Mixer knockdown embryos suggest a developmentally important role for these two transcription factors. From these two genes Mix1 was selected for further study to determine its downstream targets. The induction of similar phenotypes by different MOs targeting the same gene adds to the reliability of the results. The Mix1 MO was subject to further validations and downstream analysis in Chapters 5 and 6.

4.4.3 Several MOs do not generate a phenotype

Five of the zygotic and eight of the maternal gene knockdowns did not display a defective phenotype. It is possible that depletion of these five transcription factors does not affect the patterning of the embryo, and may have some impact on development which is unobservable in the phenotype. *In situ* hybridisation, qPCR or RNA-Seq could be used to find effects on downstream gene expression caused by knockdown of these transcription factors.

Another possibility is that the MOs are ineffective in knocking down the targeted genes. This could be the result of a robust system in which other genes compensate for the loss of a gene, or the MOs may simply not effectively block translation or splicing of the targeted transcripts. It is possible for the maternal genes that protein is also maternally contributed at fertilisation which might be the case if the maternal factors are important for development, meaning the knockdown would be harder to achieve.

4.4.4 The translation blocking MOs induced greater phenotypic effects than splice blocking MOs for genes which have maternal and zygotic transcripts

The Klf17 translation blocking MO induced a cranial oedema phenotype at higher penetrance than the splice blocking MO. This difference in penetrance is most likely an effect of the maternal transcripts which are not targeted by the splice blocking MO. Only splice blocking MO-injected embryos display a truncated phenotype. Based on the idea that the translation blocking MO has a greater impact on Klf17 depletion, it is likely that the truncated phenotype is due to off-target effects of the splice blocking MO. This could be tested using a second MO or a dominant negative Klf17 construct.

Not translation blocking MO-injected embryos displayed a phenotype whereas Not splice blocking MO-injected embryos did not. There are abundant maternal *not* transcripts present from fertilisation which are not targeted by the splice blocking MO. The maternal transcripts most likely generate sufficient levels of Not protein to maintain normal gene function, thus cancelling the effect of the splice blocking MO. The level of maternal *not* mRNA is much higher than the maternal *klf17* mRNA in the embryo, which may explain why the Klf17 splice blocking MO has an effect on the phenotype and the Not splice blocking MO does not.

4.4.5 Comparison of phenotypes from my screen to previous knockdown studies

Comparison to previous knockdown experiments gives some indication of the specificity of the MOs used in my screen. Comparisons to other MO knockdown experiments are discussed below (Table 4.3).

Gene	Species	Phenotype	<i>In situ</i> staining	Reference
<i>bix1.1</i>	No MO studies	N/A	N/A	N/A
<i>eomes</i>	<i>X. tropicalis</i>	No morphological phenotype	Reduced <i>vegt</i>	(Fukuda et al. 2010)
	zebrafish	No morphological phenotype	Reduced <i>gsc</i> and <i>flh (not)</i>	(Bruce et al. 2003)
<i>foxi4.2</i>	<i>X. laevis</i>	No morphological phenotype	Head ectoderm impaired	(Matsuo-Takasaki, Matsumura, and Sasai 2005)
<i>grh13</i>	No MO studies	N/A	N/A	N/A
<i>gsc</i>	<i>X. laevis</i>	Anterior truncations and ventralised	Expanded ventral expression and reduced <i>myod</i>	(Sander, Reversade, and De Robertis 2007)
	zebrafish	Cyclopia and loss of head development	Presumptive forebrain and midbrain markers reduced and <i>wnt8</i> up-regulated dorsally.	(Seiliez, Thisse, and Thisse 2006)
<i>klf17</i>	<i>X. laevis</i>	Loss of hatching gland and otic vesicle formation.	Reduced neural crest formation.	(Kurauchi, Izutsu, and Maeno 2010)
<i>lhx5</i>	<i>X. laevis</i>	Neural tube closure delay	N/A	(Houston and Wylie 2003)
	zebrafish	Defective forebrain development	Reduced the Wnt antagonist <i>sfrp1</i> in presumptive forebrain	(Peng and Westerfield 2006).
<i>mix1</i>	<i>X. laevis</i>	Head defects	Increased levels of <i>fgf</i> , <i>myod</i> , <i>brachyury</i> , <i>pax8</i> and <i>lhx1</i> .	(Colas et al. 2008)
<i>mixer</i>	<i>X. laevis</i>	Anterior and gut abnormalities	Reduced endodermal marker expression and increased mesodermal marker and <i>nodal</i> expression.	(Kofron, Wylie, and Heasman 2004)

<i>not</i>	No MO studies	N/A	N/A	N/A
<i>tfap2a</i>	<i>X. laevis</i>	No morphological phenotype	Loss of neural crest marker expression	(Luo et al. 2003)
<i>ventx1.2</i>	<i>X. laevis</i>	No morphological phenotype	Expanded neural plate	(Sander, Reversade, and De Robertis 2007)
<i>adnp</i>	No MO studies	N/A	N/A	N/A
<i>clock</i>	zebrafish	No morphological phenotype	N/A	(Li et al. 2008)
<i>gabpa</i>	No MO studies	N/A	N/A	N/A
<i>sp2</i>	zebrafish	Arrest at gastrulation and lethal within 24 hours. Lower doses induced a shortened axis defective head, tail and somite development.	N/A	(Xie et al. 2010)
<i>zfp64</i>	No MO studies	N/A	N/A	N/A
<i>zhx3</i>	No MO studies	N/A	N/A	N/A
<i>znf526</i>	No MO studies	N/A	N/A	N/A
<i>znf639</i>	No MO studies	N/A	N/A	N/A
<i>znf821</i>	No MO studies	N/A	N/A	N/A
<i>znf827</i>	No MO studies	N/A	N/A	N/A

Table 4.3. Knockdown phenotypes in previous studies. A summary of phenotypes identified in previous studies for all genes tested in the zygotic and maternal gene MO screens outlined in this chapter. All visible phenotypes and phenotypes identified using *in situ* hybridisation staining are outline in the table.

4.4.6 Comparing the zygotic gene screen to previous studies

Eomes: No observable phenotypic defects were found in *Eomes* knockdown embryos both in my screen and in previous studies. Although no phenotype was found in previous studies, *in situ* hybridisation revealed differential expression of *vegt*, *gsc* and *not* (Table 4.3) (Fukuda et al. 2010; Bruce et al. 2003).

Gsc: No phenotypic defects were observed in Gsc knockdown embryos in my screen, but previous MO knockdown studies in zebrafish and *X. laevis* report anterior defects in knockdown embryos (Seiliez, Thisse, and Thisse 2006; Sander, Reversade, and De Robertis 2007). Gsc has been linked to head specification and the *gsc* mis-expression in *X. laevis* induced a secondary dorsal axis (Cho et al. 1991). *X. laevis* embryos injected with *gsc* antisense RNA also displayed anterior head defects (Steinbeisser et al. 1995).

Lhx5: The head and eye defects found in Lhx5 morphants in my screen contrasts with the neural tube closure delay found in *X. laevis* morphants (Houston and Wylie 2003). The phenotypes are somewhat similar to the zebrafish Lhx5 knockdown phenotype in which forebrain development is defective and the dominant negative Lhx5 phenotype in which head and eye development are impaired (Peng and Westerfield 2006).

Mix1: MO knockdown of Mix1 in *X. laevis* generated embryos with head defects which is consistent with the head defect phenotype found in my screen. There was no report of an early gastrulation defect or ventralised phenotype in the Mix1 morphant *X. laevis* embryos (Colas et al. 2008).

Mixer: Mixer MO was injected into *X. laevis* embryos resulting in anterior and gut abnormalities, the Mixer MO-injected embryos in my screen also displayed anterior abnormalities (Kofron, Wylie, and Heasman 2004). No gut abnormalities were visible in the Mixer morphant tailbud stage embryos in my screen, which are clearly different to the ventral abnormalities presented by Kofron et al. The phenotypes are similar; both have a shortened antero-posterior axis and reduced head and eye development. Differences may be due to different penetrance of the two MOs (Kofron, Wylie, and Heasman 2004).

Tfap2a: No observable phenotypic defects were found in Tfap2a knockdown embryos both in my screen and in another knockdown study. No visible knockdown phenotype was observed, but neural crest marker gene expression was perturbed (Luo et al. 2003).

Foxi4.2, Klf17 and Ventx1.2: The zygotic gene knockdowns of Foxi4.2, Klf17 and Ventx1.2 displayed defective phenotypes in my screen but previous knockdown studies did not report any defects. *In situ* staining revealed perturbed gene expression in the Foxi4.2, Klf17 and Ventx1.2 knockdowns in *X. laevis* (Table 4.3) (Kurauchi, Izutsu, and Maeno 2010; Matsuo-Takasaki, Matsumura, and Sasai 2005) (Sander, Reversade, and De Robertis 2007).

No previous MO studies have been carried out for Bix1.1, Not and Grh3.

4.4.7 Comparing the maternal gene screen to previous studies

Clock: Clock was knocked down in Zebrafish using a MO, and no morphological defects were reported, which is inconsistent with my finding (Li et al. 2008).

Sp2: No phenotypic defects were observed in Sp2 knockdown embryos in my screen, however previous studies report defective phenotypes. Sp2 MO knockdown in zebrafish caused embryos to stall during gastrulation (Xie et al. 2010). Sp2 null mouse embryos were lethal at E9.5 of gestation, indicating an essential role for Sp2 in early development (Baur et al. 2010).

There are no published MO knockdown studies for Adnp, Gabpa, Zfp64, Zhx3, Znf526, Znf639, Znf821 and Znf827.

The differences in knockdown phenotypes found between my zygotic and maternal screen and other studies could be due to ineffective MOs, off-target effects, or could be due to species differences; most previous studies are carried out in *X. laevis* and zebrafish. An intriguing possibility is that *X. laevis* and *X. tropicalis* regulatory pathways differ as there are several more *mix* family genes in *X. laevis* than in *X. tropicalis* which may have overlapping functions (Pereira et al. 2012).

4.4.8 Conclusion

This work reveals a higher propensity for phenotypes when knocking down zygotic genes compared to maternal genes. This may reflect a more important developmental role for the zygotic transcription factors, or may simply be an effect of the technical issues of gene knockdown. The aim of using the screen to select genes with interesting phenotypes for further work has been fulfilled, and Mix1 is characterised for its downstream targets in subsequent chapters.

Chapter 5: Investigating downstream targets of Mix1

5.1. Introduction

In this chapter I investigate the downstream targets of Mix1, a transcription factor which was included in the knockdown screen described in the previous chapter. Mix1 was selected for the screen because it is one of the first transcription factors to be transcribed in early *X. tropicalis* development, and is expressed transiently. This time-limited effect may mean that Mix1 is only present for a short period and therefore its direct targets can be elucidated within a relatively short time-window. Mix1 was selected from the screen for further study because of the strong and early phenotype exhibited by Mix1 morpholino oligomer (MO)-injected embryos, indicating potential developmental significance for Mix1. Only two direct targets have been reported: *cer1* and *gsc*, and identifying more targets may reveal novel functions for Mix1 (Sudou et al. 2012).

This chapter's aims were to investigate how Mix1 regulates early development by determining its downstream targets and to evaluate the effectiveness of the time-series approach for identifying transcription factor targets. RNA was sequenced from splice blocking MO-injected Mix1 knockdown and control embryos sampled over a time-series and then analysed to find which genes showed divergent expression, suggesting up- or down-regulation by Mix1.

5.2 Results

5.2.1 Phenotypes and validation of the Mix1 splice blocking MO

5.2.1.1 Mix1 splice blocking MO-induced phenotypes

The injection of the Mix1 splice blocking MO was shown to induce phenotypic defects in the screen described in Chapter 4. With the knowledge that 10ng gives a penetrant phenotype, the MO was also injected at 10ng and at lower doses and the morphology of control and knockdown embryos was compared in order to observe the penetrance of phenotypic defects caused by depletion of Mix1. The Mix1 splice blocking MO generated defective phenotypes in injected embryos at the 5ng and 10ng doses (Fig. 5.1B, C). Of the three doses tested, 10ng of splice blocking MO was found to give the highest penetrance of phenotype (Fig. 5.1A). 95% of embryos injected with Mix1 splice blocking MO were ventralised with head

and tail defects (Fig. 5.1C), 51% failed to undergo complete blastopore closure (Fig. 5.1B) and for 100% of embryos splice blocking MO injection was lethal at the late tailbud stage.

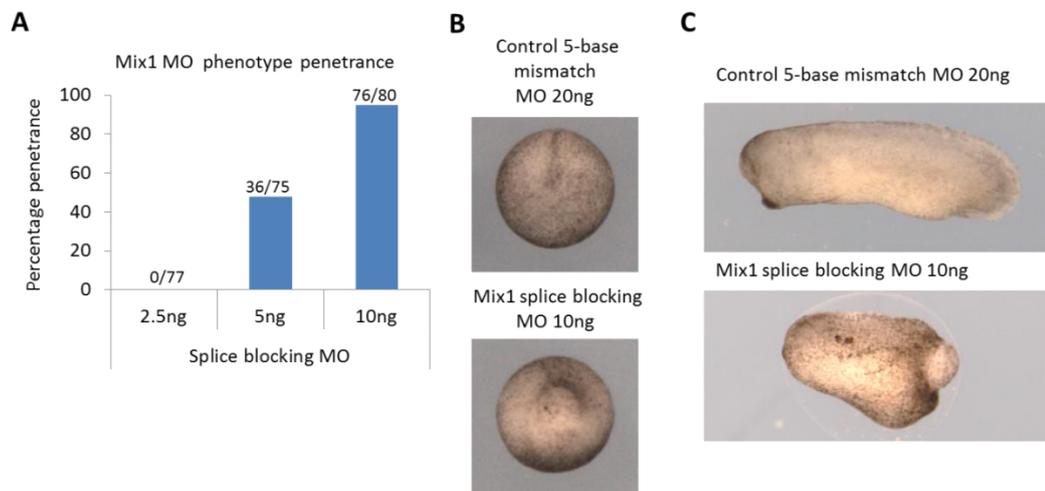


Figure 5.1. Mix1 splice blocking MO phenotype. (A) Increasing doses of Mix1 splice blocking MO were injected and phenotypic penetrance was recorded. Penetrance of ventralised phenotype for splice blocking MO is shown. Mix1 MO was injected at the three doses into three separate clutches of embryos in independent experiments, and the results were pooled to give the penetrance for each dose. (B) Vegetal view of stage 13 embryos injected with 20ng control 5-base mismatch MO (top) or 10ng Mix1 splice blocking MO (lower). (C) Lateral view of stage 26 embryos injected with 20ng 5-base mismatch control MO display no visible phenotype. Embryos injected with 10ng Mix1 splice blocking MO were ventralised in 76/80 embryos with defective blastopore closure in 41/80 embryos. The MO caused lethality at the late tailbud stage in 100% of injected embryos at 10ng and 5ng dose. 4/77 (5%) embryos injected with 2ng MO were dead at the late tailbud stage.

5.2.1.2 Mix1 splice blocking MO validation

I carried out validation experiments to test the splice blocking activity of the Mix1 splice blocking MO. I used RT-PCR to confirm the disrupted splicing of Mix1 mRNA in splice blocking MO-injected embryos. The splice blocking MO was targeted to the 1st exon-intron boundary to prevent splicing of the 1st intron. The inclusion of the first intron placed an in-frame stop codon within the mis-spliced mRNA. This stop codon was upstream of the homeodomain, which is encoded in the 2nd and 3rd exons (Fig. 5.2C).

To test the effect on splicing for the splice blocking MO, embryos were injected with either 10ng of Mix1 splice blocking MO or 20ng of 5-base mismatch negative control MO and harvested at 7.0 hpf. The 5-base mismatch control MO should not impair development, so a higher dose was injected to ensure the control had no phenotypic effect. RT-PCR using

primers annealing to the 1st and 3rd exons revealed a larger splice variant in Mix1 MO samples compared to controls. The fragment size confirmed the inclusion of the first intron. As a control, primers were targeted to the 2nd and 3rd exon which generated two bands of equal size (Fig. 5.2A, C).

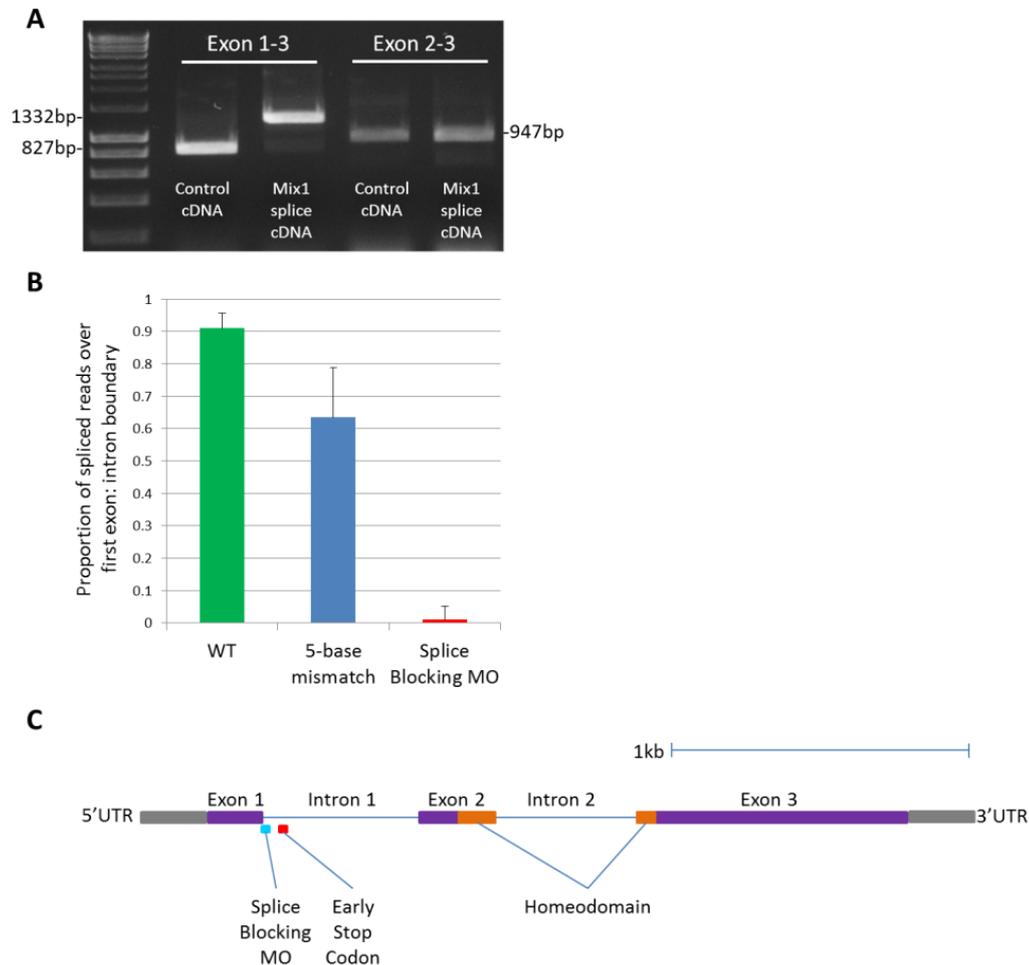


Figure 5.2. Mix1 splice blocking MO activity (A) Mix1 MO splice blocking assay by RT-PCR from embryos injected with control 5-base mismatch control MO or Mix1 splice blocking MO. Inclusion of the first intron in Mix1 splice blocking MO cDNA generated a longer fragment from exon 1- exon 3 (1332bp instead of 827bp). The fragment from exon 2- exon 3 was amplified as a control to reveal unaffected splicing beyond intron 1 (947bp). (B) RNA-seq reads were aligned to the Mix1 locus, the mean percentage of spliced reads over six 30 minute time points from 5-7.5 hpf when Mix1 is transcribed. RNA-seq samples at the first exon overhang are shown, error bars give standard deviation. (C) The Mix1 genomic locus has three exons. The MO binding sites, early stop codon induced by the splice blocking MO and homeodomain region are labelled.

As a second confirmation of Mix1 splice blocking MO activity, RNA-seq reads from Mix1 MO-injected, Mix1 5-base mismatch MO-injected and WT un-injected embryos were aligned to

the Mix1 locus (see below for details of time-series collection and sequencing). The percentage of transcripts partially overlapping the 3' end of exon 1 which were spliced out of the total reads (spliced and un-spliced) was calculated for the three different experimental conditions. This analysis was done by Nick Owens. This was used to compare the splicing efficiency between conditions (see Materials and Methods for details) (Fig. 5.2B, C). The percentage of spliced transcripts in the WT condition was 91%. Less than 1% of transcripts were spliced at this junction in Mix1 splice blocking MO-injected embryos indicative of effective splice blocking. Interestingly, in the control 5-base mismatch MO condition 64% of transcripts were spliced which indicates that the 5-base mismatch control MO has some splice blocking activity (Fig. 5.2B). Importantly, no phenotypic defects were observed in embryos injected with the 5-base mismatch MO suggesting that the partial splice blocking does not knock down Mix1 function (see below). These two assays demonstrate the ability of the Mix1 splice blocking MO to effectively inhibit splicing of *mix1* mRNA.

5.2.2 Expression of *mix1* in the early embryo

5.2.2.1 Expression of *mix1*

In order to understand where Mix1 targets might be expressed, I first explored the localisation of *mix1* mRNA within the embryo during the blastula and gastrula stages. Mix1 and its activatory targets should be expressed in overlapping regions and its repressive targets may be expressed in unshared regions. This information was used to guide later validation experiments. *In situ* hybridisation analyses showed that *mix1* expression is first detected in the dorsal marginal region in the early NF stage 9 embryo (Nieuwkoop and Faber 1994). This dorsal expression domain includes the region fated to become the blastopore lip at the start of gastrulation. By late stage 9 the *mix1* expression domain expands to occupy the entire marginal zone, but is more highly expressed in the dorsal region. At stage 10 *mix1* has prominent expression throughout the marginal zone with lower expression throughout the vegetal pole. This has been demonstrated previously using RT-PCR, in which explants from the dorsal-marginal zone and vegetal pole were assayed for *mix1* expression, and a longer gel exposure was required for the vegetal explant condition due to a lower amount of mRNA present in the vegetal cells (Lemaire et al. 1998). At stage 10.5 *mix1* is expressed around the marginal zone but is diminished in the dorsal-most region. Overall there is a shift in the *mix1* expression domain from dorsal to ventral (Fig. 5.3). The developmental stages used for *in situ* hybridisation are marked on a time scale of hours post fertilisation (Fig. 5.3).

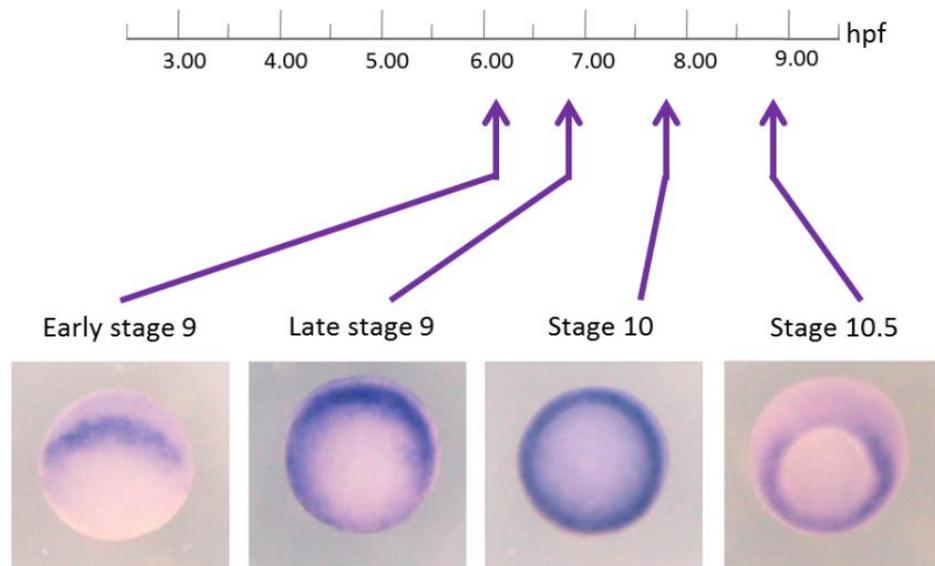


Figure 5.3. Localisation of *mix1*. The stages at which embryos were harvested and fixed for *in situ* hybridisation are marked with arrows on the time scale. Stages are marked according to the developmental rate used in this particular RNA-seq experiment. *Mix1* expression is stained by whole mount *in situ* hybridisation at stage 9 (early), stage 9 (late), stage 10 and stage 10.5. Embryos are viewed from the vegetal pole, with the dorsal side oriented at the top. *Mix1* is expressed first at the dorsal blastopore lip, the expression domain gradually expands to the entire marginal zone and dorsal expression depletes by stage 10.5. Embryos are viewed dorso-vegetally with the dorsal pole towards the top. Early stage 9 and stage 10.5 are tilted, so slightly more dorsal pole is viewed. Arrows mark the time of *in situ* (hours post fertilisation).

5.2.3 Time-series analysis of Mix1 splice blocking MO-injected knockdown embryos reveals candidate Mix1 targets

5.2.3.1 Generating knockdown and control time-series

A high resolution RNA-seq time-series was generated over several hours of early development. The aim was to use a time-series approach to identify candidate Mix1 targets by investigating differential downstream expression in knockdown embryos and to determine which genes are affected by the depletion of Mix1.

To generate the time-series data, embryos from a single clutch were either injected with Mix1 splice blocking MO, injected with 5-base mismatch control MO or were not injected (WT). By removal of asynchronously dividing embryos during the first five cleavages,

synchronous embryos were collected every 30 minutes from 2.5 hpf until 9.0 hpf (Fig. 5.4). Sampling from 2.5 hpf ensured the activation of *mix1* was captured as published wild type *X. tropicalis* time-series data reports *mix1* activation at 4.0 hpf when embryos were incubated at 23°C (Collart et al. 2014). PolyA+ mRNA from the embryos collected in this time-series was sequenced to find differences in gene expression between the knockdown and control embryos, in order to investigate the effects of Mix1 depletion.

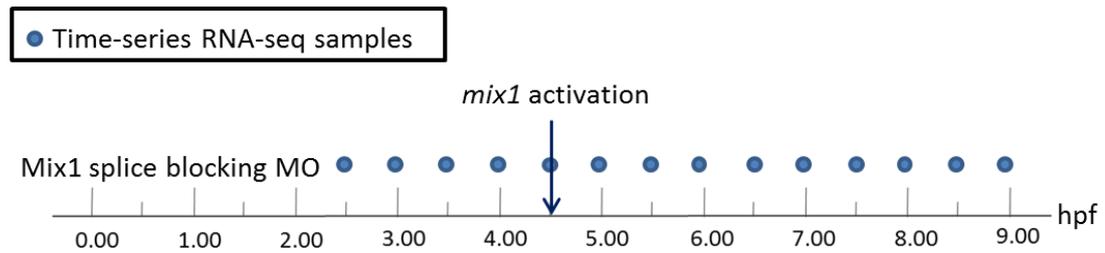


Fig. 5.4. Mix1 knockdown time-series collection. The time points collected for the Mix1 splice blocking MO time-series are marked on the time-scale (hpf).

5.2.3.2 Data normalisation and differential expression calling

RNA-seq reads were aligned to the *X. tropicalis* transcriptome and read counts for each gene were normalised against the total read counts. Normalised read counts were plotted to generate high resolution expression profiles for all detected genes (analysis by Nick Owens, see Materials and Methods for details).

Gene expression profiles from the splice blocking MO and control MO conditions were compared for each gene to determine which were differentially expressed (analysis by Nick Owens, see Materials and Methods for details). The 5-base mismatch control MO condition is the correct control for the Mix1 splice blocking MO. Some splice blocking activity was demonstrated for the 5-base mismatch control MO and the wild type condition was used to confirm that this partial splice blocking did not unduly affect gene expression. The wild type and control MO conditions showed congruent expression patterns, so the control condition was used in the differential expression analysis. The genes which are differentially expressed between knockdown and control conditions are candidate targets of Mix1. Experiments to validate these candidate targets can be found in Chapter 6.

5.2.3.3 Activation of *mix1*

Mix1 targets will be perturbed after *mix1* is activated. To limit off-target effects of the Mix1 MO, I examined the activation time of *mix1* in this time-series. The activation was calculated according to published methods (Collart et al. 2014) by measuring gene expression increases in time-series RNA-seq data, to define activation where successive time-points rise past a 10 fold threshold (analysis by Nick Owens, see Materials and Methods for details). The time-series expression profile of wild type *mix1* shows it is activated at 5.0 hours post fertilisation (hpf), just under an hour before MBT (Fig. 5.5). This is slower than in previous time-series (Collart et al. 2014) and is likely caused by low incubation temperatures, which were 21-22°C rather than 23°C.

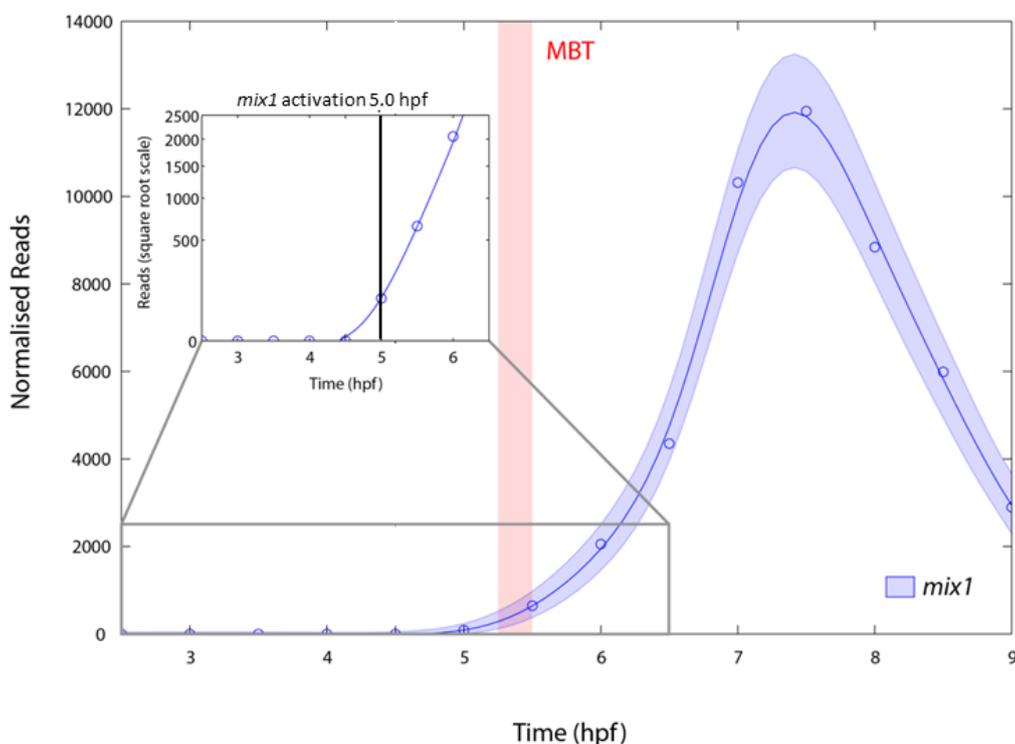
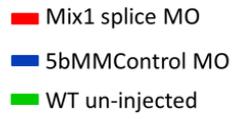


Fig. 5.5. The *mix1* expression profile. Profile derived from normalised RNA-seq reads from WT embryos shows activation at 5.0 hpf (inset) and peaks at 7.5 hpf. MBT is marked with a vertical pink bar. The line represents the median and the shaded region shows the 95% confidence intervals. Figure generated by Nick Owens.

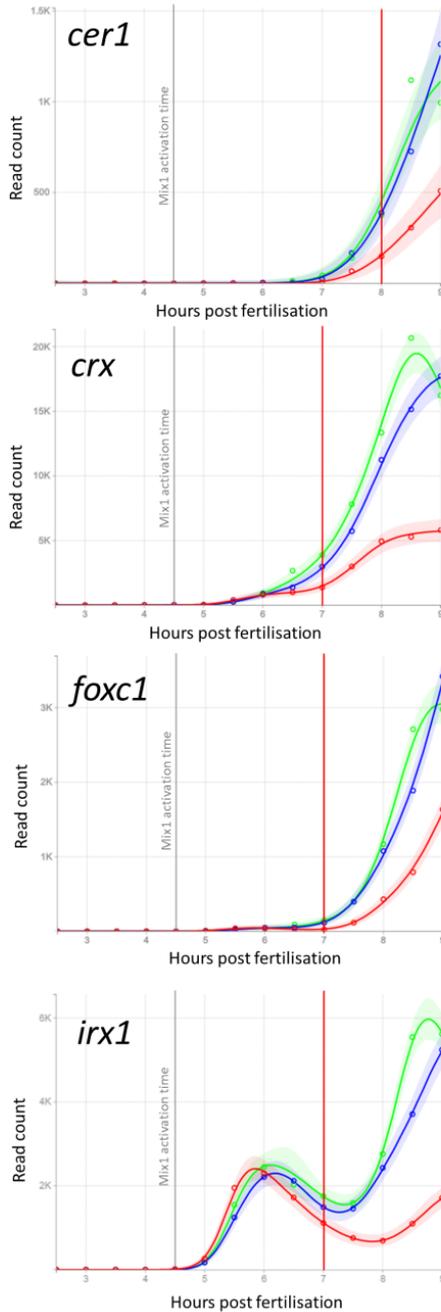
5.2.3.4 Assigning divergence times

In order to understand when Mix1 regulates its targets, I investigated the timing of differential expression events. A gene is considered to be a candidate target if its control MO and Mix1 splice blocking MO expression is equal prior to Mix1 activation and the control MO

and Mix1 splice blocking MO expression diverges after Mix1 activation. Any gene that diverges prior to Mix1 activation is likely to be an off-target effect. The divergence time is assigned as the time point after the Gaussian process models overlap by only 10% (analysis by Nick Owens, see Materials and Methods for details). This allows clear comparison of divergences. The divergence time of each differentially expressed gene is marked on the gene profile as a red vertical line. The expression profile of the Mix1 candidate activatory targets; *cer1*, *crx*, *foxc1* and *irx1* and the candidate Mix1 repressive targets; *bix1.1*, *ppp1r10*, *ventx3.2* and *szl* are shown as examples (Fig. 5.6). The majority of detected genes did not display a divergent profile (15148 genes had no differential expression) indicating that differential expression is not a result of developmental delay.



A Activatory Targets



B Repressive Targets

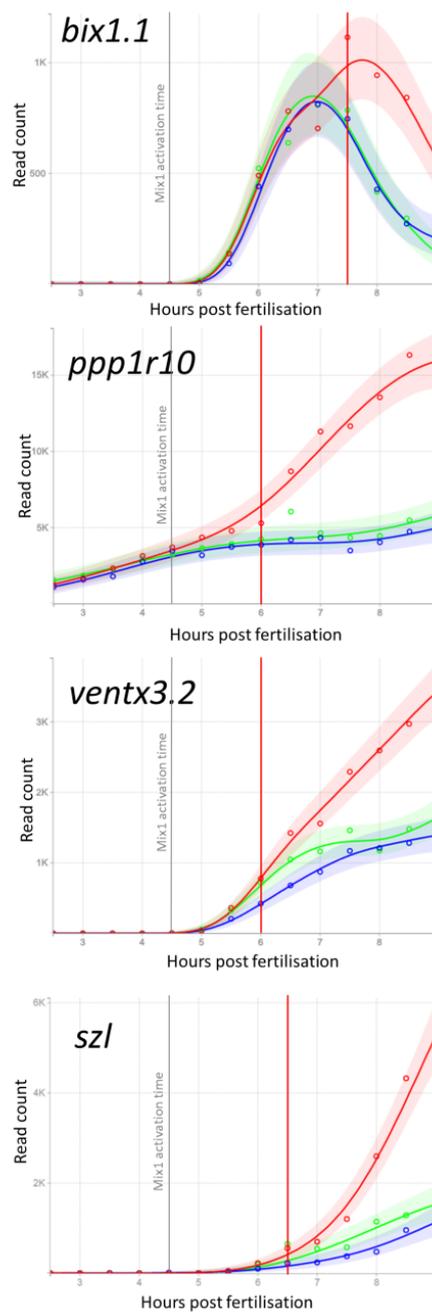


Figure 5.6. Candidate Mix1 target expression profiles. The expression profiles of (A) The candidate Mix1 activatory targets; *cer1*, *crx*, *foxc1* and *irx1*, and (B) The candidate Mix1 repressive targets; *bix1.1*, *ppp1r10*, *ventx3.2* and *szl*. These genes are differentially expressed in the Mix1 splice blocking MO time-series. Normalised read counts for Mix1 MO (red), 5-base mismatch control MO (blue) and wild type (green) conditions were plotted for all time points. Best fit lines with 95% confidence intervals were calculated. The red vertical line marks the calculated divergence time. Gene expression profiles generated by Nick Owens. This experiment was carried out in a single clutch of embryos.

5.2.3.5 Many candidate Mix1 targets were identified

In total 497 genes were differentially expressed in the Mix1 splice blocking MO knockdown time-series compared to controls. 189 genes were down-regulated after divergence in knockdown embryos and are candidate Mix1 activatory targets. 308 genes were up-regulated after divergence and are candidate Mix1 repressive targets (Fig. 5.7). The full list of differentially expressed genes can be found in the appendix on the attached CD (tab 2). Interestingly, the BMP signalling molecules, which are ventralisation markers, were not significantly upregulated, which might be expected since Mix1 knockdown embryos display a ventralised phenotype. BMP7.1 is slightly upregulated, but is not significantly differentially expressed.

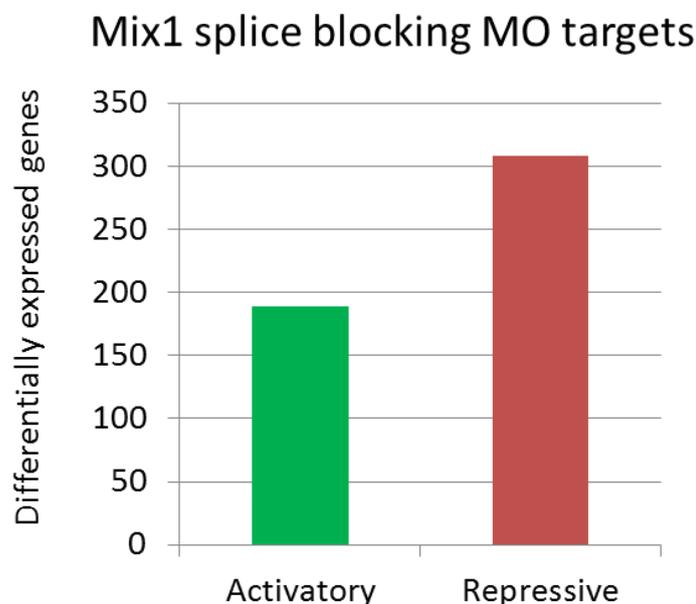


Figure 5.7. Total candidate Mix1 targets. Total activatory and repressive candidate Mix1 target genes differentially expressed in the splice blocking MO time-series. 308 genes were found to have reduced expression in knockdown embryos compared to controls and are

known as candidate Mix1 activatory targets. 497 genes were found to have increased expression in knockdown embryos compared to controls and are known as candidate Mix1 repressive targets. This experiment was carried out in a single clutch of embryos.

5.2.4 The divergence timings in the splice blocking MO time-series

To explore the timing of differential expression events in Mix1-depleted embryos, I examined the number of divergences at each time point in the splice blocking MO time-series (Fig. 5.8). The majority of divergences occur from 6.0 to 8.0 hpf. The number of divergences diminishes from 8.5 hpf. With the proximity to the end of the time-series, the capacity to detect late divergences diminishes. A number of time points are required to clearly call a divergence and the number of time-points required depends on the magnitude of the divergence. Later in the time-series, only genes that diverge rapidly can be detected.

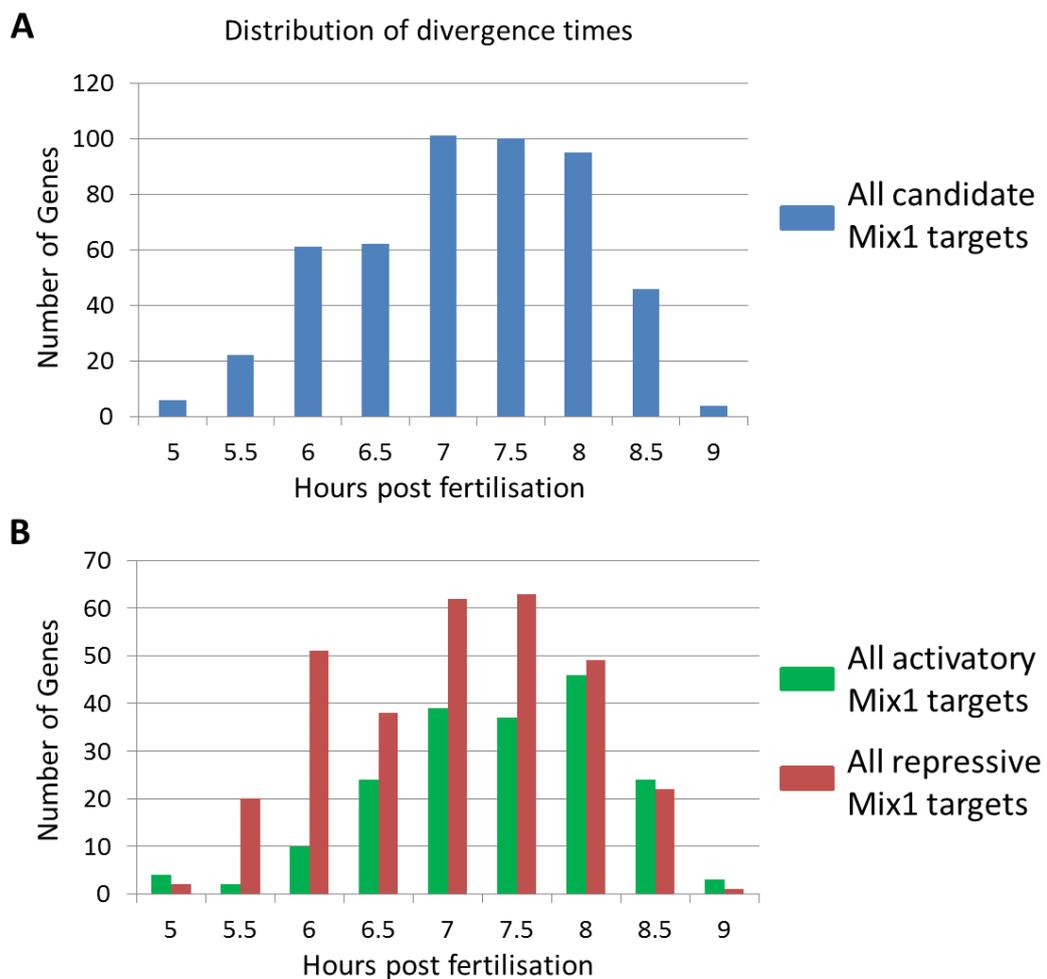


Figure 5.8. Divergence times of candidate Mix1 targets. (A) The number of divergences at each time point for candidate Mix1 targets which are differentially expressed in the splice

blocking MO time-series. (B) The number of divergences at each time point for candidate Mix1 activatory targets (green) and candidate Mix1 repressive targets (red) which are differentially expressed in the splice blocking MO time-series.

5.2.5 Gene expression perturbations are reproducible in Mix1 knockdown embryos

5.2.5.1 Gene expression perturbations found in the splice blocking MO knockdown time-series were replicated and measured by qPCR

To confirm the reproducibility of gene expression perturbations observed in the knockdown time-series, I compared expression of candidate Mix1 targets between Mix1 splice blocking MO-injected and control MO-injected embryos. I performed qPCR on cDNA from three new, separate clutches of embryos harvested at 9.0 hpf. This is when the maximum expression difference is observed in the gene expression profiles of most Mix1 candidate targets. The fold changes in expression levels between Mix1 splice blocking MO and control MO cDNA samples were calculated. All three activatory target genes tested *crx*, *foxc1* and *cer1* displayed significantly reduced expression in the Mix1 splice blocking MO-injected embryos compared to controls (Fig. 5.9). The fold changes found in the qPCR replicates were consistent with those found in the RNA-seq data (see Materials and Methods for calculation of RNA-seq fold changes and error). This demonstrates the reproducibility of Mix1 downstream target expression perturbations in knockdown embryos.

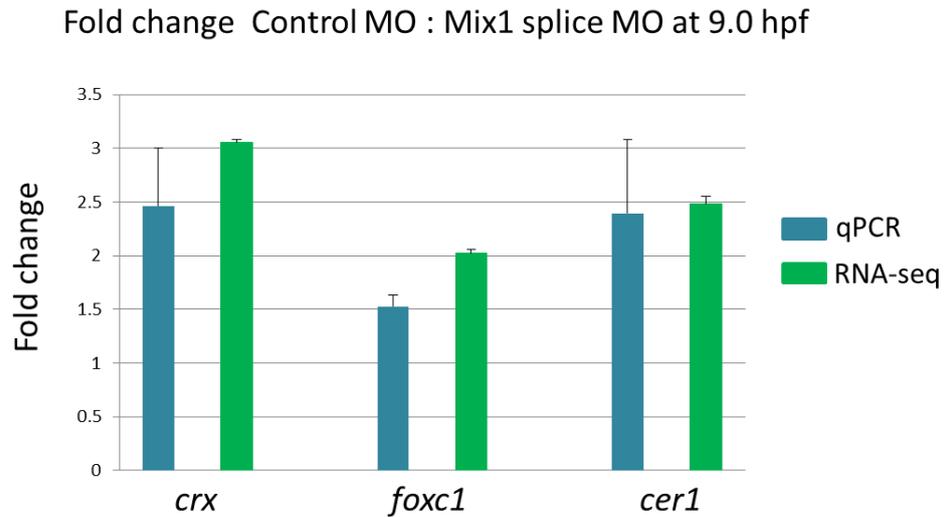


Figure 5.9. Candidate Mix1 target expression is reproducible. Quantitative PCR of three biological replicates of Mix1 splice blocking MO-injected and control MO-injected embryos at stage 10.5 (9.0 hpf) shows reproducible differential expression of the Mix1 activatory targets *crx*, *foxc1* and *cer1*. Averaged replicate values give a fold change of at least 1.5 for all genes tested. The Student's t-test was applied to control and MO concentrations normalised to *odc1* to show a significant difference p-values: *crx*: 0.0440, *foxc1*: 0.0287, *cer1*: 0.0159. See Materials and Methods for the calculation of the RNA-seq error bars. RNA-seq fold changes at 9.0 hpf are also shown and are comparable to qPCR fold changes.

5.2.5.2 Confirmation of differentially expressed genes in Mix1 depleted embryos

I used *in situ* hybridisation to validate the localised mis-expression of some Mix1 activatory targets within the *mix1* expression domain in response to MO injection. Mix1 splice blocking MO was injected unilaterally to compare gene expression between knockdown and wild-type halves of the embryo. The MO was co-injected with NLS- β -galactosidase mRNA to confirm the injected region. Embryos were then fixed at stage 10 (Fig. 5.10A).

I stained for expression of three candidate Mix1 targets *cer1*, *crx* and *foxc1* to test for altered expression in the region where *mix1* would normally have been expressed. Mix1 activatory targets would be expected to have reduced expression within the *mix1* expression domain in the Mix1-depleted half of the embryo. At stage 10 *mix1* is expressed throughout the vegetal hemisphere and is highest in the marginal zone (Fig. 5.10A).

Consistent with the reduced expression of these genes in the knockdown time-series, the Mix1 activatory targets *cer1*, *crx* and *foxc1* underwent a reduction in the intensity of expression in the Mix1 depleted half of the embryo (Fig. 5.10B, C, D). The expression domains of *foxc1* and *cer1* were reduced in intensity within the *mix1* expression domain (Fig. 5.10C, D). The domain of *crx* was almost entirely depleted in the vegetal region of its expression domain within the *mix1* expression domain and appears to be more weakly depleted towards the animal pole, which is outside of the *mix1* expression domain (Fig. 5.10B). These three candidate Mix1 targets all showed depleted expression in the *mix1* expression domain and strengthen the evidence that these are Mix1 targets.

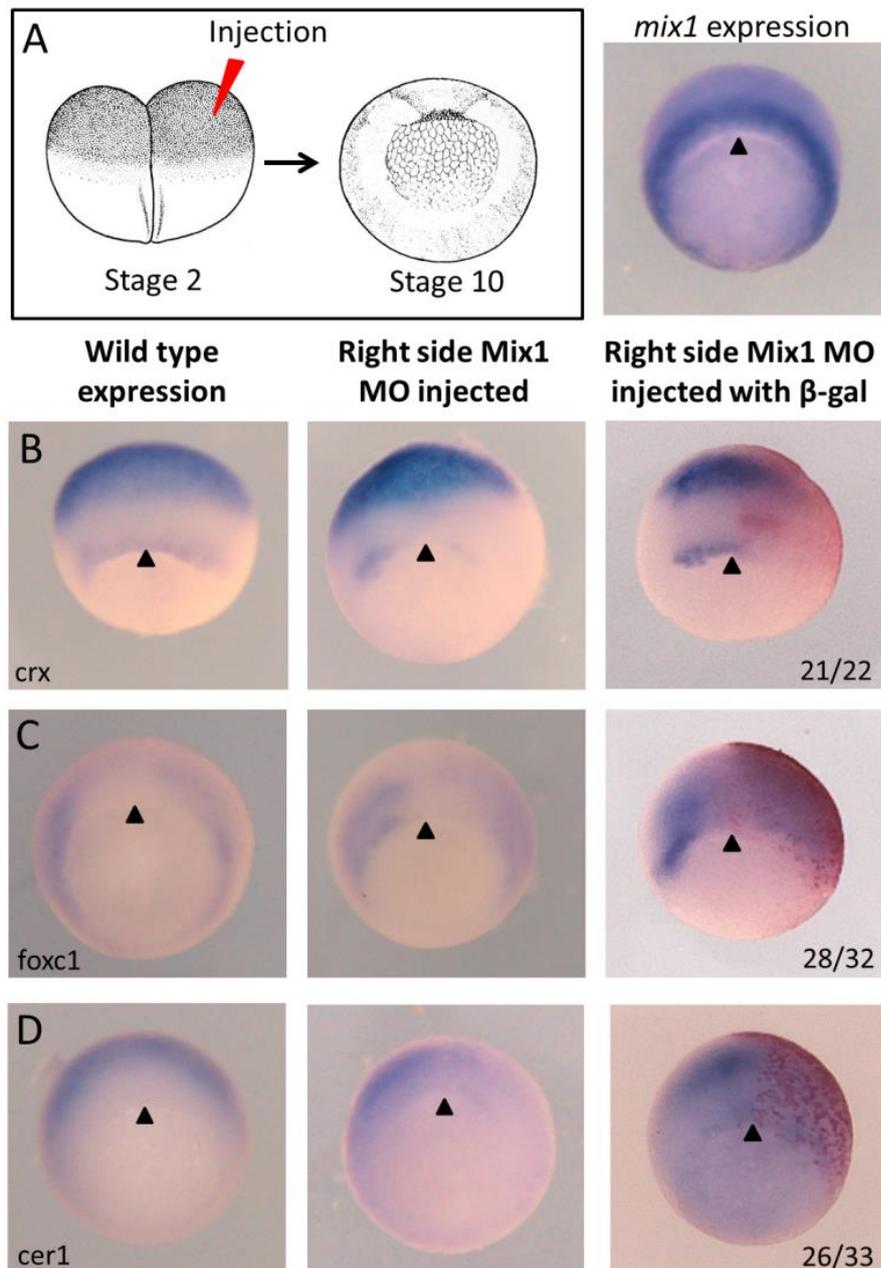


Figure 5.10. Altered expression of candidate Mix1 activatory targets in response to Mix1 depletion. (A) Embryos were injected with Mix1 MO into one blastomere at the 2-cell stage to give unilateral knockdown (on the right side of embryos). *NLS- β -galactosidase* mRNA was co-injected as a lineage tracer to give nuclear red stain seen in the injected region of embryos. The expression domain of Mix1 at stage 10 spans the marginal zone. Mix1 target mRNA was stained in stage 10 embryos by whole mount *in situ* hybridisation. (B, C, D) Expression of the candidate Mix1 activatory targets *crx*, *foxc1* and *cer1*. *In situ* hybridisations were carried out on embryos from three separate clutches in independent experiments. Stage 10 embryos are viewed from the vegetal pole, with the dorsal side oriented at the top. The blastopore lip is marked with an arrowhead. Phenotypic penetrance is shown in the bottom right corner for each gene tested. *Xenopus* stage images reproduced from Nieuwkoop and Faber (Nieuwkoop and Faber 1994) with permission Garland Science/Taylor & Francis, LLC.

5.2.6 Overexpression of *mix1* disrupts embryonic development

Mix1 mRNA was injected into embryos in order to test the effects of *mix1* overexpression. The response to different doses of *mix1* mRNA was used to guide the levels injected for rescue experiments (see below).

Control embryos were injected with GFP mRNA at 50pg, as this mRNA should not induce any phenotypic changes. As expected, control embryos displayed a normal phenotype and GFP expression was confirmed (not shown) (Fig. 5.11A). Embryos injected with 25pg of *mix1* mRNA were truncated along the antero-posterior axis and had shortened tails (37/40) (Fig. 5.10B). Embryos injected with 50pg of *mix1* mRNA failed to undergo elongation of the antero-posterior axis (43/43) and in many cases embryos completely lacked a head (27/43) (Fig. 5.11C). All injections were carried out on embryos from two separate clutches in independent experiments, total numbers were pooled to give total numbers. The overexpression phenotype demonstrates that increased expression of *mix1* has the potential to induce morphological changes in the embryo in a dose-dependent manner. Injection of a higher dose of 100pg was lethal to embryos at gastrulation (not shown).

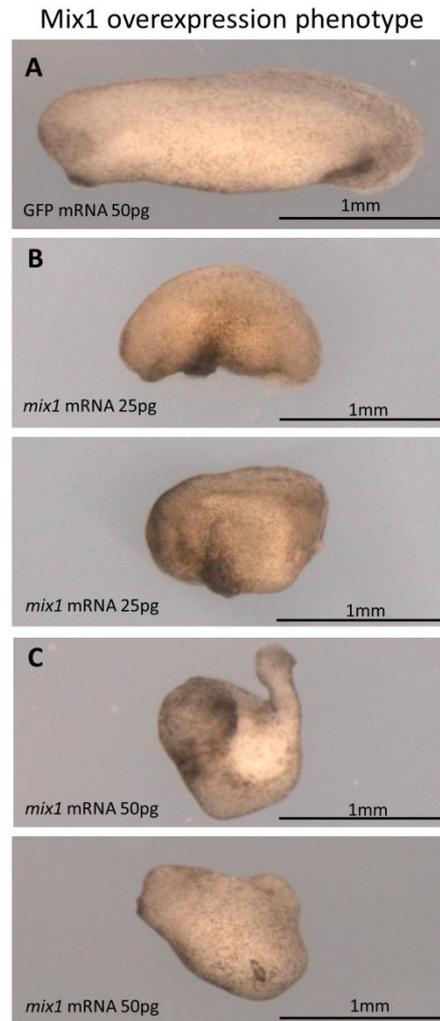


Figure 5.11. Overexpression of *mix1* disrupts development. Lateral view of stage 26 embryos. (A) Embryos injected with 50pg GFP mRNA appear normal. (B) Embryos injected with 25pg *mix1* mRNA were truncated along the antero-posterior axis and had shortened tails (37/40). (C) Embryos injected with 50pg of *mix1* mRNA failed to undergo elongation of the antero-posterior axis (43/43) and in many cases embryos did not form a head (27/43). Embryos from two separate clutches were injected with each dose in independent experiments.

5.2.7 The Mix1 splice blocking MO knockdown phenotype and downstream gene expression perturbations were not rescued

5.2.7.1 The Mix1 splice blocking MO knockdown phenotype was not rescued

To validate that the effects of Mix splice blocking MO-injection were caused by the depletion of Mix1, I attempted to rescue the phenotype and gene expression changes found in

knockdown embryos by co-injecting *mix1* mRNA with the MO. A range of doses of *mix1* mRNA were injected alongside 10ng splice blocking MO, from 2.5pg-100pg. The mRNA was resistant to the MO as it contained no introns. When the Mix1 splice blocking MO and mRNA were injected into embryos, no doses rescued the morphant phenotype. For the lower doses of *mix1* mRNA 2.5pg, 5pg and 10pg embryos appeared similar to MO-injected embryos (not shown). At the higher doses of *mix1* mRNA 15pg, 25pg, 50pg and 100pg, embryos were more defective than morphant embryos and the rate of lethality was higher. (Observed phenotypes at 24 hpf- 10ng MO 40/45 truncated, 5/45 dead, 10ng MO + 2.5pg mRNA 26/26 truncated, 10ng MO + 5pg mRNA 31/32 truncated, 1/32 dead, 10ng MO + 10pg mRNA 28/31 truncated, 3/31 dead, 10ng MO + 15pg 25/36 truncated, 11/36 dead, 10ng MO + 25pg 30/49 truncated, 19/49 dead, 10ng MO + 50pg 32/32 dead, 10ng MO +100pg 34/34 dead.)

5.2.7.2 The downstream gene expression perturbations in Mix1 knockdown embryos were not rescued

Following the failure of phenotypic rescue, I tried the rescue experiment again and measured the transcriptional output of several candidate Mix1 targets identified in the splice blocking MO time-series to see if their expression levels could be restored towards control expression levels. Even a partial rescue of these downstream effects would indicate the specificity of the MO. Embryos were injected with Mix1 splice blocking MO, Mix1 splice blocking MO +25pg mRNA, Mix1 splice blocking MO +50pg mRNA or Mix1 splice blocking MO +100pg mRNA and harvested at 9.0 hpf. This timing coincides with the end of the time-series when the difference in expression between the knockdown and control conditions is at the maximum for these two genes. I used qPCR to measure the fold change difference in the expression level of candidate Mix1 targets between rescue and knockdown embryos (Fig 5.12). The fold changes of *cer1*, *zic3*, *crx*, *tnfrsf21*, *ppp1r10*, *szl* and *ventx3.2* expression were below 1.6 and a Student's t-test fails to find a significant difference between the splice blocking MO-injected and attempted rescue embryos. Three biological replicates of this experiment were performed. This experiment confirms that Mix1 target expression is not rescued in these experiments. There is no validation that depletion of Mix1 in MO-injected embryos is the cause of phenotypic or downstream expression changes.

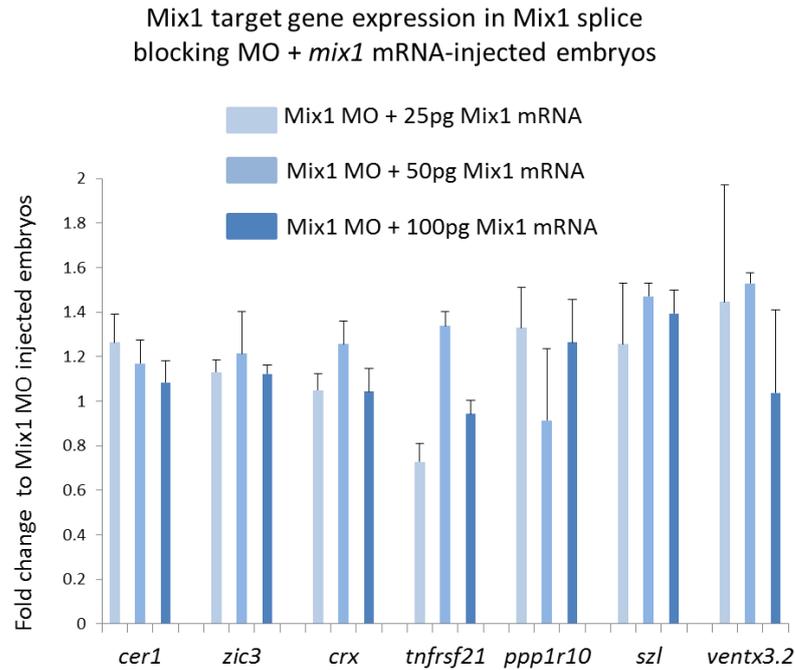


Figure 5.12. Candidate Mix1 target expression was not rescued. Mix1 rescue was attempted, qPCR shows no significant fold change difference in the expression levels of the candidate Mix1 activatory targets *cer1*, *zic3*, *crx* and *tnfrsf21*, and the candidate Mix1 repressive targets *ppp1r10*, *szl* and *ventx3.2* between MO-injected and MO + Mix1 mRNA-injected embryos at three doses ($n=3$ biological replicates). The Student's t-test was applied to MO and MO+mRNA concentrations normalised to *odc1* to show no significant difference. p-values: all >0.05 .

5.2.8 The expression of *mix1* is perturbed in the Mix1 knockdown time-series

The expression of *mix1* is up-regulated in the splice blocking MO time-series, indicating the possible loss of a negative feedback loop (Fig. 5.13). This could be regulated by direct feedback where Mix1 normally represses its own transcription or could be regulated by indirect feedback, where a Mix1 activatory target represses the transcription of Mix1. Mix1 falls into the list of candidate repressive targets because it is up-regulated in the knockdown time-series.

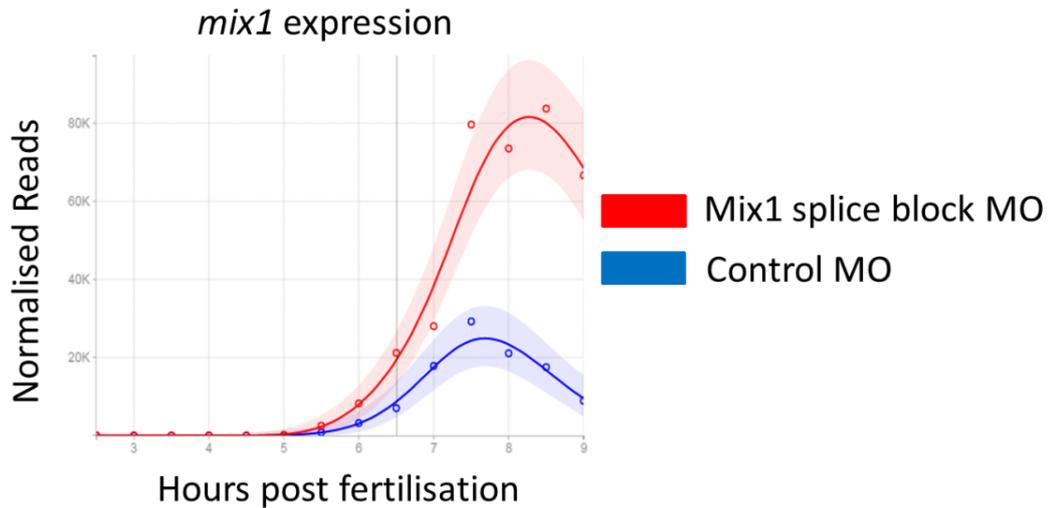


Figure 5.13. Mix1 is up-regulated in knockdown embryos. Time-series expression profile of *mix1* in splice blocking MO time-series reveals up-regulation of *mix1* expression in knockdown embryos compared to controls.

5.3 Discussion

5.3.1 Time-series analysis reveals candidate Mix1 targets

This work demonstrates the effectiveness of knockdown time-series data for identifying gene expression changes. Time-series sequencing reveals the timing and magnitude of these expression changes. The methodology for collecting synchronous embryos and determining gene divergences is established here. Despite the failure of the rescue experiment, I identified 497 candidate Mix1 targets. This work forms the basis for further experiments described in Chapter 6. In order to control for possible off-target effects, I generated two new Mix1 MO time-series using different MOs. To validate candidate Mix1 targets, *mix1*-expressing animal caps were sequenced (see Chapter 6). It should be noted that some of the candidate Mix1 targets mentioned in this chapter are not validated in subsequent experiments.

5.3.2 Mix1 knockdown phenotype

5.3.2.1 The Mix1 splice blocking MO phenotype

Mix1 splice blocking MO-injected embryos appear similar to UV irradiated embryos, which are ventralised as shown on the dorso-anterior deficiency index at 3 and 4 and are

characterised by axis deficiency and reduced eyes and forehead or cyclopia (Kao and Elinson 1988). The ventralised phenotype is also reminiscent of the ventralised phenotype found in *chordin* knockdown embryos (Oelgeschlager et al. 2003). Interestingly, *chordin* is not differentially expressed in either Mix1 knockdown time-series. The ventralised phenotype of Mix1 knockdown embryos may be caused by the reduced expression of other genes which are expressed by the Spemann organiser in the dorsal mesoderm and have dorsalising potential e.g. *gsc* and *cer1* (Cho et al. 1991; Kuroda, Wessely, and De Robertis 2004).

5.3.2.2 Differences to the Mix1 knockdown phenotype found in *X. laevis*

The phenotypes generated here with a Mix1 splice blocking MO appear to be different from the Mix1 knockdown phenotype found in *X. laevis* embryos injected in the dorsal marginal zone with a translation blocking MO. These embryos developed abnormal heads at the tailbud stage but no ventralisation or blastopore defects were reported (Colas et al. 2008). An MO targeting the *X. laevis* homeologs, Mix1 and Mix2, was injected and no phenotypic defects were reported, however it seems as though this MO was not injected into the dorsal marginal zone to test for head defects. It is therefore possible that Mix1 and Mix2 knockdown would have the same effect on head development as Mix1 knockdown. The absence of ventralised phenotypes and blastopore defects in the *X. laevis* study suggests that these defects are specific to my Mix1 knockdown (Colas et al. 2008). The differences between the phenotype for my splice blocking MO and the Mix1 knockdown phenotype found in *X. laevis* may be due to species differences.

5.3.3 Divergences give insight into the dynamic regulation by Mix1

The divergence timings reflect the timing of interactions between Mix1 and its targets. The peak of divergences is found from 6.0 hpf to 8.0 hpf which indicates that Mix1 may have the greatest effect on downstream gene expression after 1.5 hours of its own activation. The timing of these gene divergences is in agreement with previous work showing that Brachyury and Mixer activate targets within 1.5-3.5 hours of their own activation (Collart et al. 2014). A major advantage of the time-series approach is that we gain a unique insight into the timing of transcription factor interactions. Some of these targets have to be direct targets and the earlier targets are more likely to be direct. However, this cannot be determined from this data. Further experiments would be needed to determine whether Mix1 targets are direct, e.g. reporter assays or CHIP experiments.

5.3.4 Reproducibility of Mix1 target expression perturbations in splice blocking MO-injected embryos

The fold changes found in the qPCR data comparing knockdown and control expression of candidate Mix1 targets were consistent with fold changes found in the RNA-seq data. The gene expression perturbations observed in the time-series can be closely replicated which suggests that the changes in gene expression are not the effects of variation between clutches of embryos - but effects caused by MO injection.

The *in situ* hybridisation images of Mix1 half-depleted embryos show that Mix1 activatory targets underwent expression changes within the normal *mix1* expression domain, with a reduction in expression which is consistent with perturbations seen in the RNA-seq time-series (Fig. 5.10). This is consistent with the idea that these are true Mix1 targets, although it does not validate the specific gene interactions because the same MO is used in the qPCR and RNA-seq experiments. I observed the apparent weak depletion of *crx* expression in the ectoderm on the periphery of the *mix1* expression domain. *Crx* may be a direct target of Mix1, and *mix1* may be weakly expressed outside of the visible expression domain. *Crx* could be depleted indirectly in the ectoderm by signalling molecules and transcription factors downstream of Mix1. The differential expression of *crx* could be an off-target effect. In subsequent experiments (see Chapter 6), *crx* is upregulated in *mix1*-expressing animal caps, but is not differentially expressed in all MO knockdown experiments, so it unclear whether *crx* is a genuine Mix1 target.

5.3.5 Mix1 overexpression

The overexpression phenotype is similar to one previously found, in which embryos undergo a reduction in the dorsal axial structures notochord and muscle (Lemaire et al. 1998). In contrast to Lemaire et al. who found that tail formation was completely inhibited in *mix1*-overexpressing embryos, I found that tail formation was only disrupted. This difference in phenotype may be because Lemaire et al. injected dorsally at the 4-cell stage to give localised, ectopic expression, whereas I injected *mix1* mRNA vegetally at the 1-cell stage (Lemaire et al. 1998). More localised injection of *mix1* mRNA at later stages may have been a better way to test the effects of *mix1* overexpression.

5.3.6 Failure of Mix1 rescue

There is no commercially available antibody for *Xenopus* Mix1. Without a Mix1 antibody, it was not possible to test for the depletion of Mix1 protein in MO-injected embryos. The

validation of Mix1 splice blocking MO through rescue would have confirmed both the specificity of the phenotype and the correct dosing of the MO. Without these control experiments, there is a possibility that Mix1 splice blocking MO may incorrectly target transcripts, leading to off-target effects.

There are several possible reasons for the failure to rescue Mix1 knockdown embryos. When injecting mRNA into an embryo it does not disperse efficiently, so it is difficult to recapitulate the wild type localisation of transcripts. MOs disperse more effectively than mRNA within the embryo, so knockdown is easier than rescue. As *mix1* is expressed throughout the marginal zone in the vegetal hemisphere, the mRNA was injected into the vegetal region -but may have been insufficiently distributed in the marginal zone to rescue the effects of Mix1 depletion. The timing and dose of endogenous gene expression are difficult to replicate through microinjection of mRNA. *Mix1* is rapidly activated at 4.5 hpf, so injecting *mix1* mRNA into the stage 1 embryo allows aberrant expression of *mix1* for over four hours. This could disrupt early gene expression and developmental processes. Injection of the *mix1* mRNA causes embryonic defects at certain doses and therefore *mix1* mRNA may not be suitable for rescue. The phenotype is more severe in “rescue” embryos than knockdown embryos leading to earlier lethality in 100% of embryos.

I assayed the expression of several Mix1 candidate target genes in “rescue” embryos, and found that gene expression perturbations were not rescued. However, the qPCR was only carried out on embryos injected with 25pg and 50pg of Mix1 mRNA which caused severe phenotypes. The dose for rescue was most likely disruptive to development and given more time, I would attempt qPCR for embryos injected with lower doses of mRNA with Mix1 MO.

5.3.7 The limitations of sequencing whole embryos

As I sequenced whole embryos, I could not detect whether expression changes were localised in a particular region of the embryo. In the case of *irx1* the gene expression profile that shows two “waves” of activation in the knockdown and control time-series (Fig. 5.6A), I cannot determine whether these are two separate activation events in different regions of the embryo or spatially identical events. It is also possible that small localised changes in expression occur, but are too small to be detected when the whole embryo is sequenced. This is a limitation of this approach which has been addressed in zebrafish by taking fine sections of the embryo across three body axes and sequencing the sections to generate a 3D digital image of gene expression (Junker et al. 2014). This provides rich spatial information

for each gene at three developmental stages. If used in knockdown embryos, the specific localised changes in expression could be accurately measured - however this would be a very expensive and time-consuming method to characterise knockdowns.

5.3.8 Conclusions

In this chapter I used a time-series sequencing approach to identify a number of candidate Mix1 targets. In order to determine whether the expression changes found in knockdown embryos are caused by the depletion of Mix1 rather than off-target effects, the candidate Mix1 targets identified here must be compared to other datasets (see Chapter 6).

Chapter 6 – Controlling for MO off-target effects to determine Mix1 targets

6.1 Introduction

In this chapter I build on data described in Chapter 5 where I used a Mix1 splice blocking morpholino (MO) to identify several hundred candidate activatory and repressive Mix1 targets. This set of targets may include off-target effects caused by the MO. The strategy in this chapter was to generate additional lists of Mix1 targets using different MOs to create knockdown and control time-series. The assumption is that only genuine Mix1 targets will appear in all lists of targets.

I used two different MOs which target Mix1 and the related transcription factor Mixer; translation blocking MO1 and translation blocking MO2. These two MOs were originally intended for identifying the targets of these two transcription factors separately, however, it came to light that the 5'UTR and initial part of the coding sequence is 100% identical for Mix1 and Mixer (Fig. 6.1). This region covers all possible binding sites for translation blocking MOs in both genes. Each MO was only tested for the capacity to block translation of both of these transcription factors but given the sequence similarity the assumption is that the two MOs target both Mix1 and Mixer.

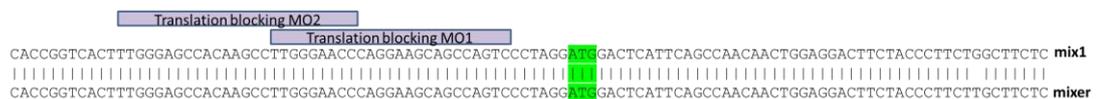


Fig. 6.1. Mixer and Mix1 share 5'UTR sequences. The alignment of the *mix1* (upper) and *mixer* (lower) cDNA sequences which have 100% identity over the 5'UTR and initial coding sequence. Binding sites of the translation blocking MO1 and MO2 are illustrated. The start codon is highlighted in green.

This sequence similarity of the Mix1 and Mixer 5'UTRs poses challenges, but I use these MOs to identify candidate Mix1 targets which are then used in comparative analysis with the Mix1 splice blocking MO data. The expectation for these two translation blocking MOs is that they will each block the translation of both Mix1 and Mixer. It is also possible that each translation blocking MO will cause off-target effects and it is likely that these off-target effects will be MO-specific. Genes that are consistently differentially expressed in both translation blocking

MO1 and MO2 time-series are candidate Mix1 and/or Mixer targets. Genes that are differentially expressed in just one MO time-series may be off-target effects.

The candidate Mix1 and/or Mixer targets identified in the comparison of the two translation blocking MO time-series were then compared to the candidate targets identified in the Mix1 splice blocking MO time-series. The assumption is that the Mix1 targets identified using the splice blocking MO will contain Mix1 targets and off-target genes. Genes that are consistently differentially expressed in all three MO time-series are therefore strong candidate Mix1 targets. These genes are assigned to Group 1 (See table 6.1). This strategy should eliminate off-target effects because the three MOs have different sequences and therefore should not target the same off-target sequences. I found statistically significant numbers of genes consistently differentially expressed in all three MO time-series

In order to validate the candidate Mix1 targets identified in the three MO time-series, *mix1*-expressing animal caps were sequenced. If the differential expression of a gene was in agreement in all experimental conditions tested, which includes loss-of-function and gain-of-function analysis, I think that this is strong evidence that this gene is a true target of Mix1 and I refer to these genes as validated Mix1 targets. By injecting *X. laevis* animal caps with *mix1* mRNA, Mix1 protein is produced and can induce or up-regulate expression of Mix1 activatory targets. Likewise, the expression of Mix1 repressive targets which are endogenously expressed in the animal cap cells may be down-regulated. Limitations of this assay are that the timing of the expression of *mix1* and the environment e.g. co-factors and chromatin state are not recapitulated in the animal cap, so some Mix1 targets might not be mis-expressed in animal caps.

A list of candidate Mix1 targets was generated based on differential expression detected in the animal cap sequencing data. These candidate Mix1 targets were compared to the strong candidate targets identified by comparing the three MO knockdown time-series. Nine genes were consistently differentially regulated in all four experimental conditions are validated Mix1 targets which is significantly more than would be expected by overlap between conditions by chance. These validated Mix1 targets were assigned to Group 2 (see table. 6.1). This is because their differential expression has been demonstrated in response to Mix1 depletion and *mix1* expression using different experimental approaches.

Group	No. Genes	Target strength	Overlap criteria
Group 1 targets	33	Strong candidate Mix1 targets	Consistently differentially expressed in all three MO time-series.
Group 2 targets	9	Validated Mix1 targets	Consistently differentially regulated in all three MO time-series and <i>mix1</i> -expressing animal caps.

Table 6.1. Group 1 strong candidate and Group 2 validated Mix1 targets. Groups of strong candidate and validated Mix1 targets identified through overlap of genes differentially expressed in different experimental conditions. The full list of Group 1 and Group 2 genes can be found in the appendix on the attached CD (tab 6).

6.2 Results

6.2.1 Translation blocking MO1 validation

In order to test the activity of the translation blocking MO1, an *in vitro* assay was performed using the TNT transcription/translation system. A *mix1* HA-tagged vector was used to generate Mix1 protein. A series of reactions was set up, with increasing concentrations of Mix1 translation blocking MO and western blot was used to detect the HA-epitope. With the addition of MO, the intensity of the band for HA-tagged Mix1 was reduced compared to the MO free condition in a dose dependent manner. At the highest concentration of 40mM, the Mix1-HA band disappears almost entirely. This assay demonstrates the ability of MO1 to block translation of Mix1 (Fig. 6.2).



Figure 6.2. Translation blocking MO1 activity. Tagged *mix1*-HA western blot of translation blocking MO1 activity demonstrated by *in vitro* SP6 coupled transcription/translation (TNT) assay. Translation of HA-tagged *mix1* vector produces a band of Mix1 protein at 49kDa. Intensity of the band decreases with increasing concentrations of MO.

6.2.2 Translation blocking MO2 validation

The translation blocking activity of the translation blocking MO2 was tested *in vivo* by injecting HA-tagged *mixer* mRNA into embryos and with increasing doses of the MO. Mixer depletion was tested rather than Mix1, because the MO was originally intended to target *mixer*. Embryos were harvested at 7.0 hpf and an immunoprecipitation was performed with an HA antibody followed by western blot to assay the level of translation blocking activity by the MO at different doses. I found that at the 15ng dose, no band for tagged Mixer protein was visible, indicating efficient translation blocking. At lower doses (5ng and 10ng) the amount of tagged Mixer protein was lower than in the MO-free condition, indicating some translation blocking activity (Fig. 6.3A). This demonstrates the ability of the MO to block the translation of Mixer *in vivo*. The dose of 15ng was sufficient to deplete the injected HA-tagged *mixer* mRNA in the presence of endogenous *mixer* mRNA. It is therefore likely that a 15ng dose is sufficient to block translation of endogenous Mixer and Mix1 protein in the absence of injected mRNA.

In addition, an *in vitro* assay was performed using the TNT transcription/translation system to test the MO translation blocking activity. As described for the translation blocking MO1, increasing concentrations of translation blocking MO2 were used to set up several reactions to produce HA-tagged Mixer protein and western blot was used to detect the HA-epitope. I found that higher concentrations (20nM and 40nM) were sufficient to entirely block translation of Mixer (Fig. 6.3B).

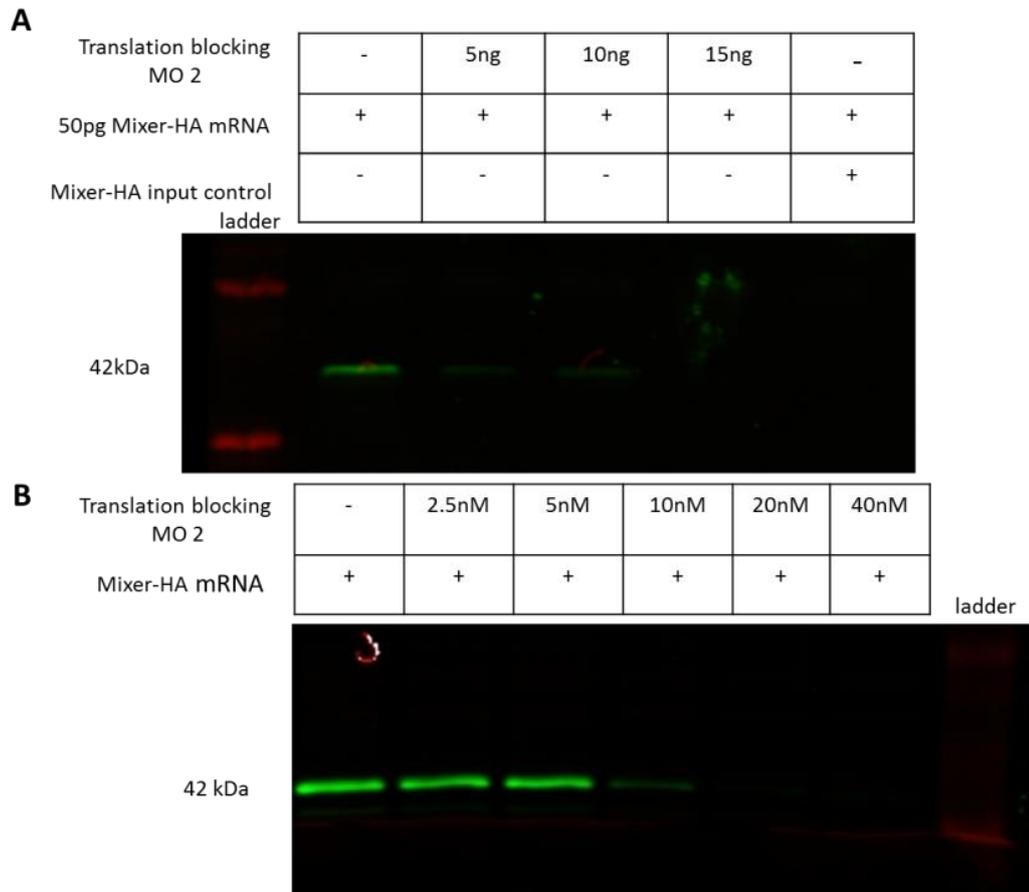


Figure 6.3. Translation blocking MO2 activity. (A) Tagged *mixer*-HA immunoprecipitation and western blot of translation blocking MO2 activity *in vivo* by injection of *mixer* mRNA and increasing doses of translation blocking MO2. A non-immunoprecipitated input sample in the far right lane gives no band. (B) Tagged *mixer*-HA western blot of translation blocking MO2 activity demonstrated by in vitro SP6 coupled transcription/translation (TNT) assay. Translation of HA-tagged *mixer* vector produces a band of Mixer protein at 42kDa. Intensity of the band decreases with increasing concentrations of MO.

6.2.3 Phenotypes generated by the two MOs

I injected different doses of each translation blocking MO into embryos to determine the appropriate dose to use in later time-series experiments. Of the three doses tested for each MO, 15ng was found to give the highest penetrance of phenotype. The 15ng dose was used in later validation and time-series experiments as this likely indicates efficient knockdown of Mix1.

Embryos were injected with translation blocking MO1; at a 15ng dose 45/50 (90%) were truncated with a blastopore closure defect, at a 10ng dose 12/56 (21%) were truncated and none had blastopore closure defects, at a 5ng dose 16/57 (28%) were truncated and 2/57

(4%) also had blastopore defects (Fig. 6.4A, C). Morpholino injection was lethal to defective embryos during the late tailbud stage. Embryos were injected with translation blocking MO2, this generated a truncated phenotype and embryos displayed head and tail defects; at a 15ng dose 35/43 (81%) were truncated and 27/43 (63%) had a blastopore closure defect, at a 10ng dose 27/48 (56%) were truncated and 10/48 (21%) had blastopore closure defects, at a 5ng 17/33 (51%) were truncated and none had a blastopore closure defect. Morpholino injection was lethal to defective embryos during the late tailbud stage (Fig. 6.4B, C). All injections were carried out on embryos from two separate clutches in independent experiments, total numbers were pooled to give penetrance for each MO.

The phenotypes generated by translation blocking MO1 and MO2 were similar and highly penetrant at the 15ng dose so this dose was used in further experiments. This dose of MO2 was sufficient to block in vivo translation of injected *mixer* mRNA, so it is likely that 15ng is sufficient to block translation of endogenous Mix1 and Mixer.

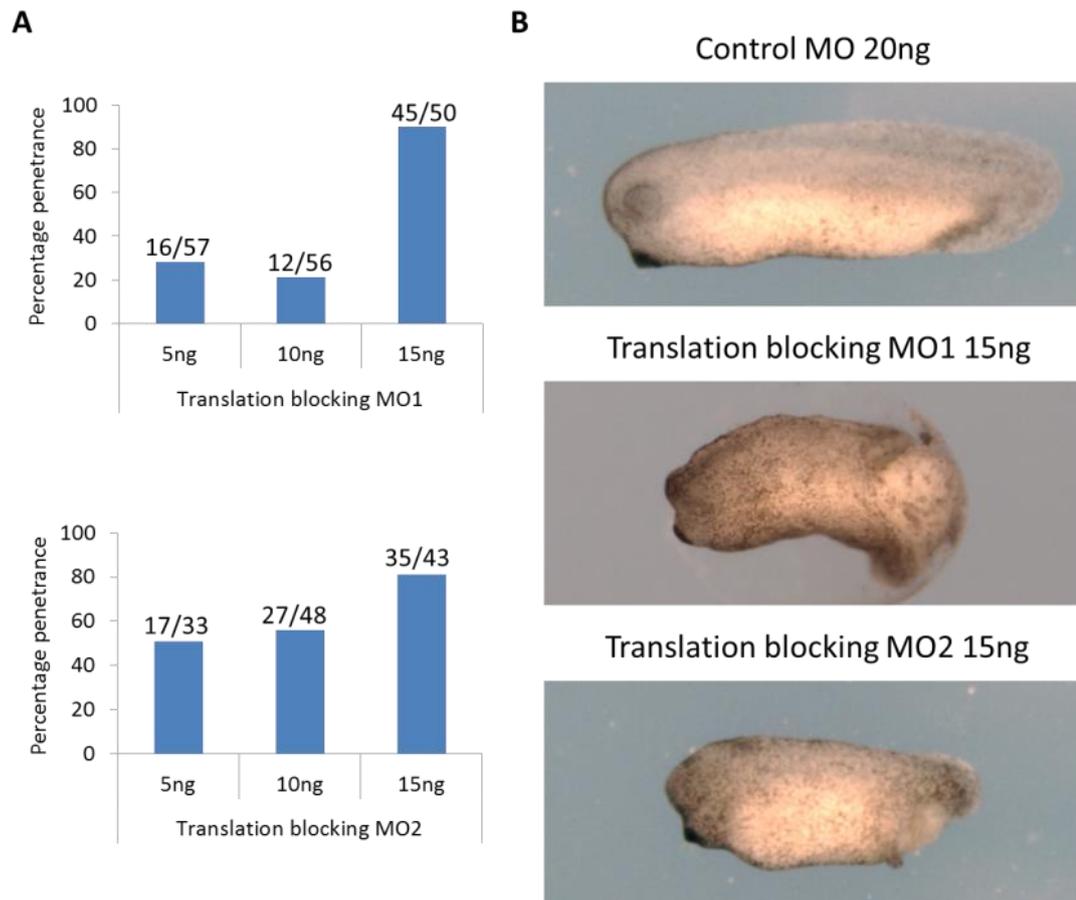


Figure 6.4. Phenotypes induced by translation blocking MO1 and MO2. (A) Increasing doses of translation blocking MO1 and MO2 were injected and the penetrance of the truncated phenotype was recorded. (B) Embryos injected with 20ng control MO have a normal phenotype. Embryos injected with 15ng translation blocking MO1 and MO2 have a truncated phenotype and head and tail defects. MO1 and MO2 were injected at the three doses into two separate clutches of embryos in independent experiments, and the results were pooled to give the penetrance for each dose.

6.2.4 Time-series analysis of gene expression in translation blocking MO1 and MO2-injected embryos

The Mix1 splice blocking MO knockdown time-series data was used to identify candidate Mix1 targets (see Chapter 5). To control for off-target effects caused by the splice blocking MO, the two Mix1/Mixer translation blocking MOs were used to generate two additional RNA-seq time-series. As before, embryos were synchronised and sampled at 30-minute intervals from a single clutch for knockdown, control and wild type conditions in each time-

series. In the translation blocking MO1 time-series, embryos were collected from 5.5 hpf to 9.0 hpf. In the translation blocking MO2 time-series, embryos were collected from 4.0 hpf to 10.0 hpf (Fig. 6.5). The MO1 time-series covered fewer time-points because of lower numbers of embryos in the clutch used for collection. The MO2 time series is longer as there were more embryos in that particular clutch and this allowed me to extend the time-series so I might identify later diverging genes.

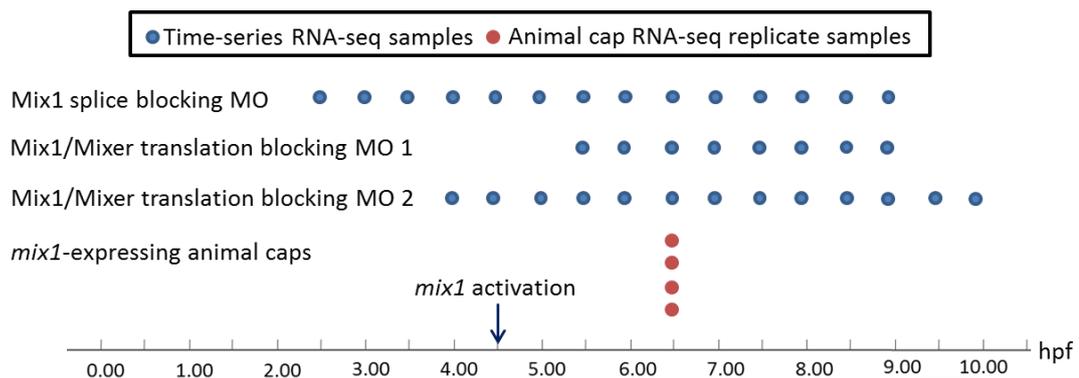


Figure 6.5. Time-series and animal cap collection times. The time points collected for the three MO knockdown time-series are marked on the time-scale (hpf). The *X. laevis mix1*-expressing animal caps were harvested at stage 9 which is equivalent to 6.5 hpf in *X. tropicalis* (see below for details of animal cap experiment).

Instead of using a 5-base mismatch MO as used for the for the Mix1 splice blocking time-series in Chapter 5, a standard control MO from Gene Tools was used for both time-series. This is because the 5-base mismatch MO for the Mix1 splice blocking MO was found to have some splice blocking activity in the RNA-seq read alignments (see Chapter 5). Because of this I was concerned that any 5-base mismatch MOs designed for the translation blocking MOs would have some translation blocking activity. Wild type un-injected embryos were collected as an additional control in each time-series. As described for the splice blocking MO time-series, RNA was sequenced, reads aligned, normalised and analysed for differential expression (see Materials and Methods for detail). The wild-type and control MO expression profiles were similar in both translation blocking MO time-series. The control profiles were used in the differential expression analysis.

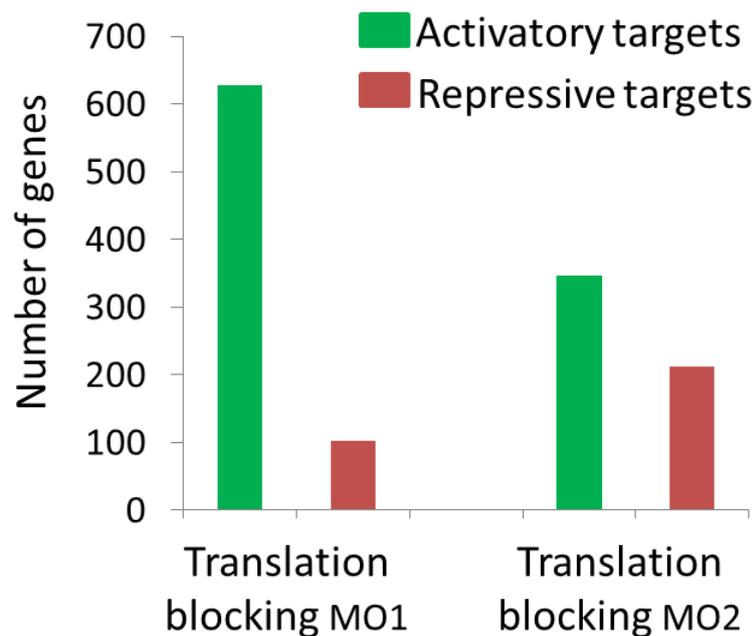


Figure 6.6. Total differentially expressed genes in the translation blocking MO1 and MO2 time-series. Numbers of differentially expressed genes in the translation blocking MO1 and MO2 time-series. 628 candidate activatory and 103 candidate repressive targets were identified in the translation blocking MO1 time-series. 346 candidate activatory and 212 repressive targets were identified in the translation blocking MO2 time-series. These experiments were each carried out in a single clutch of embryos.

In the translation blocking MO1 time-series 731 genes were detected with differential expression in knockdown embryos compared to controls. Of these genes, 628 were down-regulated after divergence and 103 were up-regulated after divergence (Fig. 6.6). In the translation blocking MO2 time-series 558 genes were differentially expressed, of these 346 were down-regulated after divergence and 212 were up-regulated after divergence (Fig.6.6). The full lists of differentially expressed genes can be found in the appendix on the attached CD (tab 3 and 4).

There are clear differences in the number of targets differentially expressed in two different translation blocking MO time-series and differences in the proportions of activatory and repressive targets. In total 81% of all differentially expressed genes in the MO1 time-series and 57% of all differentially expressed genes in the MO2 time-series were not differentially expressed in the other translation blocking MO time-series.

Because the MO1 time-series was sampled from after *mix1* was activated, it was not possible to test whether control MO and MO1 expression profiles were equal prior to *mix1* activation.

For this reason, different parameters had to be applied, which required the confidence intervals of the control MO and MO1 data points to be within close proximity at the start of the time-series (see Materials and Methods for details). This contributes to there being more differentially expressed genes in the MO1 compared to the MO2 time-series. I found more genes differentially expressed at later time points compared to the MO1 time-series because the MO2 time-series extends by an extra two time-points. This has an effect on the overall numbers of differentially expressed genes in each time-series. However, these non-overlapping differentially expressed genes may reflect that off-target effects are prevalent in each MO knockdown time-series.

6.2.5 Comparison of the two translation blocking MO time-series

As a control for off-target effects, the differentially expressed genes from the two translation blocking MO time-series were compared, as it is unlikely that two different MOs targeting different sequences would influence the same off-target gene expression. 126 overlapping activatory targets and 15 overlapping repressive targets were found (Fig. 6.7). There are significantly more genes overlapping than would be expected to overlap by chance i.e. if all effects were off-target: only 14.4 activatory targets and 1.4 repressive targets would be expected to overlap by chance in the two conditions (8.8x and 10.7x enrichment of activatory and repressive targets respectively, chi square test $p < 0.0001$). The consistent regulation by the two MOs indicates that these genes are likely to be targets of Mix1, Mixer or both transcription factors, rather than off-target effects.

Genes up-regulated in both time-series are candidate activatory targets of Mix1 and/or Mixer, and genes down-regulated in both time-series are candidate repressive targets of Mix1 and/or Mixer. Only the genes which were differentially expressed in both translation blocking MO time-series were included in further analysis to investigate which genes might be Mix1 targets. This was achieved first by comparison to the Mix1 specific splice blocking MO data and then to *mix1*-expressing animal cap sequencing data.

6.2.6 Comparison to the splice blocking MO time-series

To investigate which genes identified in the Mix1 splice blocking MO time-series (see Chapter 5) are Mix1 targets, I compared these genes to the candidate Mix1 targets identified in the two Mix1 translation blocking MO time-series. I found 32 activatory targets and 3 repressive targets which are consistently differentially expressed in the three MO time-series (Fig. 6.7). These overlaps are significantly higher than would be expected by chance in the three time-

series; for activatory targets 0.16 genes would be expected to overlap at random and for repressive targets 0.03 genes would be expected to overlap at random (200x enrichment of activatory targets and 115x enrichment of repressive targets). Interestingly, 6 genes are found in all three MO time-series with conflicting directionality in one time-series. As only 0.79 genes would be expected to overlap by chance, these genes are 7.6x enriched which is significantly more than would be expected to overlap by chance, but is considerably less enrichment than is found for the activatory and repressive overlapping targets.

The overlapping genes from the three knockdown time-series experiments are strong candidate Mix1 targets, as they exhibit consistent differential expression in three different experimental conditions. These strong candidate Mix1 targets are assigned to Group 1 (See table 6.1). Most of the Group 1 strong candidate Mix1 targets are activatory targets and this suggests that Mix1 primarily functions as a transcriptional activator.

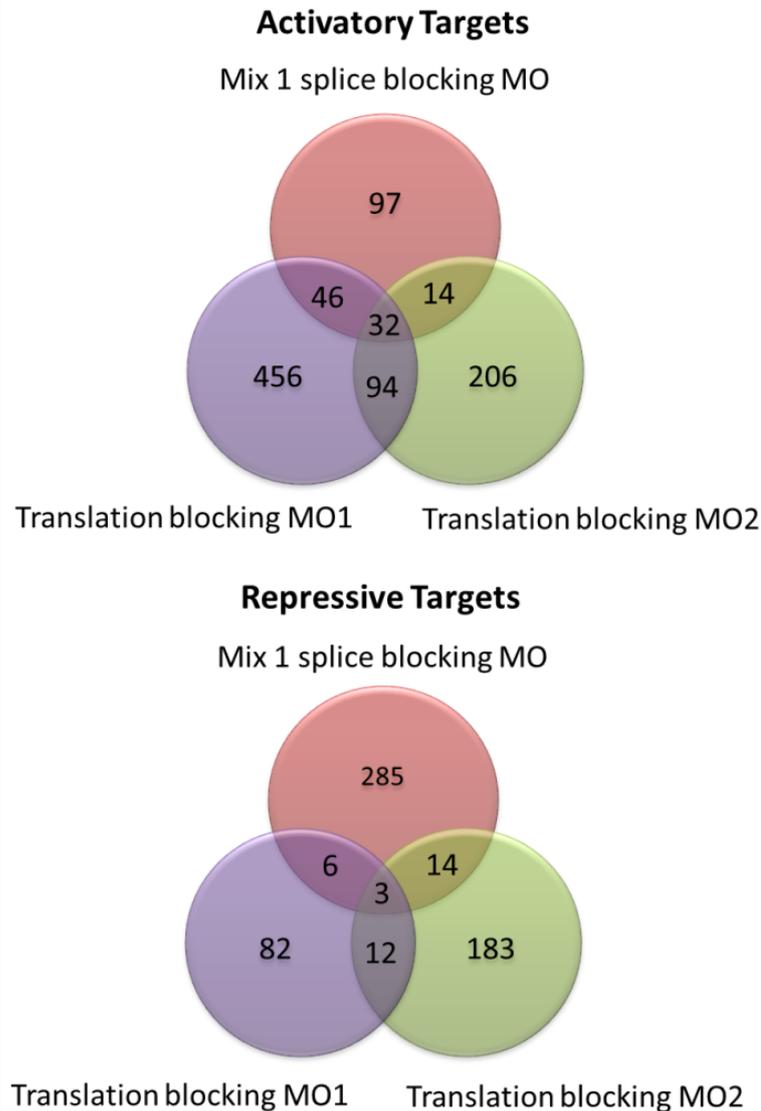


Figure 6.7. Comparison of candidate targets from the three MO time-series. Total activatory and repressive candidate Mix1 target genes differentially expressed in the splice blocking MO time-series and translation blocking MO1 and MO2 time-series and overlap between the conditions (Chi square test $p < 0.0001$ for activatory and repressive overlaps in all three conditions).

6.2.7 Mix1-expressing animal caps were sequenced to validate Mix1 target genes

Next I used a gain-of-function approach to determine Mix1 targets using an animal cap assay. For practicality, *X. laevis* embryos were used instead of *X. tropicalis* because slower cell divisions allow more time for animal cap dissection and the larger embryos are easier to dissect. *X. tropicalis mix1* mRNA was injected into *X. laevis* embryos, the animal caps were dissected at stage 8 and then cultured for two hours and harvested during stage 9. For each sample, 30 animal caps were collected from each of four separate clutches to give four

replicates of *mix1*-expressing animal caps but only three replicates of WT caps due to failed RNA extraction of one sample (Fig. 6.5). Mix1 activatory targets are likely to be induced in animal caps in response to ectopic Mix1 protein. Mix1 repressive targets might be down-regulated in animal caps in response to ectopic Mix1 protein, provided they were expressed in animal cap cells in the first place.

A preliminary experiment was performed to check that the dose of *mix1* mRNA injected was sufficient to induce the expression of Mix1 targets in animal caps. The expression of *gsc*, a known Mix1 target from previous work (Latinkic and Smith 1999), was measured by qPCR to confirm a significant up-regulation in *mix1*-expressing animal caps (Fig. 6.8). *Vegt* expression was used as a negative control gene, as this gene was not differentially expressed in the Mix1 splice blocking MO time-series and is expressed vegetally and is not expressed in animal caps (Sudou et al. 2012). As expected, *vegt* was not induced in *mix1*-expressing animal caps (Fig. 6.8). In the wild type embryo, *gsc* is expressed in dorsal-vegetal tissue and *vegt* is expressed throughout the marginal zone at stage 10 (Sudou et al. 2012). CDNA from dorsal-vegetal tissue dissected from stage 10 *X. laevis* embryos was used as a positive control for *gsc* and *vegt* expression. The expression level of *gsc* in *mix1*-expressing animal caps was approximately half as much as in the positive control and *vegt* expression was much lower in *mix1*-expressing animal caps (Fig. 6.8). This indicates that biologically relevant levels of *gsc* were induced by Mix1 in stage 9 animal caps. Expression of the “housekeeping” gene *odc1* was measured and used to normalise expression levels of *gsc* and *vegt*.

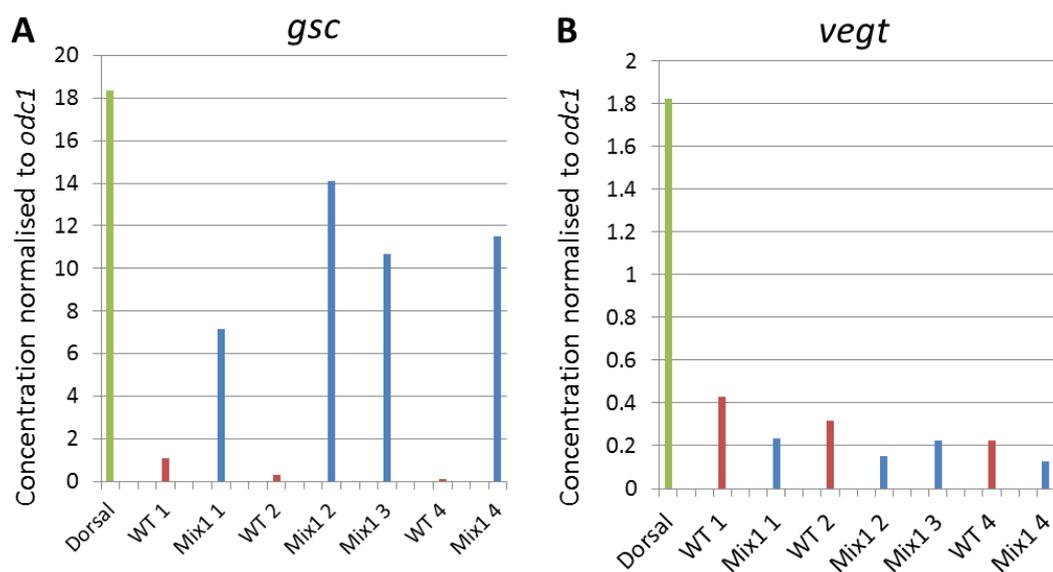


Figure 6.8. Mix1 induces *gsc* expression in animal cap explants. Gene expression in dorsal tissue and animal cap explants (A) *gsc* expression in stage 10 dorsal tissue (positive control) and stage 9 animal caps, either expressing *mix1* (n=4) or wild type (n=3). Expression is induced beyond background level in *mix1*-expressing animal caps. The Student's t-test was applied to Mix1 and WT concentrations to show a significant difference, p-value: 0.0009 (B) *Vegt* expression in cDNA from dorsal tissue from stage 10 embryos and animal caps at stage 9, either expressing *mix1* (n=4) or wild type (n=3). Expression is not induced in animal caps. The Student's t-test was applied to Mix1 and WT concentrations to show no significant difference, p-value: 0.0692.

Following the preliminary experiment which confirmed the induction of the Mix1 target *gsc*, the same animal cap RNA samples used for qPCR were also sequenced. Genes differentially expressed between WT and *mix1*-expressing animal caps are candidate Mix1 targets. The fold changes in expression between the replicate WT and *mix1*-expressing animal caps were calculated. Genes with an FDR<0.1 meet the criteria for differential expression. The FDR<0.1 was used as a threshold to correct for the large number of genes in the data set and allows genes with an expected false discovery rate of 10% to meet criteria, which is acceptable since the genes are to be used for comparative analysis with other data sets. The *X. laevis* differentially expressed genes were then mapped to the *X. tropicalis* genes for comparison with time-series data (analysis by Nick Owens, see Materials and Methods for detail). 236 genes were up-regulated in *mix1*-expressing animal caps and were identified as candidate Mix1 activatory targets. 118 genes were down-regulated in *mix1*-expressing animal caps and were identified as candidate Mix1 repressive targets (Fig. 6.9). The full list of differentially expressed genes can be found in the appendix on the attached CD (tab 5).

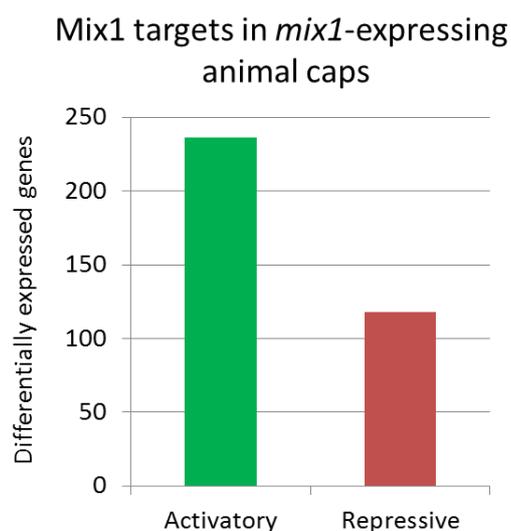


Figure 6.9. Total differentially expressed genes in mix1-expressing animal caps. Numbers of differentially expressed genes detected in RNA-seq of *mix1*-expressing animal caps. 236 candidate activatory and 118 candidate repressive targets were identified. This experiment was carried out in four clutches of embryos (*mix1* mRNA injected) and three clutches of embryos (control un-injected).

6.2.8 Comparison of differentially expressed genes in the three knockdown time-series and *mix1*-expressing animal caps

To validate candidate Mix1 targets, the Group 1 strong candidate Mix1 targets and candidate targets from the animal cap sequencing were compared. Candidate targets which overlap in these two sets are considered to be validated because the perturbed expression of these genes has been demonstrated in response to both depletion of Mix1 protein in whole embryos and to *mix1* expression in animal caps.

Activatory targets are likely to be up-regulated in animal caps, due to activation by Mix1, and down-regulated in MO knockdowns in response to depletion of Mix1. The opposite is true for repressive targets. However, repressive targets must be expressed in the animal cap to be down-regulated by Mix1 in animal cap explants, therefore some repressive targets will not be observed in this data.

Nine genes were consistently differentially regulated in all conditions and therefore I designate these genes as validated Mix1 targets and assign them to Group 2 (See table 6.1). The nine validated genes are all Mix1 activatory targets and demonstrate a significant overlap between the different datasets. Less than one gene would be expected to overlap by chance (Fig. 6.10). There were no repressive targets consistently differentially regulated in all experimental conditions. There are two Group 1 Mix1 targets for which no orthologue in *X. laevis* could be found. Therefore, there could be more Group 2 validated Mix1 targets which cannot be identified in the *X. laevis* animal caps.

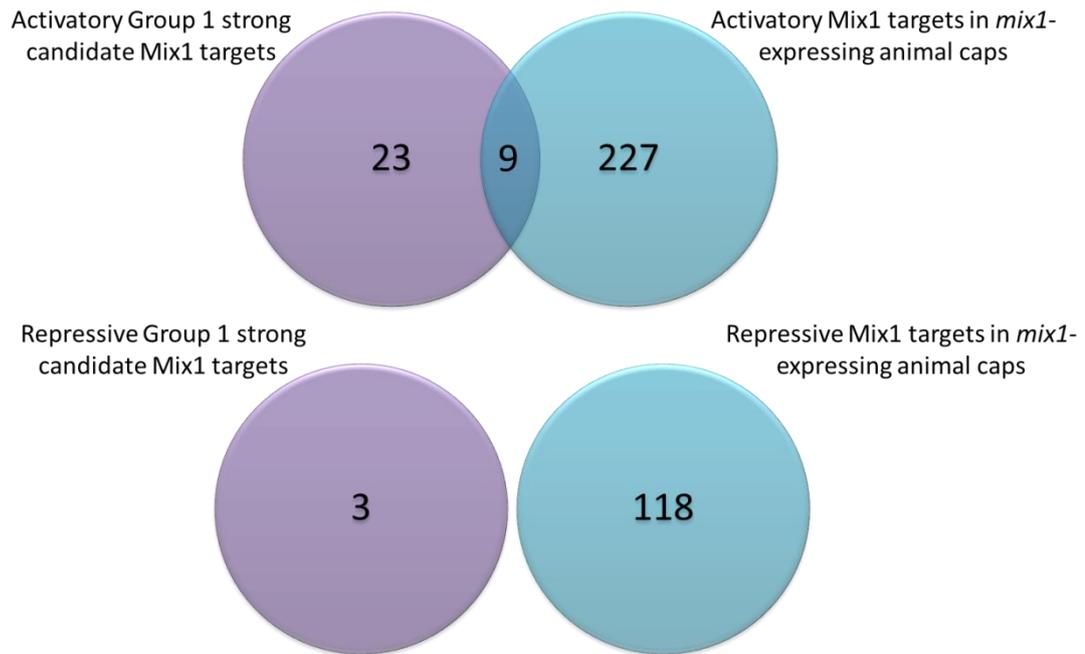


Figure 6.10. Comparison of Group 1 strong candidate Mix1 targets and *mix1*-expressing animal cap targets. Total activatory and repressive Group 1 Mix1 strong candidate targets and genes differentially expressed in *mix1*-expressing animal caps and overlap between the conditions (Activatory target overlap: Chi square test $p < 0.0001$).

As the nine Group 2 targets are the only validated Mix1 targets, these are the focus of later analysis to understand likely functions of Mix1 through GO analysis and literature searching. The validated targets are all activatory targets, six of these are transcription factors; *arx*, *foxc1*, *irx1*, *irx3*, *phox2b* and *msx2*, two are signalling molecules; *admp* and *cer1* and finally *gjb2* is a gap junction protein (Fig. 6.11).

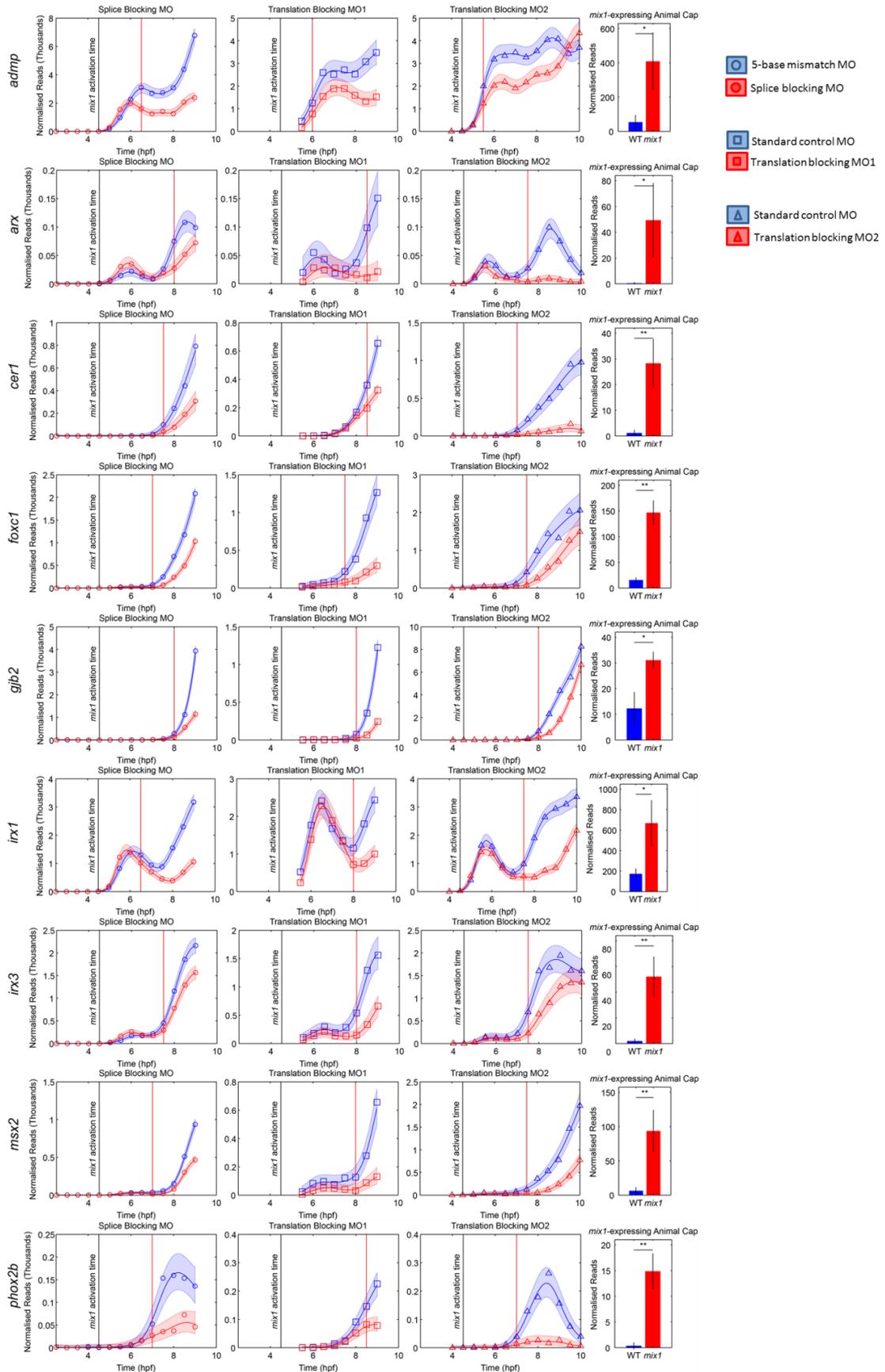


Figure 6.11. Group 2 validated Mix1 target expression profiles. Profiles of validated Mix1 targets which were differentially expressed with consistent differential regulation in the splice blocking MO time-series, the translation blocking MO1 time-series, the translation blocking MO2 time-series and the animal cap sequencing. The nine validated activatory targets *admp*, *arx*, *cer1*, *foxc1*, *gjb2*, *irx1*, *irx3*, *msx2* and *phox2b* are shown. Normalised read counts for Mix1 splice blocking MO (red circles), 5-base mismatch control MO (blue circles), Mix1 translation blocking MO1 (red squares), the standard control MO (blue squares), Mix1 translation blocking MO2 (red triangles) and the standard control MO (blue triangles) were plotted for all time points in the two time-series. Gaussian Process median lines and 95% confidence intervals are shown. Red vertical lines mark divergence times and grey vertical lines mark the activation time of Mix1 at 4.5 hpf. Four replicates of *mix1*-expressing animal caps and three replicates of WT animal caps were sequenced to give average normalised reads. Animal cap error bars mark standard deviation, * and ** mark $p < 0.05$ and $p < 0.01$ respectively. Gene expression profiles and histograms generated by Nick Owens.

Then I examined genes which were not consistently differentially regulated in all four conditions. There are 24 genes in Group 1 that are differentially expressed in all three MO time-series but not in animal caps. These 24 candidate Mix1 targets found in Group 1 are not validated (see discussion).

Mix1 expression is up-regulated in all three MO time-series and is therefore identified as a repressive target. In animal caps *mix1* expression is higher in *mix1*-expressing animal caps than in wild types because the mRNA is injected, so *mix1* is therefore untestable in the animal caps.

6.2.9 Conflicting differential expression of *gsc* in the MO time-series

The Group 1 genes are strong candidate Mix1 targets in the sense that off-target effects are well controlled for. However, an issue with the two translation blocking MOs is that they target Mix1 and Mixer so the depletion of Mixer may alter the expression of Mix1 targets. The expression of the established Mix1 target *gsc* (Latinkic and Smith 1999) is down-regulated in the splice blocking MO time-series and up-regulated in both of the translation blocking MO time-series. Although *gsc* expression does not diverge sufficiently to be considered differentially expressed in the translation blocking MO1 time-series, it seems likely that a longer time-series would reveal differential expression of *gsc* (Fig. 6.12). The opposing regulation of this gene indicates that *gsc* may be a repressive target of Mixer and an activatory target of Mix1.

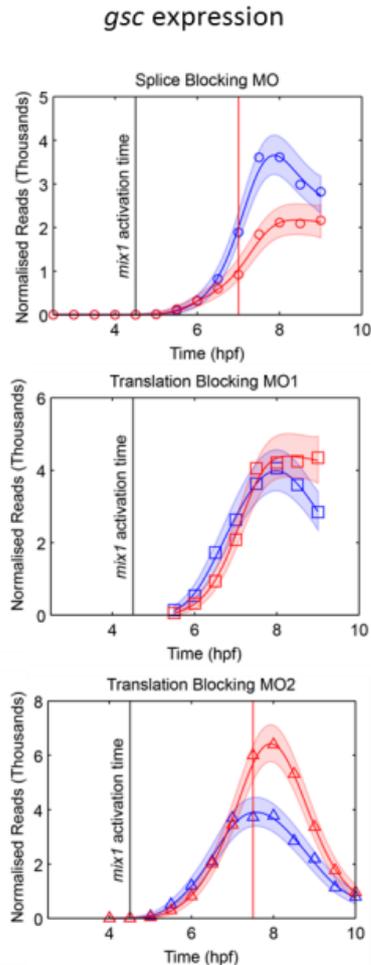


Figure 6.12. Differential expression of *gsc* reveals different regulation in the splice and translation blocking MO time-series. *Gsc* expression is down-regulated in the splice blocking MO time-series indicating activation by Mix1. *Gsc* expression is up-regulated in the MO1 and MO2 time-series, in which both Mixer and Mix1 are knocked out. These opposing effects indicate that Mixer may be responsible for repressing *gsc* expression in wild type embryos. Gene expression profiles generated by Nick Owens.

6.2.10 Mix1 targets are enriched for transcription factors

In order to understand more about the role of Mix1, I investigated the number of transcription factors in the Group 1 and Group 2 Mix1 targets which contain consistently differentially regulated genes (see Materials and Methods for detail). I found significant enrichment of transcription factors in both sets, compared to inconsistently differentially regulated genes (non-overlapping genes) that did not fall in to Group 1 or Group 2. The highest proportion of transcription factors is found in the Group 2 validated Mix1 targets (Fig. 6.13).

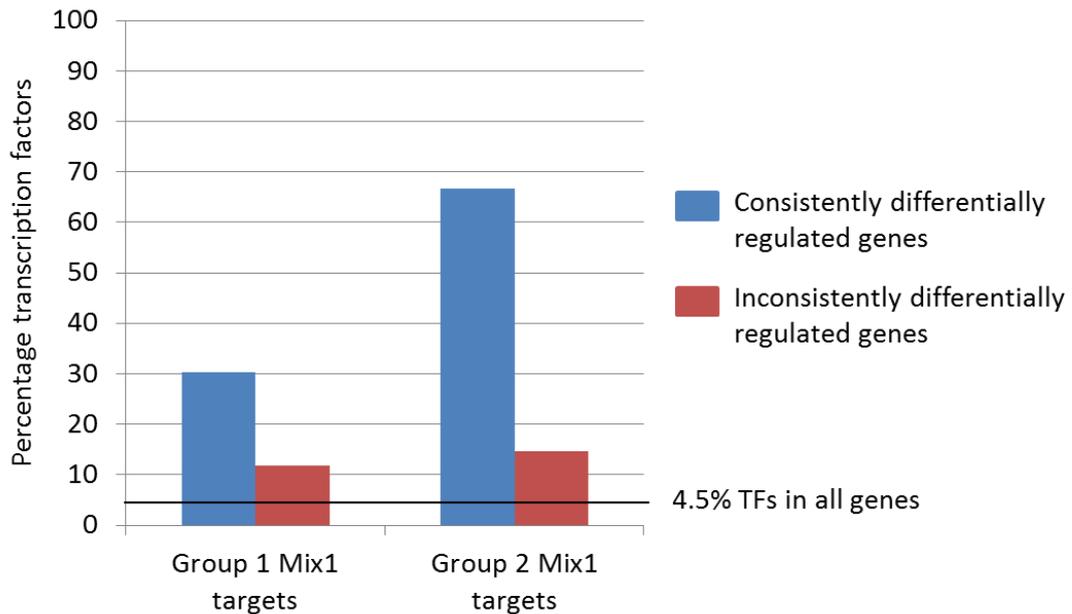


Figure 6.13. Many consistently differentially regulated genes in Group 1 and Group 2 are transcription factors. Percentage of transcription factors within the Group 1 strong candidate Mix1 targets and the Group 2 validated Mix1 targets. Group 1: 10/33 (30%) are TFs, Group 2: 6/9 (67%) are TFs. The number of TFs in the consistently differentially regulated and inconsistently differentially regulated genes was compared, the Fisher's exact two tailed test p values are: Group 1; 0.0045, Group 2; 0.0006. 4.5% represents the percentage of transcription factors present in the *X. tropicalis* transcriptome.

6.2.11 Gene ontology analysis of Mix1 targets reveals potential functions of Mix1

The Mix1 targets were examined using gene ontology (GO) analysis by searching for functional annotations of Mix1 target genes (GO analysis by Nick Owens, see Materials and Methods for detail). The full list of enriched GO terms can be found in the appendix on the attached CD (tabs 7 and 8). Using this information, I can infer the specific functions of Mix1 as a transcriptional regulator. The Group 1 genes were included in the GO analysis even though not all of these genes are validated, this is because they might be absent from the animal cap data for various reasons (see discussion). The comparison between the candidate and validated targets might reveal similarities between the two sets of genes. Group 2 is a subset of Group 1, but some genes which are exclusive to Group 1 are annotated with some of the same gene ontology enrichments. A large number of highly enriched GO terms were found to be enriched in the Group 1 and Group 2 Mix1 targets.

The biological process GO terms enriched in the Group 1 and Group 2 Mix1 targets include; *dorsal/ventral pattern formation, regulation of BMP signalling pathway, mesoderm development, cell fate commitment, anterior/posterior pattern specification, regionalization,*

ossification, brain development, central nervous system development, sensory organ development, organ morphogenesis, cell proliferation and locomotion (Fig. 6.14).

Several GO terms were enriched in Group 1 strong candidate Mix1 targets but not in the Group 2 validated Mix1 targets; *Somitogenesis, gastrulation* and *regulation of Wnt receptor signaling pathway* (Fig. 6.14).

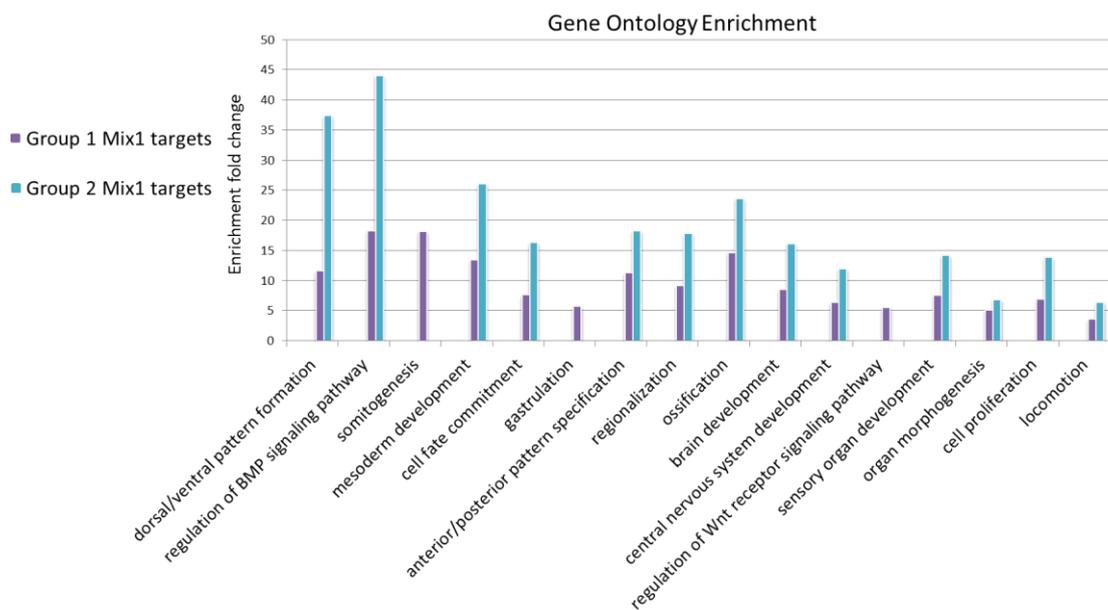


Figure 6.14. Annotated GO terms for biological processes enriched in the Group 1 and Group 2 Mix1 targets. Enriched GO terms in Group 1 Mix1 targets (purple) which includes validated Mix1 targets with consistent differential expression in all three MO time-series and in *mix1*-expressing animal caps. Enriched terms in Group 2 Mix1 targets (blue) which includes strong candidate Mix1 targets with consistent differential expression in all three MO time-series. Fold changes represent the observed/expected count for each term.

6.3 Discussion

6.3.1 Genes are validated through perturbation in various experimental conditions

My experimental design of combining data from four separate datasets facilitates robust identification of Mix1 targets. I sequenced three different MO time-series as a strategy to control for off-target effects and then sequenced *mix1*-expressing animal caps as an additional control to validate the strong candidate Mix1 targets in Group 1. Nine genes were identified as activatory targets in the Group 2 Mix1 targets. These are validated Mix1 targets since their perturbed expression has been induced by both loss-of-function and gain-of-function of Mix1. Many of these validated targets are novel Mix1 targets.

In support of the validation of these Mix1 targets, I found that a significantly greater number of genes are differentially expressed with consistent differential regulation under different experimental conditions than would be expected by chance. The prevalence of transcription factors within the validated Mix1 targets, all of which are activatory targets, indicates that Mix1 functions to positively regulate early developmental gene regulatory networks.

The time-series approach facilitates the detection of small changes in expression by using sequential time points to calculate the error across the data points, and then to call differential expression taking this into account. This means the approach is much more powerful than sequencing a single time-point, as small divergences in gene expression might be attributed to random variation. Furthermore, as genes have different optimal time-points, a single time point approach would not detect the maximum differential expression of many genes.

6.3.2 Limitations of using sequencing data from mix1-expressing animal caps to identify Mix1 targets in this analysis

There are 24 genes from the Group 1 candidate Mix1 targets which are differentially expressed in the three MO knockdown time-series but not in *mix1*-expressing animal caps. There are several reasons which may explain why these genes are not differentially expressed in animal caps:

1. Mix1 repressive targets would not be down-regulated in animal caps unless already expressed there. There are two repressive targets in Group 1.
2. No *X. laevis* orthologue could be found for two of the Group 1 genes, so they could not be validated this way.
3. The animal cap does not recapitulate the conditions of the embryo, so chromatin may not be accessible at gene regulatory regions, and co-factors involved in gene regulation may not be present in animal cap cells. Targets which are cooperatively activated by Mix1 and other transcription factors therefore may not be identified using this approach.

For these reasons, despite not being confirmed by the animal cap overexpression experiment, any of these 24 candidate Mix1 targets may be real targets of Mix1.

Furthermore, in the comparison between time-series and animal cap data, numerous genes are differentially expressed in animal caps but are not differentially expressed in any of the MO knockdown time-series:

1. These could reflect random variation in animal cap gene expression.
2. These may be targets that are robustly regulated by a gene regulatory network that can compensate for the depletion of Mix1 in knockdown embryos.
3. These may be genes that are not endogenous Mix1 targets, but are regulated by Mix1 when it is ectopically expressed in animal caps, potentially due to the different availability of co-factors.

6.3.3 Combining data from three conditions controls for MO off-target effects

In each of the three MO time-series, the majority of differentially expressed genes do not overlap with targets identified in the other MO time-series. These are weak candidate targets and it is possible that differential expression of some of these genes is caused by off-target effects, due to the unintended blocking of gene expression in a MO-specific manner. Another possibility is that clutch variation exists, resulting in different genes being differentially expressed in separate experiments. This could be caused by varying quantities of maternal proteins and transcripts which affects the downstream changes in gene expression. Ideally, at least two replicate time-series from different clutches of embryos should have been collected for each MO in order to control for this variation. The lack of biological repeats in this work may contribute to the lack of overlap between the different conditions. This means that the MOs may not each induce numerous off-target effects, and the problem may be caused by the experimental design.

A technical factor which contributes to the numbers of non-overlapping targets found in the three MO time-series is the different lengths of the time-series. The different lengths meant that a different detection threshold was used for the MO1 time-series because the time-series was collected after *mix1* was activated (see Materials and Methods for detail). The different analysis meant more targets could be detected, which could be real targets or off-target effects. Furthermore, the longer time-series for MO2 means that more late-diverging genes were detected than in the other two time-series. This means that, for technical reasons, the three time-series are not directly comparable and therefore I cannot determine which differentially expressed genes are off-target effects in each time-series.

Recent work in zebrafish comparing MO knockdown phenotypes to mutant phenotypes generated using ZFNs, TALENs and CRISPR has found that approximately 80% of MO phenotypes were not recapitulated in mutant embryos (Kok et al. 2015). This study suggests that genome targeting knockout approaches are more reliable for phenotypic analysis. This

study emphasises the need to exercise caution when using MOs, and I do so by generating three independent MO knockdown time-series, and examining only genes which are perturbed in all three MO knockdown conditions to control for off-target effects. Also by sequencing *mix1*-expressing animal caps, Mix1 targets are validated, which adds further robustness to my approach.

However, it has been recognised that compensatory pathways can be activated in mutant lines which are not activated in MO knockdowns. MO knockdown of the endothelial extracellular matrix protein *Egfl7* caused severe vascular defects in morphants, whereas mutants generated using TALENs displayed no defects apart from brain haemorrhage in 5% of cases. These differences were attributed to the up-regulation of the related *emilin* genes which were able to rescue the morphant phenotype (Rossi et al. 2015). Knockout lines may have epigenetic alterations which are heritable and alter gene expression to obscure the effects of gene deletion. This indicates that MO studies may be more reliable than has been suggested in recent knockout studies, since compensatory pathways are not activated in morphants.

In zebrafish embryos, the apoptotic pathway can be activated by MOs, prompting the use of p53 knockout zebrafish lines in MO studies. Surprisingly, apoptotic genes were found to induce neurogenic gene expression in zebrafish hindbrain boundaries (Gerety and Wilkinson 2011). This indicates that apoptotic genes have the potential to activate different regulatory pathways in MO knockdowns. In both Mix1 knockdown time-series, I found *tp53*, *mdm2* and *cdkn1a* (*p21*) were not differentially expressed. *Bcl3* was up-regulated in the splice blocking MO knockdowns and just slightly up-regulated in the translation blocking MO knockdowns (although not classified as differentially expressed). *Bcl3* is an apoptotic gene, so could be the cause of off-target effects in any of the knockdown time-series, which might account for the genes differentially expressed in just one MO time-series (Brocke-Heidrich et al. 2006).

In light of these numerous possible causes of off-target effects, and significant overlap between experimental conditions, it is possible that the three MOs target Mix1, but also have individual off-target effects. My experimental design addresses this problem by selecting only genes that are differentially expressed in all three time-series to control for off-target effects to ensure robust identification of Mix1 targets. However, this may eliminate some real Mix1 targets that do not appear in one time-series for various reasons (see below).

6.3.4 Targets not differentially expressed in all three MO time-series

The experiments carried out in this thesis may give variable results due to several possible factors. Here I explore reasons that might explain why targets might not be identified as differentially expressed in all three time-series.

1. On a technical basis, the differing lengths of time-series mean analysis thresholds and detection of targets is different in each time-series experiment.
2. In the case of the two translation blocking MO time-series, Mixer targets will be present which are not found in the splice blocking MO time-series.
3. False negative and false positive results can be produced as artefacts of the experimental method. Each MO injected may have differing efficacy to regulate Mix1 depletion, and the effect of Mix1 depletion on gene expression may be non-linear with respect to the Mix1 concentration where specific thresholds of transcription factor binding produce a transcriptional output.
4. Off-target effects caused by specific MOs blocking expression of unintended genes can generate false negative or false positive results, and off-target genes can have secondary effects which can lead to further false negative or positive results.
5. Variation in gene expression between clutches of embryos could mean some genes are regulated differently and pathways may be more or less susceptible to Mix1 depletion.

6.3.5 Mix1 targets not found in this work

Previous studies have showed that the mesoderm-expressed genes *brachyury*, *fgf4* and *fgf8* were repressed in *mix1*-overexpressing gastrula stage *X. laevis* embryos (Colas et al. 2008) and *a2m* (*endodermin*) and *hba1* (α -*globin*) were marginally induced in *mix1*-expressing animal caps (Henry and Melton 1998). *Hba1* is activated much later than could be detected by the three MO time-series, and *a2m* is activated close to the end of the time-series (Owens 2015), so differential expression of these genes is difficult to determine. *Fgf4* is expressed at a very low level in wild type embryos (Owens 2015), so may have fallen below the RNA-seq detection threshold in either time-series.

The MO time-series data does not support the finding that *fgf8* and *brachyury* are repressive Mix1 targets. *Fgf8* is not differentially expressed in any of the MO time-series or in *mix1*-expressing animal caps and *brachyury* was differentially expressed in the translation blocking

MO2 time-series where it was up-regulated, but was not differentially expressed in the other two MO time-series or in the *mix1*-expressing animal caps.

The previous MO study in *X. laevis* measured the expression of Mix1 target genes after 3 hours incubation of stage 10 embryos, meaning they were likely stage 11 when the targets *brachyury*, *fgf4* and *fgf8* were identified (Colas et al. 2008). The longer nature of the MO2 time-series seems to explain why *brachyury* was detected here, but not in the other MO time-series. It is likely that the time-series are not sufficiently long enough to detect all Mix1 targets. Interestingly, *brachyury* has been found to be up-regulated in response to Mixer knockdown (Kofron, Wylie, and Heasman 2004), so the differential expression seen in the MO2 time-series may be an effect of Mixer depletion.

Brachyury is thought to be indirectly repressed by Mix1 which fits with the idea that Mix1 represses mesoderm through its activatory function (Lemaire et al. 1998; Latinkic and Smith 1999; Fletcher, Baker, and Harland 2006). Two suggested mechanisms of the indirect repression of *brachyury* by Mix1 are (i) through repression of *fgf8* which is an activator of *brachyury* (Fletcher, Baker, and Harland 2006) or (ii) through activation of *gsc*, a repressor of *brachyury* (Latinkic and Smith 1999). I find support for neither mechanism, as *brachyury* and *fgf8* are not differentially expressed in the Mix1 splice blocking MO knockdown time-series. It is possible that Gsc activates *brachyury*, but since *gsc* was only partially repressed in the splice MO time-series the levels of Gsc might be too high to allow the de-repression of *brachyury* expression. Alternatively, the absence of *brachyury* and *fgf8* in the Mix1 repressive target set could potentially arise from gene interaction differences between *X. laevis* and *X. tropicalis* (Colas et al. 2008).

6.3.6 Gene ontology analysis of Mix1 targets

6.3.6.1 Analysis of Mix1 targets reveals a likely function for Mix1 in neural development

GO analysis of the Mix1 candidate and validated targets revealed enrichment for functional annotations in neural development, the terms *brain development* and *central nervous system development* were enriched in Group 1 and Group 2 Mix1 targets. Previous studies have identified functional roles for the nine validated Mix1 targets in neural development. Here I outline some of the functions of the validated Mix1 targets in regulation of neural development.

In *X. laevis*, *cer1* is expressed in the anterior endomesoderm at gastrula and neurula stages (Bouwmeester et al. 1996). Cer1 is a secreted antagonist of BMP, Wnt and Nodals and is

necessary for induction of neuroectoderm (Silva et al. 2003). *Irx1* is expressed in the presumptive midbrain in the *X. laevis* embryo and regulates the midbrain-hindbrain boundary (Glavic, Gomez-Skarmeta, and Mayor 2001). *Irx1* represses transcription of *BMP* and is necessary for neural development (Gomez-Skarmeta, de La Calle-Mustienes, and Modolell 2001). The related gene *irx3* is expressed in the presumptive midbrain, hindbrain and spinal cord in zebrafish (Tan, Korzh, and Gong 1999). *Irx3* regulates the patterning of the antero-posterior neural axis downstream of Wnt signals in *X. laevis* (Janssens et al. 2010). *Arx* is expressed in the diencephalon and telencephalon in *X. laevis* and is necessary for normal forebrain development (Seufert, Prescott, and El-Hodiri 2005). Interestingly, *Arx* antagonises *irx1* and *irx3* in the regionalisation of neural plate. *Arx* represses the expression of *irx* genes in the anterior forebrain, and *Irx* factors repress posterior expression of *arx* (Rodriguez-Seguel, Alarcon, and Gomez-Skarmeta 2009). In *X. laevis*, *foxc1* is expressed in the marginal zone at the gastrula stage and in the lateral neural plate at the neurula stage and is required for normal neural tube development (Koster, Dillinger, and Knochel 1998; Cha et al. 2007). In *X. laevis*, *phox2b* is expressed in the hindbrain and neural tube as well as the ventral heart field (Talikka et al. 2004). *Phox2b* regulates differentiation of neuronal progenitor cells in chick (Dubreuil et al. 2000) and regulates differentiation of motor neurons in mouse hindbrain (Pattyn et al. 2000). In *X. laevis* *msx2* is expressed in the epidermal-neural border where neural crest is induced. Both *msx2* and *msx1* are essential for neural crest induction in *X. laevis* (Khadka, Luo, and Sargent 2006). *Admp* antagonises head formation which contrasts with the pro-anterior function of other Mix1 targets (Dosch and Niehrs 2000). In *X. laevis* *Admp* is expressed in the Spemann organiser and later in the prechordal plate (Moos, Wang, and Krinks 1995). The gap junction protein *Gjb2* is not well characterised for neural functions, but mutation in *gjb2* has been linked to neural deafness indicating a likely role in sensory organ development (Lee, Derosa, and White 2009).

Two studies have revealed a potential neural-inductive function of Mix1. In support of a pro-neural role for Mix1, neural expression of *pax2* and *pax8* was down-regulated in Mix1 and Mix1/Mix2 depleted embryos at the neurula and tailbud stages respectively (Colas et al. 2008). I did not detect differential expression of *pax2* and *pax8* because their late expression was not within the time-series. The expression of these two genes may be regulated by some of the validated Mix1 targets outlined above.

The zebrafish gene *mixl1* is an orthologue of *Xenopus mix1* (Ensembl 2015). *Mix1* knockout disrupted forebrain development and reduced the expression of the neural plate marker *otx2*

and the dorso-anterior gene *gsc* (Trinh, Meyer, and Stainier 2003). *Mix1* is not expressed in the forebrain, but could regulate forebrain development indirectly through *otx2* which has an overlapping expression domain with *mix1* in the mesoderm, but is also expressed in the ectoderm. This is in accordance with the splice blocking MO time-series where *otx2* and *gsc* are down-regulated, and the animal cap data where they are up-regulated. *Gsc* and *otx2* are both up-regulated in both translation blocking MO time-series, although only *gsc* in the translation blocking MO2 time-series diverges sufficiently to be considered differentially expressed. This suggests that Mixer depletion may be responsible for this up-regulation (see below for Mixer discussion).

Given that most of the validated Mix1 targets are expressed in neural tissues and have established roles in regulating neural development it seems likely that, through regulation of these targets, Mix1 promotes neural development. *Admp* is an exception, as this is an anti-neural BMP signalling molecule, but is important in regulating the BMP gradient which fine tunes dorsal and neural patterning (Reversade and De Robertis 2005).

6.3.6.2 Mix1 as a regulator of dorso-anterior development

The GO terms *dorsal/ventral pattern formation* and *anterior/posterior pattern specification* are enriched in the Group 1 and Group 2 Mix1 targets. Many of the Group 2 Mix1 targets are expressed in the Spemann organiser at the gastrula stage; *admp* (Reversade and De Robertis 2005), *gjb2* (Bowes et al. 2010), *irx1* (Rodriguez-Seguel, Alarcon, and Gomez-Skarmeta 2009), *irx3* (Rodriguez-Seguel, Alarcon, and Gomez-Skarmeta 2009) and *cer1* (Wills et al. 2008). The Spemann organiser has the potential to induce the dorsal axis, which forms anterior structures (Cooke 1972), so the expression of the validated Mix1 targets in this region supports the idea that these genes contribute to the establishment of the dorsal and anterior poles..

Support for Mix1 as a regulator of anterior development comes from Mix1 loss-of-function studies in which anterior structures were disrupted in *X. laevis* embryos. Injection of mRNA encoding dominant negative Mix1 or antisense *mix1* caused embryos to develop dorso-anterior defects, ranging from reduced anterior structures to complete loss of head development (Latinkic and Smith 1999; Lemaire et al. 1998). The two known direct Mix1 targets, *gsc* and *cer1* (Sudou et al. 2012), are expressed in the dorsal organizer region (De Roberts et al. 1992; Bouwmeester et al. 1996) and they both have the ability to induce a secondary dorsal axis with ectopic head structures when expressed in the ventral

blastomeres (Cho et al. 1991; Bouwmeester et al. 1996). *Cer1* is a BMP, Wnt and nodal antagonist, and blocking these two signals is sufficient for head induction in ventral marginal zone explants (Piccolo et al. 1999). These experiments identify a strong dorso-anterior inducing potential for *Gsc* and *Cer1*. *Cer1* is in the validated Mix1 target set, but *gsc* is only identified as a Mix1 activatory target in the splice blocking MO time-series and animal cap data. As discussed earlier, Mixer may affect the expression of *gsc*, and antagonise the effect of Mix1 depletion in the two translation blocking MO time-series.

In further support for the role of Mix1 in dorso-anterior development, Mix1 can form heterodimers with *Gsc* (Wilson et al. 1993) and *Sia* (Mead et al. 1996). Ventral injection of *gsc* or *sia* mRNA was sufficient to induce a secondary dorsal axis including axial and head structures (Cho et al. 1991; Lemaire, Garrett, and Gurdon 1995).

6.3.6.3 Mix1 as a regulator of endoderm and mesoderm development

The GO term *mesoderm development* is enriched in the Group 1 and Group 2 Mix1 targets. As most Mix1 targets are activatory, this suggests a pro-mesodermal function for Mix1. Published studies are somewhat conflicted over the role of Mix1 in mesoderm development. Studies have shown that Mix1 can repress mesoderm marker genes in whole embryo overexpression experiments (Lemaire et al. 1998; Colas et al. 2008). Some studies propose a role for Mix1 in promoting mesoderm development, as Mix1 overexpression can increase blood formation, and loss-of-function can block heart development (Mead et al. 1996; Latinkic and Smith 1999). It is possible that Mix1 has a dose-dependent effect on mesoderm formation, exerting a repressive or activatory effect on genes which specify mesoderm, depending on expression level in particular regions.

6.3.7 The translation blocking MOs target Mix1 and Mixer

The two translation blocking MOs target Mix1 and Mixer, these two transcription factors may have some common and different targets. The homeodomains in Mix1 and Mixer share 84% identity in the protein sequences in *X. laevis* and 77% in *X. tropicalis*. Mix1 alone does not have endoderm inducing potential whereas Mixer does (Doherty et al. 2006). The different amino acids in the Mix1 and Mixer homeodomain do not affect endoderm induction, rather the conserved c-terminal acidic domain found in Mixer was found to be necessary for endoderm induction, but the Mix1 acidic domain with the Mixer homeodomain was insufficient for endoderm induction (Doherty et al. 2006). The similarity of the homeodomain within the Mix1 and Mixer proteins may account for their similar roles in mesoderm

repression (Lemaire et al. 1998; Kofron, Wylie, and Heasman 2004). Given the differing endoderm-inducing potential of these two proteins, I would expect them to have some different targets.

The two translation blocking MOs target Mixer and Mix1, so more targets would be expected to be shared between the two translation blocking MO time-series than either with the Mix1 splice blocking MO time-series. I found a slightly larger enrichment in the number of genes overlapping between the two translation blocking MOs compared to genes overlapping between the splice blocking MO and either one of the translation blocking MOs (5.4x enrichment in splice-MO1 overlap, 5.7x enrichment in splice-MO2 overlap, 6.5x enrichment in MO1-MO2 overlap). This fits with the idea that the translation blocking MOs would share more targets with each other than with the splice blocking MO, because both Mixer and Mix1 targets are shared.

The 7.6x enrichment found in the overlap of targets with conflicting directionality of differential expression between the three MO time-series (section 6.2.6) may be caused by interference by Mixer depletion in the two translation blocking MO time-series. Importantly, this enrichment is much lower than is found for the consistently differentially regulated targets which supports the idea that these genes are genuinely differentially expressed in response to Mix1 depletion.

Gsc may be an example of a gene that is a target of both Mix1 and Mixer; it is identified as a candidate activatory target in the splice blocking MO time-series data and is up-regulated in *mix1*-expressing animal caps. In contrast, *gsc* is identified as a repressive target in the translation blocking MO2 time-series and begins to be up-regulated at the end of the MO1 time-series. The up-regulation in the translation blocking MO time-series suggests that *gsc* may be a repressive target of Mixer and an activatory target of Mix1. *Gsc* is an established activatory target of Mix1, demonstrated by its expression in *mix1*-expressing animal caps and through cooperative binding of Mix1 to the *gsc* promoter region (Latinkic and Smith 1999; Sudou et al. 2012). It is known that *gsc* is not an activatory target of Mixer: in an animal cap assay in *X. laevis*, *gsc* was induced in *sox17b*-expressing animal caps, but not in *mixer*-expressing animal caps (Dickinson, Leonard, and Baker 2006). It is possible that *gsc* is a repressive target of Mixer and that Mixer has a stronger effect than Mix1 on *gsc* expression, which would explain the different regulation between time-series.

6.3.8 Some known Mixer targets were consistently differentially expressed in the two translation blocking MO time-series but not in the Mix1 splice blocking MO time-series

To test for Mixer depletion in the two translation blocking MO time-series, I examined differential expression of known Mixer targets within the two translation blocking MO time-series. Using a set of Mixer targets identified previously in a MO knockdown study in *X. laevis*, I searched for differential expression of the activatory targets *gata5*, *cer1*, *a2m*, *sox17*, *brachyury*, *vegt* and *fgf4* and the repressive targets *eomes*, *bix1*, *not*, *gata2*, *fgf3*, *fgf8*, *nodal1* and *nodal5* in the two translation blocking MO time-series (Kofron, Wylie, and Heasman 2004).

The activatory targets *gata5*, *sox17a* and *cer1* were down-regulated in both of the translation blocking MO time-series. *Cer1* is an established Mix1 target both in previous work and is differentially expressed in the splice blocking MO time-series so is most likely a target of both Mix1 and Mixer (Sudou et al. 2012). The differential expression of the known Mixer activatory targets *gata5* and *sox17a* in the two translation blocking MO time-series but not in the Mix1 splice blocking MO time-series suggests that these genes are differentially expressed in response to Mixer depletion in MO1 and MO2-injected embryos.

The activatory targets *brachyury* and *vegt* were up-regulated only in the MO2 time-series and the repressive target *eomes* was up-regulated only in the MO1 time-series. The activatory target *sox17b* and the repressive targets *fgf8*, *not*, *gata2*, *nodal1*, *nodal5* were not differentially expressed in either time-series. The activatory targets *a2m* and *fgf4* and the repressive targets *bix1.1*, *bix1.2* and *fgf3* expression were not detected in either time-series. The reason for the absence of the *bix* genes is unclear, but the late activation of *fgf3* and *a2m*, and the low expression level of *fgf4* (Owens 2015), may explain why they were not detected in either time-series.

It is likely that timing is important in identifying transcription factor targets, and the regulation of other known targets by Mixer may occur later than the time points covered in the two time-series. The depletion of Mixer appears to have a more marked effect on target gene expression in the previous MO study at stage 11 than at stage 10 (Kofron, Wylie, and Heasman 2004). The 10 hpf time-point roughly coincides with stage 11 in the time-series used here, due to slow developing clutches of embryos, so it is likely that more Mixer targets would be detected in a longer time-series.

6.3.9 The Mix1 translation blocking MO phenotypes

The two translation blocking MOs targeting Mix1 and Mixer both generated similar phenotypic defects when injected. The phenotype is different to the ventralised phenotype found in embryos injected with the Mix1 splice blocking MO. There were some similarities, as all three MOs generate a truncated phenotype and head and tail defects. The difference in phenotype is therefore likely to be caused by the additional Mixer depletion in the two translation blocking MO time-series. As mentioned in Chapter 5, the ventralised phenotype induced by the splice blocking MO was not found in *X. laevis* when a Mix1 MO was injected (Colas et al. 2008). The ventralised phenotype may then be caused by off-target effects of the Mix1 splice blocking MO.

When comparing the phenotype of the Mix1/Mixer translation blocking MOs to the individual Mixer splice blocking MO-induced phenotype (see Chapter 4), there are clear similarities. All embryos are truncated with defective head and tail development. Interestingly, the loss of head development seems to be more severe in the individual splice blocking MOs, whereas one might expect the double knockdown phenotype to be more severe. Head development was affected in the three MO knockdowns, which may be caused by some loss of neural induction in Mix1 depleted embryos.

6.3.10 Conclusions

My experimental design uses multiple knockdown time-series and animal cap expression analysis to robustly identify targets of Mix1, an early-activated transcription factor. I identified and validated nine Mix1 targets and analysed their known functions in order to better understand the role of Mix1 in early development. I found that many Mix1 of these promote neural development, suggesting a novel function for Mix1. I identified many weaker candidate Mix1 targets, which have similar functional annotation to the validated targets. These candidate targets must be validated e.g. using CHIP-seq or mutant embryos to demonstrate the transcription factor-target interaction without using MOs. Time-series analysis is particularly useful for observing precise timings of gene interactions within a short window of development in great detail. This approach could be applied to other early transcription factors and used to delineate the GRNs which regulate embryonic development.

Chapter 7: Final summary and discussion

7.1 Chapter 3: mechanisms of gene activation

In the first results chapter I differentiated between the two mechanisms of gene activation in the early embryo; polyadenylation and transcription. I used non-polyA selective gene expression analysis to characterise the switch from early polyadenylation of maternally deposited transcripts, to *de novo* transcription shortly before MBT. This builds on previous work which suggests there are two early, distinct waves of gene activation and reveals that polyadenylation is the predominant mechanism of activation in the first wave and zygotic transcription is the predominant mechanism of activation in the second wave.

7.2 Chapter 4: Morpholino screen of early-activated genes

In this chapter I set out to select early activated transcription factors for later knockdown sequencing analysis to determine transcription factor targets. I screened a selection of maternal polyadenylated and zygotically transcribed transcription factors. From this screen I selected the zygotically transcribed transcription factor Mix1 due to its transient, early activation and the early, and highly penetrant, phenotype produced in response to morpholino knockdown.

7.3 Chapter 5: Investigating downstream targets of Mix1

The aim of this chapter was to investigate the targets of Mix1 using morpholino knockdown and time-series RNA-seq. By comparing divergent expression between the control MO and Mix1 splice blocking MO conditions I was able to identify almost 500 candidate Mix1 targets. I demonstrated the mis-expression of some of these candidate targets within the *mix1* expression domain. The Mix1 knockdown phenotype was not rescued and without this validation, I decided that additional experimental conditions must be tested to control for off-target effects of the MO, and to validate the candidate targets using an alternative approach.

I demonstrated the effectiveness of using morpholino knockdown time-series RNA-seq to identify transcription factor targets. By sequencing embryos collected at regular time-points gene expression changes can be identified and the timing at which regulation events occur can be understood. Gaussian process models are used to evaluate and handle experimental noise, and call differential expression based on the confidence intervals calculated from the

noise. This enables the detection of small changes in gene expression in a way that is not reliant on arbitrary thresholds. This approach is more powerful than sequencing a single time-point as it is possible to ensure that differential expression is caused by a divergence in gene expression over time. Different targets would have different optimal time-points for measuring differential expression, therefore using a time-series captures a larger range of differential expression than a single time-point could.

7.4 Chapter 6 – Controlling for MO off-target effects to determine Mix1 targets

In the final results chapter, I introduced two additional morpholinos that target Mix1 and Mixer, and also sequence *mix1*-expressing animal caps. I used the two morpholinos to generate two further knockdown time-series, which were sequenced to identify additional sets of candidate Mix1 targets: 731 and 558 candidate targets were identified in these two experiments. I then compared the candidate targets from the three MO time-series and found significant numbers of genes consistently differentially expressed in all three conditions; 32 activatory targets and 3 repressive targets. I then tested the effect of Mix1 gain-of-function by sequencing *mix1*-expressing animal caps. From this data I identified an additional set of 354 candidate Mix1 targets, which I used to validate the targets identified in the three MO time-series. Targets were designated as validated if consistently differentially regulated in the four experimental conditions which included gain-of-function and loss-of-function approaches. Nine genes had consistent differential regulation in all four experimental conditions, and are therefore considered *validated* Mix1 targets. However, to truly validate these targets, the genomic interaction of Mix1 with the regulatory regions of these genes should be demonstrated using ChIP analysis. The validated Mix1 targets represent a statistically significant overlap between experimental conditions, and are enriched for transcription factors indicating that Mix1 has a role in establishing early gene regulatory networks. The strong candidate and validated Mix1 targets were enriched for gene ontology terms linked to neural development, indicating a novel function for Mix1. Gene ontology terms also support the findings from previous studies that suggest Mix1 regulates dorsal-ventral and antero-posterior patterning.

Overall this work contributes to the understanding of Mix1 function and for the first time, has produced a list of validated Mix1 targets. I detected a large number of differentially expressed genes which were unique to each morpholino tested, which suggests that off-target effects are prevalent in each knockdown. These off-target effects are most likely sequence specific for each morpholino, through binding to off-target mRNAs and

unintentionally blocking gene expression. I propose that multiple morpholino time-series should be used to control for off-target effects.

7.5 Limitations of the approaches used in this thesis

Although the approaches outlined in this thesis have been used to identify targets of Mix1, and to suggest likely roles for this transcription factor, some approaches could have been carried out more effectively. These are described in the following sections.

7.5.1 The morpholino screen

Firstly, in the early morpholino screen to select for phenotypes, screening for phenotypes using just one MO allows morpholinos with off target effects to be selected. It was not possible to validate all MOs in the screen because of the time-consuming nature of this task. Similarly, it was not possible to screen each knockdown for gene expression changes by *in situ* hybridisation. It may have been more practical to rely on published studies to indicate which transcription factors might have a significant developmental role accompanied by validations such as rescue. This may have introduced a selection bias, however, for the purposes of testing the knockdown time-series RNA-seq approach for the ability to identify transcription factor targets, this is not necessarily important. It is likely that interesting data could have been generated for any early transcription factor tested, as long as an early knockdown phenotype was observed.

7.5.2 Additional experiments to complement the Mix1 splice blocking morpholino time-series

Given the scale of genes with inconsistent differential expression in the three MO time-series outlined in this thesis, it is likely that each MO has a large repertoire of off-target effects. Based on the small sample of three MOs, I find similar numbers of possible off-target effects for each. This supports the idea that these effects are quite general and therefore a good strategy for knockdown studies would be to use more than one MO. Here I suggest a series of experiments which could be used to further understand the scale of off-target effects, and strategies for controlling for these effects.

Following the Mix1 splice blocking morpholino time-series, in order to fully understand the implications of using a morpholino for such a study, a biological replicate time-series using the same Mix1 splice blocking MO would have been a good approach for testing the variability of gene expression within the system. This biological replicate would allow me to

distinguish between clutch-specific effects and MO-specific off-target effects. Then to control for MO-specific off-target effects, instead of using two alternative MOs which block translation of Mix1 and Mixer, alternative non-overlapping Mix1 splice blocking MO could be designed. A morpholino targeting the intron 1-exon 2 boundary would cause the deletion of exon 2 which contains a large portion of the homeodomain sequence. This second splice blocking MO would be equivalent to the original splice blocking MO so would be a good control for off-target effects. As the two translation blocking MOs used in this work target both Mix1 and Mixer, due to an identical region in the 5'UTR of both mRNAs, the splice MO is preferred as it is specific to Mix1.

Furthermore, in order to differentiate between off-target effects and technical differences between time-series, the collection of the time-series should be as similar as possible. Ideally, the two Mix1 splice blocking MOs and control MOs would be injected into embryos from the same clutch to control for gene expression variation between clutches and the differences in developmental rates. At the very least, clutches of embryos should be collected at the same time-points to allow equivalent analysis to allow comparison of targets from different time-series. Biological replicates should be generated for each time-series. Any genes differentially expressed in only one of the Mix1 splice blocking MO time-series would likely be off-target effects.

A ChIP-seq experiment would complement the findings of the Mix1 time-series data. A Mix1 antibody must first be obtained, and then ChIP-seq could be used to identify genomic binding locations of Mix1. An antibody was generated for Mix1; however, it was not immunoreactive against Mix1 protein (data not shown). The genes found closest to these binding regions would make up a list of candidate Mix1 targets. Comparison to the Mix1 candidate targets identified in the knockdown time-series could validate the direct targets of Mix1.

Mix1 mutant embryos could be generated using the CRISPR-Cas9 technology to induce targeted genomic alterations. If the homozygous mutant embryos exhibit the same early lethal phenotype as the MO-injected embryos this would indicate that the Mix1 MO is responsible for the depletion of Mix1 and that the phenotype is an effect of this. If the homozygous mutant embryos survived through part of development they could be sequenced to identify Mix1 targets using the time-series approach. It would be particularly interesting to examine whether the targets identified using Mix1 knockdown were the same. If compensatory networks were activated in the knockouts, then this would not be useful for time-series analysis to identify Mix1 targets (Rossi et al. 2015).

If the homozygous mutant embryos did not exhibit any phenotype, this would either indicate that the deletion of Mix1 has no major effect on development or that compensatory mechanisms are present in the knockout embryos which replace the function of Mix1 allowing normal development (Rossi et al. 2015).

7.5.3 Additional experiments to investigate Mix1 function

The biggest question specific to Mix1 which arises from this work is whether it is essential for neural development, or whether it simply contributes to the gene regulatory networks which regulate neural development. In order to understand the physiological effects of Mix1 depletion on neural development, *in situ* hybridisation for neural markers e.g. the pan-neural marker N-CAM, the hindbrain marker Krox-20 and the anterior neural markers En-2, Otx2 and Foxg1 (Bolce et al. 1992) should be performed. This would reveal whether Mix1 is necessary for formation of specific neural tissues.

The work presented here does not measure combinatorial effects with other transcription factors. Mix1 has been shown to bind cooperatively to genomic regulatory regions with other transcription factors (Sudou et al. 2012), so future ChIP-seq studies examining the cooperative binding of multiple transcription factors will contribute to a richer understanding of Mix1 function.

Mix1 has been shown to dimerise with Gsc (Wilson et al. 1993) and Sia1 (Mead et al. 1996), both of which have the ability to induce a secondary dorsal axis with head structures (Cho et al. 1991; Lemaire, Garrett, and Gurdon 1995). Furthermore, these three transcription factors have been shown to cooperatively regulate the migration of the head mesoderm (Luu et al. 2008). For these reasons, a ChIP-seq experiment for these three transcription factors could be used to determine shared genomic binding regions which might be used to identify targets involved in the regulation of dorsal and anterior development.

Double morpholino knockdown studies of Mix1 with Gsc or Sia1 would be interesting, particularly if this data could be compared to ChIP-seq data. Double knockdown time-series RNA-seq studies would most likely be problematic because of the potential for both MOs to induce off-target effects. Multiple MOs should be used to identify real and off-target effects caused by MOs targeting a single gene, then double knockdown time-series sequencing could be used, and the known off-target genes could be removed during analysis. The double knockdown data could then be compared to the single knockdown data to identify common targets which might be compensated for by the other gene in single knockdown experiments.

Finally, an interesting experiment would be to dissect Mix1 knockdown embryos and sequence different regions of the embryo, e.g. animal and vegetal halves, dorsal and ventral halves, and the marginal zone. This might eliminate the problem of small gene expression changes not being detectable when whole embryos are sequenced. Since many of the validated Mix1 targets are dorsally expressed, sequencing the dorsal half of the embryo might reveal more Mix1 targets that exhibit weaker differential expression in response to Mix1 depletion.

7.5.4 Limitations of control morpholinos

The standard control MO has limited practicality because the off-target effects of morpholinos seem to be sequence specific. This means that the same problem may apply to 5-base mismatch morpholinos. As these are closer in sequence to the morpholino than the standard control morpholino, 5 bases may be sufficiently similar to exert an effect on the gene of interest. I identified some splice blocking activity for my 5-base mismatch Mix1 morpholino. Evidence suggests that 4-base mismatch morpholinos have the ability to induce phenotypic defects in injected embryos (Wickstrom et al. 2004). Mismatching more bases may generate a morpholino so dissimilar to the original that it is not an effective control.

7.6 Conclusion

To conclude, I present a strategy for identifying transcription factor targets over a time-series. The use of multiple morpholinos controls for off-target effects, which may be prevalent in each morpholino knockdown. I used this approach to identify a list of candidate Mix1 targets, and validate nine of these targets using animal cap explants. Most of the validated Mix1 targets have established functions in neural development. This reveals likely novel role for Mix1 in the regulation of neural development. This approach could be applied to other early transcription factors in order to delineate the early gene regulatory networks which control embryonic development.

Bibliography

- Aanes, H., C. L. Winata, C. H. Lin, J. P. Chen, K. G. Srinivasan, S. G. Lee, A. Y. Lim, H. S. Hajan, P. Collas, G. Bourque, Z. Gong, V. Korzh, P. Alestrom, and S. Mathavan. 2011. 'Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition', *Genome Res*, 21: 1328-38.
- Abe, K., R. Yamamoto, V. Franke, M. Cao, Y. Suzuki, M. G. Suzuki, K. Vlahovicek, P. Svoboda, R. M. Schultz, and F. Aoki. 2015. 'The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3' processing', *EMBO J*, 34: 1523-37.
- Agius, E., M. Oelgeschlager, O. Wessely, C. Kemp, and E. M. De Robertis. 2000. 'Endodermal Nodal-related signals and mesoderm induction in *Xenopus*', *Development*, 127: 1173-83.
- Akkers, R. C., S. J. van Heeringen, U. G. Jacobi, E. M. Janssen-Megens, K. J. Francoijs, H. G. Stunnenberg, and G. J. Veenstra. 2009. 'A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in *Xenopus* embryos', *Dev Cell*, 17: 425-34.
- Alexander, J., and D. Y. Stainier. 1999. 'A molecular pathway leading to endoderm formation in zebrafish', *Curr Biol*, 9: 1147-57.
- Amaya, E., T. J. Musci, and M. W. Kirschner. 1991. 'Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in *Xenopus* embryos', *Cell*, 66: 257-70.
- Amaya, E., P. A. Stein, T. J. Musci, and M. W. Kirschner. 1993. 'FGF signalling in the early specification of mesoderm in *Xenopus*', *Development*, 118: 477-87.
- Amodeo, A. A., D. Jukam, A. F. Straight, and J. M. Skotheim. 2015. 'Histone titration against the genome sets the DNA-to-cytoplasm threshold for the *Xenopus* midblastula transition', *Proc Natl Acad Sci U S A*, 112: E1086-95.
- Ariizumi, T., S. Takahashi, T. C. Chan, Y. Ito, T. Michiue, and M. Asashima. 2009. 'Isolation and differentiation of *Xenopus* animal cap cells', *Curr Protoc Stem Cell Biol*, Chapter 1: Unit 1D 5.
- Audic, Y., F. Omilli, and H. B. Osborne. 1997. 'Postfertilization deadenylation of mRNAs in *Xenopus laevis* embryos is sufficient to cause their degradation at the blastula stage', *Mol Cell Biol*, 17: 209-18.
- Barkoff, A., S. Ballantyne, and M. Wickens. 1998. 'Meiotic maturation in *Xenopus* requires polyadenylation of multiple mRNAs', *EMBO J*, 17: 3168-75.
- Baur, F., K. Nau, D. Sadic, L. Allweiss, H. P. Elsasser, N. Gillemans, T. de Wit, I. Kruger, M. Vollmer, S. Philipsen, and G. Suske. 2010. 'Specificity protein 2 (Sp2) is essential for mouse development and autonomous proliferation of mouse embryonic fibroblasts', *PLoS One*, 5: e9587.
- Beelman, C. A., A. Stevens, G. Caponigro, T. E. LaGrandeur, L. Hatfield, D. M. Fortner, and R. Parker. 1996. 'An essential component of the decapping enzyme required for normal rates of mRNA turnover', *Nature*, 382: 642-6.
- Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander. 2006. 'A bivalent chromatin structure marks key developmental genes in embryonic stem cells', *Cell*, 125: 315-26.
- Blythe, S. A., S. W. Cha, E. Tadjuidje, J. Heasman, and P. S. Klein. 2010. 'beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2', *Dev Cell*, 19: 220-31.

- Bogdanovic, O., S. W. Long, S. J. van Heeringen, A. B. Brinkman, J. L. Gomez-Skarmeta, H. G. Stunnenberg, P. L. Jones, and G. J. Veenstra. 2011. 'Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis', *Genome Res*, 21: 1313-27.
- Bogdanovic, O., S. J. van Heeringen, and G. J. Veenstra. 2012. 'The epigenome in early vertebrate development', *Genesis*, 50: 192-206.
- Bolce, M. E., A. Hemmati-Brivanlou, P. D. Kushner, and R. M. Harland. 1992. 'Ventral ectoderm of *Xenopus* forms neural tissue, including hindbrain, in response to activin', *Development*, 115: 681-8.
- Bouvet, P., J. Paris, M. Phillippe, and H. B. Osborne. 1991. 'Degradation of a developmentally regulated mRNA in *Xenopus* embryos is controlled by the 3' region and requires the translation of another maternal mRNA', *Mol Cell Biol*, 11: 3115-24.
- Bouwmeester, T., S. Kim, Y. Sasai, B. Lu, and E. M. De Robertis. 1996. 'Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer', *Nature*, 382: 595-601.
- Bowes, J. B., K. A. Snyder, E. Segerdell, C. J. Jarabek, K. Azam, A. M. Zorn, and P. D. Vize. 2010. 'Xenbase: gene expression and improved integration', *Nucleic Acids Res*, 38: D607-12.
- Brocke-Heidrich, K., B. Ge, H. Cvijic, G. Pfeifer, D. Loffler, C. Henze, T. W. McKeithan, and F. Horn. 2006. 'BCL3 is induced by IL-6 via Stat3 binding to intronic enhancer HS4 and represses its own transcription', *Oncogene*, 25: 7297-304.
- Bruce, A. E., C. Howley, Y. Zhou, S. L. Vickers, L. M. Silver, M. L. King, and R. K. Ho. 2003. 'The maternally expressed zebrafish T-box gene *eomesoderm* regulates organizer formation', *Development*, 130: 5503-17.
- Cao, Q., and J. D. Richter. 2002. 'Dissolution of the maskin-eIF4E complex by cytoplasmic polyadenylation and poly(A)-binding protein controls cyclin B1 mRNA translation and oocyte maturation', *EMBO J*, 21: 3852-62.
- Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang. 2002. 'Role of histone H3 lysine 27 methylation in Polycomb-group silencing', *Science*, 298: 1039-43.
- Castagnetti, S., and A. Ephrussi. 2003. 'Orb and a long poly(A) tail are required for efficient oskar translation at the posterior pole of the *Drosophila* oocyte', *Development*, 130: 835-43.
- Cha, J. Y., B. Birsoy, M. Kofron, E. Mahoney, S. Lang, C. Wylie, and J. Heasman. 2007. 'The role of FoxC1 in early *Xenopus* development', *Dev Dyn*, 236: 2731-41.
- Cho, K. W., B. Blumberg, H. Steinbeisser, and E. M. De Robertis. 1991. 'Molecular nature of Spemann's organizer: the role of the *Xenopus* homeobox gene *gooseoid*', *Cell*, 67: 1111-20.
- Christian, J. L., and R. T. Moon. 1993. 'Interactions between Xwnt-8 and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of *Xenopus*', *Genes Dev*, 7: 13-28.
- Chua, G., Q. D. Morris, R. Sopko, M. D. Robinson, O. Ryan, E. T. Chan, B. J. Frey, B. J. Andrews, C. Boone, and T. R. Hughes. 2006. 'Identifying transcription factor functions and targets by phenotypic activation', *Proc Natl Acad Sci U S A*, 103: 12045-50.
- Clute, P., and Y. Masui. 1995. 'Regulation of the appearance of division asynchrony and microtubule-dependent chromosome cycles in *Xenopus laevis* embryos', *Dev Biol*, 171: 273-85.
- Colas, A., J. Cartry, I. Buisson, M. Umbhauer, J. C. Smith, and J. F. Riou. 2008. 'Mix.1/2-dependent control of FGF availability during gastrulation is essential for pronephros development in *Xenopus*', *Dev Biol*, 320: 351-65.

- Coll, O., A. Villalba, G. Bussotti, C. Notredame, and F. Gebauer. 2010. 'A novel, noncanonical mechanism of cytoplasmic polyadenylation operates in *Drosophila* embryogenesis', *Genes Dev*, 24: 129-34.
- Collart, C., G. E. Allen, C. R. Bradshaw, J. C. Smith, and P. Zegerman. 2013. 'Titration of four replication factors is essential for the *Xenopus laevis* midblastula transition', *Science*, 341: 893-6.
- Collart, C., N. D. Owens, L. Bhaw-Rosun, B. Cooper, E. De Domenico, I. Patrushev, A. K. Sesay, J. N. Smith, J. C. Smith, and M. J. Gilchrist. 2014. 'High-resolution analysis of gene activity during the *Xenopus* mid-blastula transition', *Development*, 141: 1927-39.
- Conlon, F. L., S. G. Sedgwick, K. M. Weston, and J. C. Smith. 1996. 'Inhibition of Xbra transcription activation causes defects in mesodermal patterning and reveals autoregulation of Xbra in dorsal mesoderm', *Development*, 122: 2427-35.
- Cooke, J. 1972. 'Properties of the primary organization field in the embryo of *Xenopus laevis*. II. Positional information for axial organization in embryos with two head organizers', *J Embryol Exp Morphol*, 28: 27-46.
- Cornell, R. A., T. J. Musci, and D. Kimelman. 1995. 'FGF is a prospective competence factor for early activin-type signals in *Xenopus* mesoderm induction', *Development*, 121: 2429-37.
- Dale, L., G. Howes, B. M. Price, and J. C. Smith. 1992. 'Bone morphogenetic protein 4: a ventralizing factor in early *Xenopus* development', *Development*, 115: 573-85.
- Davidson, E. H., J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C. H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, A. G. Rust, Zj Pan, M. J. Schilstra, P. J. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood, and H. Bolouri. 2002. 'A genomic regulatory network for development', *Science*, 295: 1669-78.
- De Domenico, E., N. D. Owens, I. M. Grant, R. Gomes-Faria, and M. J. Gilchrist. 2015. 'Molecular asymmetry in the 8-cell stage *Xenopus tropicalis* embryo described by single blastomere transcript sequencing', *Dev Biol*.
- De Robertis, E. M. 2006. 'Spemann's organizer and self-regulation in amphibian embryos', *Nat Rev Mol Cell Biol*, 7: 296-302.
- De Roberts, E. M., M. Blum, C. Niehrs, and H. Steinbeisser. 1992. 'Goosecoid and the organizer', *Dev Suppl*: 167-71.
- Del Giacco, L., A. Pistocchi, and A. Ghilardi. 2010. 'prox1b Activity is essential in zebrafish lymphangiogenesis', *PLoS One*, 5: e13170.
- Detivaud, L., G. Pascreau, A. Karaïskou, H. B. Osborne, and J. Z. Kubiak. 2003. 'Regulation of EDEN-dependent deadenylation of Aurora A/Eg2-derived mRNA via phosphorylation and dephosphorylation in *Xenopus laevis* egg extracts', *J Cell Sci*, 116: 2697-705.
- Dickinson, K., J. Leonard, and J. C. Baker. 2006. 'Genomic profiling of mixer and Sox17beta targets during *Xenopus* endoderm development', *Dev Dyn*, 235: 368-81.
- Dickson, K. S., A. Bilger, S. Ballantyne, and M. P. Wickens. 1999. 'The cleavage and polyadenylation specificity factor in *Xenopus laevis* oocytes is a cytoplasmic factor involved in regulated polyadenylation', *Mol Cell Biol*, 19: 5707-17.
- Doherty, J. R., H. Zhu, E. Kulyev, and P. E. Mead. 2006. 'Determination of the minimal domains of Mix.3/Mixer required for endoderm development', *Mech Dev*, 123: 56-66.
- Dosch, R., and C. Niehrs. 2000. 'Requirement for anti-dorsalizing morphogenetic protein in organizer patterning', *Mech Dev*, 90: 195-203.

- Dubreuil, V., M. R. Hirsch, A. Pattyn, J. F. Brunet, and C. Goriadis. 2000. 'The Phox2b transcription factor coordinately regulates neuronal cell cycle exit and identity', *Development*, 127: 5191-201.
- Duncan, D. S., A. Ruzov, J. A. Hackett, and R. R. Meehan. 2008. 'xNmt1 regulates transcriptional silencing in pre-MBT *Xenopus* embryos independently of its catalytic function', *Development*, 135: 1295-302.
- Duval, C., P. Bouvet, F. Omilli, C. Roghi, C. Dorel, R. LeGuellec, J. Paris, and H. B. Osborne. 1990. 'Stability of maternal mRNA in *Xenopus* embryos: role of transcription and translation', *Mol Cell Biol*, 10: 4123-9.
- Eisen, J. S., and J. C. Smith. 2008. 'Controlling morpholino experiments: don't stop making antisense', *Development*, 135: 1735-43.
- Ensembl. 2015. 'Ensembl Genome Browser'.
- Fletcher, R. B., J. C. Baker, and R. M. Harland. 2006. 'FGF8 spliceforms mediate early mesoderm and posterior neural tissue formation in *Xenopus*', *Development*, 133: 1703-14.
- Fletcher, R. B., and R. M. Harland. 2008. 'The role of FGF signaling in the establishment and maintenance of mesodermal gene expression in *Xenopus*', *Dev Dyn*, 237: 1243-54.
- Fukuda, M., S. Takahashi, Y. Haramoto, Y. Onuma, Y. J. Kim, C. Y. Yeo, S. Ishiura, and M. Asashima. 2010. 'Zygotic VegT is required for *Xenopus* paraxial mesoderm formation and is regulated by Nodal signaling and Eomesodermin', *Int J Dev Biol*, 54: 81-92.
- Gao, F., B. C. Foat, and H. J. Bussemaker. 2004. 'Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data', *BMC Bioinformatics*, 5: 31.
- Garner, M. M., and A. Revzin. 1981. 'A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system', *Nucleic Acids Res*, 9: 3047-60.
- Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D. L. Dunaway, H. P. Fell, S. Ferree, R. D. George, T. Grogan, J. J. James, M. Maysuria, J. D. Mitton, P. Oliveri, J. L. Osborn, T. Peng, A. L. Ratcliffe, P. J. Webster, E. H. Davidson, L. Hood, and K. Dimitrov. 2008. 'Direct multiplexed measurement of gene expression with color-coded probe pairs', *Nat Biotechnol*, 26: 317-25.
- Gentsch, G. E., N. D. Owens, S. R. Martin, P. Piccinelli, T. Faial, M. W. Trotter, M. J. Gilchrist, and J. C. Smith. 2013. 'In vivo T-box transcription factor profiling reveals joint regulation of embryonic neuromesodermal bipotency', *Cell Rep*, 4: 1185-96.
- Gerety, S. S., and D. G. Wilkinson. 2011. 'Morpholino artifacts provide pitfalls and reveal a novel role for pro-apoptotic genes in hindbrain boundary development', *Dev Biol*, 350: 279-89.
- Gilchrist, M. J., A. M. Zorn, J. Voigt, J. C. Smith, N. Papalopulu, and E. Amaya. 2004. 'Defining a large set of full-length clones from a *Xenopus tropicalis* EST project', *Dev Biol*, 271: 498-516.
- Glavic, A., J. L. Gomez-Skarmeta, and R. Mayor. 2001. 'Xiro-1 controls mesoderm patterning by repressing *bmp-4* expression in the Spemann organizer', *Dev Dyn*, 222: 368-76.
- Glenwinkel, L., D. Wu, G. Minevich, and O. Hobert. 2014. 'TargetOrtho: a phylogenetic footprinting tool to identify transcription factor targets', *Genetics*, 197: 61-76.
- Gomez-Skarmeta, J., E. de La Calle-Mustienes, and J. Modolell. 2001. 'The Wnt-activated Xiro1 gene encodes a repressor that is essential for neural development and downregulates *Bmp4*', *Development*, 128: 551-60.
- Graindorge, A., O. Le Tonqueze, R. Thuret, N. Pollet, H. B. Osborne, and Y. Audic. 2008. 'Identification of CUG-BP1/EDEN-BP target mRNAs in *Xenopus tropicalis*', *Nucleic Acids Research*, 36: 1861-70.

- Graindorge, A., R. Thuret, N. Pollet, H. B. Osborne, and Y. Audic. 2006. 'Identification of post-transcriptionally regulated *Xenopus tropicalis* maternal mRNAs by microarray', *Nucleic Acids Res*, 34: 986-95.
- Graphpad. 2015. 'graphpad quick calcs, analyze a 2x2 contingency table'. <http://graphpad.com/quickcalcs/contingency2/>.
- Groisman, I., Y. S. Huang, R. Mendez, Q. Cao, W. Theurkauf, and J. D. Richter. 2000. 'CPEB, maskin, and cyclin B1 mRNA at the mitotic apparatus: implications for local translational control of cell division', *Cell*, 103: 435-47.
- Guo, W., A. P. Chan, H. Liang, E. D. Wieder, J. J. Molldrem, L. D. Etkin, and L. Nagarajan. 2002. 'A human Mix-like homeobox gene MIXL shows functional similarity to *Xenopus Mix.1*', *Blood*, 100: 89-95.
- Hair, A., M. N. Prioleau, Y. Vassetzky, and M. Mechali. 1998. 'Control of gene expression in *Xenopus* early development', *Dev Genet*, 22: 122-31.
- Hake, L. E., R. Mendez, and J. D. Richter. 1998. 'Specificity of RNA binding by CPEB: requirement for RNA recognition motifs and a novel zinc finger', *Mol Cell Biol*, 18: 685-93.
- Hallikas, O., K. Palin, N. Sinjushina, R. Rautiainen, J. Partanen, E. Ukkonen, and J. Taipale. 2006. 'Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity', *Cell*, 124: 47-59.
- Hart, A. H., L. Hartley, K. Sourris, E. S. Stadler, R. Li, E. G. Stanley, P. P. Tam, A. G. Elefanty, and L. Robb. 2002. 'Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo', *Development*, 129: 3597-608.
- Harvey, S. A., I. Sealy, R. Kettleborough, F. Fenyes, R. White, D. Stemple, and J. C. Smith. 2013. 'Identification of the zebrafish maternal and paternal transcriptomes', *Development*, 140: 2703-10.
- Hawley, S. H., K. Wunnenberg-Stapleton, C. Hashimoto, M. N. Laurent, T. Watabe, B. W. Blumberg, and K. W. Cho. 1995. 'Disruption of BMP signals in embryonic *Xenopus* ectoderm leads to direct neural induction', *Genes Dev*, 9: 2923-35.
- Heasman, J. 2006. 'Patterning the early *Xenopus* embryo', *Development*, 133: 1205-17.
- Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, and D. S. Rokhsar. 2010. 'The genome of the Western clawed frog *Xenopus tropicalis*', *Science*, 328: 633-6.
- Hemmati-Brivanlou, A., O. G. Kelly, and D. A. Melton. 1994. 'Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity', *Cell*, 77: 283-95.
- Henry, G. L., and D. A. Melton. 1998. 'Mixer, a homeobox gene required for endoderm development', *Science*, 281: 91-6.
- Heyn, P., M. Kircher, A. Dahl, J. Kelso, P. Tomancak, A. T. Kalinka, and K. M. Neugebauer. 2014. 'The earliest transcribed zygotic genes are short, newly evolved, and different across species', *Cell Rep*, 6: 285-92.
- Horton, J. D., N. A. Shah, J. A. Warrington, N. N. Anderson, S. W. Park, M. S. Brown, and J. L. Goldstein. 2003. 'Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes', *Proc Natl Acad Sci U S A*, 100: 12027-32.

- Houston, D. W., and C. Wylie. 2003. 'The *Xenopus* LIM-homeodomain protein *Xlim5* regulates the differential adhesion properties of early ectoderm cells', *Development*, 130: 2695-704.
- Hu, Z. W., and B. B. Hoffman. 1993. 'Cycloheximide induces the alpha 1B adrenergic receptor gene by activation of transcription in DDT1 MF-2 smooth muscle cells', *Mol Pharmacol*, 44: 1105-12.
- Isaacs, H. V., M. E. Pownall, and J. M. Slack. 1994. 'eFGF regulates *Xbra* expression during *Xenopus* gastrulation', *EMBO J*, 13: 4469-81.
- Janssens, S., T. Denayer, T. Deroo, F. Van Roy, and K. Vleminckx. 2010. 'Direct control of *Hoxd1* and *Irx3* expression by Wnt/beta-catenin signaling during anteroposterior patterning of the neural axis in *Xenopus*', *Int J Dev Biol*, 54: 1435-42.
- Jones, C. M., K. M. Lyons, P. M. Lapan, C. V. Wright, and B. L. Hogan. 1992. 'DVR-4 (bone morphogenetic protein-4) as a posterior-ventralizing factor in *Xenopus* mesoderm induction', *Development*, 115: 639-47.
- Junker, J. P., E. S. Noel, V. Guryev, K. A. Peterson, G. Shah, J. Huisken, A. P. McMahon, E. Berezikov, J. Bakkers, and A. van Oudenaarden. 2014. 'Genome-wide RNA Tomography in the zebrafish embryo', *Cell*, 159: 662-75.
- Kao, K. R., and R. P. Elinson. 1988. 'The entire mesodermal mantle behaves as Spemann's organizer in dorsoanterior enhanced *Xenopus laevis* embryos', *Dev Biol*, 127: 64-77.
- Karczewski, K. J., M. Snyder, R. B. Altman, and N. P. Tatonetti. 2014. 'Coherent functional modules improve transcription factor target identification, cooperativity prediction, and disease association', *PLoS Genet*, 10: e1004122.
- Karpinka, J. B., J. D. Fortriede, K. A. Burns, C. James-Zorn, V. G. Ponferrada, J. Lee, K. Karimi, A. M. Zorn, and P. D. Vize. 2015. 'Xenbase, the *Xenopus* model organism database; new virtualized system, data types and genomes', *Nucleic Acids Res*, 43: D756-63.
- Kel, A. E., O. V. Kel-Margoulis, P. J. Farnham, S. M. Bartley, E. Wingender, and M. Q. Zhang. 2001. 'Computer-assisted identification of cell cycle-related genes: new targets for E2F transcription factors', *J Mol Biol*, 309: 99-120.
- Khadka, D., T. Luo, and T. D. Sargent. 2006. '*Msx1* and *Msx2* have shared essential functions in neural crest but may be dispensable in epidermis and axis formation in *Xenopus*', *Int J Dev Biol*, 50: 499-502.
- Khokha, M. K., J. Yeh, T. C. Grammer, and R. M. Harland. 2005. 'Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures', *Dev Cell*, 8: 401-11.
- Kikuchi, Y., L. A. Trinh, J. F. Reiter, J. Alexander, D. Yelon, and D. Y. Stainier. 2000. 'The zebrafish *bonnie* and *clyde* gene encodes a Mix family homeodomain protein that regulates the generation of endodermal precursors', *Genes Dev*, 14: 1279-89.
- Kimelman, D., M. Kirschner, and T. Scherson. 1987. 'The events of the midblastula transition in *Xenopus* are regulated by changes in the cell cycle', *Cell*, 48: 399-407.
- Knochel, S., A. Schuler-Metz, and W. Knochel. 2000. 'c-Jun (AP-1) activates BMP-4 transcription in *Xenopus* embryos', *Mech Dev*, 98: 29-36.
- Kofron, M., C. Wylie, and J. Heasman. 2004. 'The role of Mixer in patterning the early *Xenopus* embryo', *Development*, 131: 2431-41.
- Kok, F. O., M. Shin, C. W. Ni, A. Gupta, A. S. Grosse, A. van Impel, B. C. Kirchmaier, J. Peterson-Maduro, G. Kourkoulis, I. Male, D. F. DeSantis, S. Sheppard-Tindell, L. Ebarasi, C. Betsholtz, S. Schulte-Merker, S. A. Wolfe, and N. D. Lawson. 2015. 'Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish', *Dev Cell*, 32: 97-108.
- Koster, M., K. Dillinger, and W. Knochel. 1998. 'Expression pattern of the winged helix factor XFD-11 during *Xenopus* embryogenesis', *Mech Dev*, 76: 169-73.

- Kuge, H., and J. D. Richter. 1995. 'Cytoplasmic 3' poly(A) addition induces 5' cap ribose methylation: implications for translational control of maternal mRNA', *EMBO J*, 14: 6301-10.
- Kumano, G., and W. C. Smith. 2000. 'FGF signaling restricts the primary blood islands to ventral mesoderm', *Dev Biol*, 228: 304-14.
- Kurauchi, T., Y. Izutsu, and M. Maeno. 2010. 'Involvement of Neptune in induction of the hatching gland and neural crest in the *Xenopus* embryo', *Differentiation*, 79: 251-9.
- Kuroda, H., O. Wessely, and E. M. De Robertis. 2004. 'Neural induction in *Xenopus*: requirement for ectodermal and endomesodermal signals via Chordin, Noggin, beta-Catenin, and Cerberus', *PLoS Biol*, 2: E92.
- Kwon, T., M. I. Chung, R. Gupta, J. C. Baker, J. B. Wallingford, and E. M. Marcotte. 2014. 'Identifying direct targets of transcription factor Rfx2 that coordinate ciliogenesis and cell movement', *Genom Data*, 2: 192-94.
- Lane, M. C., and M. D. Sheets. 2000. 'Designation of the anterior/posterior axis in pregastrula *Xenopus laevis*', *Dev Biol*, 225: 37-58.
- Latinkic, B. V., and J. C. Smith. 1999. 'Goosecoid and mix.1 repress Brachyury expression and are required for head formation in *Xenopus*', *Development*, 126: 1769-79.
- Latinkic, B. V., M. Umbhauer, K. A. Neal, W. Lerchner, J. C. Smith, and V. Cunliffe. 1997. 'The *Xenopus* Brachyury promoter is activated by FGF and low concentrations of activin and suppressed by high concentrations of activin and by paired-type homeodomain proteins', *Genes Dev*, 11: 3265-76.
- Lee, J. R., A. M. Derosa, and T. W. White. 2009. 'Connexin mutations causing skin disease and deafness increase hemichannel activity and cell death when expressed in *Xenopus* oocytes', *J Invest Dermatol*, 129: 870-8.
- Lee, M. T., A. R. Bonneau, C. M. Takacs, A. A. Bazzini, K. R. DiVito, E. S. Fleming, and A. J. Giraldez. 2013. 'Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition', *Nature*, 503: 360-4.
- Lee, T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. 2002. 'Transcriptional regulatory networks in *Saccharomyces cerevisiae*', *Science*, 298: 799-804.
- Legagneux, V., F. Omilli, and H. B. Osborne. 1995. 'Substrate-specific regulation of RNA deadenylation in *Xenopus* embryo and activated egg extracts', *RNA*, 1: 1001-8.
- Lemaire, P., S. Darras, D. Caillol, and L. Kodjabachian. 1998. 'A role for the vegetally expressed *Xenopus* gene Mix.1 in endoderm formation and in the restriction of mesoderm to the marginal zone', *Development*, 125: 2371-80.
- Lemaire, P., N. Garrett, and J. B. Gurdon. 1995. 'Expression cloning of Siamois, a *Xenopus* homeobox gene expressed in dorsal-vegetal cells of blastulae and able to induce a complete secondary axis', *Cell*, 81: 85-94.
- Lemaire, P., and J. B. Gurdon. 1994. 'A role for cytoplasmic determinants in mesoderm patterning: cell-autonomous activation of the goosecoid and Xwnt-8 genes along the dorsoventral axis of early *Xenopus* embryos', *Development*, 120: 1191-9.
- Li, P., S. S. Chaurasia, Y. Gao, A. L. Carr, P. M. Iuvone, and L. Li. 2008. 'CLOCK is required for maintaining the circadian rhythms of Opsin mRNA expression in photoreceptor cells', *J Biol Chem*, 283: 31673-8.
- Liang, H. L., C. Y. Nien, H. Y. Liu, M. M. Metzstein, N. Kirov, and C. Rushlow. 2008. 'The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*', *Nature*, 456: 400-3.
- Lim, S. M., L. Pereira, M. S. Wong, C. E. Hirst, B. E. Van Vranken, M. Pick, A. Trounson, A. G. Elefanty, and E. G. Stanley. 2009. 'Enforced expression of Mixl1 during mouse ES

- cell differentiation suppresses hematopoietic mesoderm and promotes endoderm formation', *Stem Cells*, 27: 363-74.
- Lindeman, L. C., I. S. Andersen, A. H. Reiner, N. Li, H. Aanes, O. Ostrup, C. Winata, S. Mathavan, F. Muller, P. Alestrom, and P. Collas. 2011. 'Prepatterning of developmental gene expression by modified histones before zygotic genome activation', *Dev Cell*, 21: 993-1004.
- Loose, M., and R. Patient. 2004. 'A genetic regulatory network for *Xenopus* mesendoderm formation', *Dev Biol*, 271: 467-78.
- Lowry, Richard. 2015. 'Log-Linear Analysis for an AxBxC Contingency Table'. <http://vassarstats.net/abc.html>.
- Lund, E., M. Liu, R. S. Hartley, M. D. Sheets, and J. E. Dahlberg. 2009. 'Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos', *RNA*, 15: 2351-63.
- Luo, T., Y. H. Lee, J. P. Saint-Jeannet, and T. D. Sargent. 2003. 'Induction of neural crest in *Xenopus* by transcription factor AP2alpha', *Proc Natl Acad Sci U S A*, 100: 532-7.
- Luu, O., M. Nagel, S. Wacker, P. Lemaire, and R. Winklbauer. 2008. 'Control of gastrula cell motility by the Goosecoid/Mix.1/ Siamois network: basic patterns and paradoxical effects', *Dev Dyn*, 237: 1307-20.
- Luxardi, G., L. Marchal, V. Thome, and L. Kodjabachian. 2010. 'Distinct *Xenopus* Nodal ligands sequentially induce mesendoderm and control gastrulation movements in parallel to the Wnt/PCP pathway', *Development*, 137: 417-26.
- Matsuo-Takasaki, M., M. Matsumura, and Y. Sasai. 2005. 'An essential role of *Xenopus* Foxi1a for ventral specification of the cephalic ectoderm during gastrulation', *Development*, 132: 3885-94.
- Mead, P. E., I. H. Brivanlou, C. M. Kelley, and L. I. Zon. 1996. 'BMP-4-responsive regulation of dorsal-ventral patterning by the homeobox protein Mix.1', *Nature*, 382: 357-60.
- Mead, P. E., C. M. Kelley, P. S. Hahn, O. Piedad, and L. I. Zon. 1998. 'SCL specifies hematopoietic mesoderm in *Xenopus* embryos', *Development*, 125: 2611-20.
- Mead, P. E., Y. Zhou, K. D. Lustig, T. L. Huber, M. W. Kirschner, and L. I. Zon. 1998. 'Cloning of Mix-related homeodomain proteins using fast retrieval of gel shift activities, (FROGS), a technique for the isolation of DNA-binding proteins', *Proc Natl Acad Sci U S A*, 95: 11251-6.
- Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein. 2007. 'Genome-wide maps of chromatin state in pluripotent and lineage-committed cells', *Nature*, 448: 553-60.
- Mohn, D., S. W. Chen, D. C. Dias, D. C. Weinstein, M. A. Dyer, K. Sahr, C. E. Ducker, E. Zahradka, G. Keller, K. S. Zaret, L. J. Gudas, and M. H. Baron. 2003. 'Mouse Mix gene is activated early during differentiation of ES and F9 stem cells and induces endoderm in frog embryos', *Dev Dyn*, 226: 446-59.
- Moos, M., Jr., S. Wang, and M. Krinks. 1995. 'Anti-dorsalizing morphogenetic protein is a novel TGF-beta homolog expressed in the Spemann organizer', *Development*, 121: 4293-301.
- Nascone, N., and M. Mercola. 1995. 'An inductive role for the endoderm in *Xenopus* cardiogenesis', *Development*, 121: 515-23.
- Newport, J., and M. Kirschner. 1982a. 'A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage', *Cell*, 30: 675-86.
- . 1982b. 'A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription', *Cell*, 30: 687-96.

- Nieuwkoop, P. D. , and J. Faber. 1994. *Normal Table of Xenopus Laevis* (Garland Science).
- Nutt, S. L., O. J. Bronchain, K. O. Hartley, and E. Amaya. 2001. 'Comparison of morpholino based translational inhibition during the development of *Xenopus laevis* and *Xenopus tropicalis*', *Genesis*, 30: 110-3.
- Oelgeschlager, M., H. Kuroda, B. Reversade, and E. M. De Robertis. 2003. 'Chordin is required for the Spemann organizer transplantation phenomenon in *Xenopus* embryos', *Dev Cell*, 4: 219-30.
- Owens, N. D. L., Ira L. Blitz, Maura A. Lane, Ilya Patrushev, John D. Overton, Michael J. Gilchrist, Ken W. Y. Cho, Mustafa K. Khokha. 2015. 'Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development', *Under Review to Cell Reports*.
- Paranjpe, S. S., U. G. Jacobi, S. J. van Heeringen, and G. J. Veenstra. 2013. 'A genome-wide survey of maternal and embryonic transcripts during *Xenopus tropicalis* development', *BMC Genomics*, 14: 762.
- Paris, J., H. B. Osborne, A. Couturier, R. Le Guellec, and M. Philippe. 1988. 'Changes in the polyadenylation of specific stable RNA during the early development of *Xenopus laevis*', *Gene*, 72: 169-76.
- Paris, J., and M. Philippe. 1990. 'Poly(A) metabolism and polysomal recruitment of maternal mRNAs during early *Xenopus* development', *Dev Biol*, 140: 221-4.
- Pattyn, A., M. Hirsch, C. Goridis, and J. F. Brunet. 2000. 'Control of hindbrain motor neuron differentiation by the homeobox gene *Phox2b*', *Development*, 127: 1349-58.
- Peale, F. V., Jr., L. Sugden, and M. Bothwell. 1998. 'Characterization of CMIX, a chicken homeobox gene related to the *Xenopus* gene *mix.1*', *Mech Dev*, 75: 167-70.
- Peng, G., and M. Westerfield. 2006. 'Lhx5 promotes forebrain development and activates transcription of secreted Wnt antagonists', *Development*, 133: 3191-200.
- Pereira, L. A., M. S. Wong, S. Mei Lim, E. G. Stanley, and A. G. Elefanty. 2012. 'The Mix family of homeobox genes--key regulators of mesendoderm formation during vertebrate development', *Dev Biol*, 367: 163-77.
- Piccioni, F., V. Zappavigna, and A. C. Verrotti. 2005. 'Translational regulation during oogenesis and early development: the cap-poly(A) tail relationship', *C R Biol*, 328: 863-81.
- Piccolo, S., E. Agius, L. Leyns, S. Bhattacharyya, H. Grunz, T. Bouwmeester, and E. M. De Robertis. 1999. 'The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals', *Nature*, 397: 707-10.
- Piccolo, S., Y. Sasai, B. Lu, and E. M. De Robertis. 1996. 'Dorsoventral patterning in *Xenopus*: inhibition of ventral signals by direct binding of chordin to BMP-4', *Cell*, 86: 589-98.
- Plouhinec, J. L., D. D. Roche, C. Pegoraro, A. L. Figueiredo, F. Maczkowiak, L. J. Brunet, C. Milet, J. P. Vert, N. Pollet, R. M. Harland, and A. H. Monsoro-Burq. 2014. 'Pax3 and Zic1 trigger the early neural crest gene regulatory network by the direct activation of multiple key neural crest specifiers', *Dev Biol*, 386: 461-72.
- Poulain, M., and T. Lepage. 2002. 'Mezzo, a paired-like homeobox protein is an immediate target of Nodal signalling and regulates endoderm specification in zebrafish', *Development*, 129: 4901-14.
- Radford, H. E., H. A. Meijer, and C. H. de Moor. 2008. 'Translational control by cytoplasmic polyadenylation in *Xenopus* oocytes', *Biochim Biophys Acta*, 1779: 217-29.
- Ramagopalan, S. V., A. Heger, A. J. Berlanga, N. J. Mauger, M. R. Lincoln, A. Burrell, L. Handunnetthi, A. E. Handel, G. Disanto, S. M. Orton, C. T. Watson, J. M. Morahan, G. Giovannoni, C. P. Ponting, G. C. Ebers, and J. C. Knight. 2010. 'A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution', *Genome Res*, 20: 1352-60.

- Reversade, B., and E. M. De Robertis. 2005. 'Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field', *Cell*, 123: 1147-60.
- Reversade, B., H. Kuroda, H. Lee, A. Mays, and E. M. De Robertis. 2005. 'Depletion of Bmp2, Bmp4, Bmp7 and Spemann organizer signals induces massive brain formation in *Xenopus* embryos', *Development*, 132: 3381-92.
- Richter, J. D. 1999. 'Cytoplasmic polyadenylation in development and beyond', *Microbiol Mol Biol Rev*, 63: 446-56.
- Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O. L. Griffith, A. He, M. Marra, M. Snyder, and S. Jones. 2007. 'Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing', *Nat Methods*, 4: 651-7.
- Rodriguez-Seguel, E., P. Alarcon, and J. L. Gomez-Skarmeta. 2009. 'The *Xenopus* *Irx* genes are essential for neural patterning and define the border between prethalamus and thalamus through mutual antagonism with the anterior repressors *Fezf* and *Arx*', *Dev Biol*, 329: 258-68.
- Rosa, F. M. 1989. 'Mix.1, a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumptive endodermal cells of *Xenopus* embryos', *Cell*, 57: 965-74.
- Rossi, A., Z. Kontarakis, C. Gerri, H. Nolte, S. Holper, M. Kruger, and D. Y. Stainier. 2015. 'Genetic compensation induced by deleterious mutations but not gene knockdowns', *Nature*, 524: 230-3.
- Sagata, N., K. Shiokawa, and K. Yamana. 1980. 'A study on the steady-state population of poly(A)+RNA during early development of *Xenopus laevis*', *Dev Biol*, 77: 431-48.
- Sahr, K., D. C. Dias, R. Sanchez, D. Chen, S. W. Chen, L. J. Gudas, and M. H. Baron. 2002. 'Structure, upstream promoter region, and functional domains of a mouse and human Mix paired-like homeobox gene', *Gene*, 291: 135-47.
- Sakabe, N. J., I. Aneas, T. Shen, L. Shokri, S. Y. Park, M. L. Bulyk, S. M. Evans, and M. A. Nobrega. 2012. 'Dual transcriptional activator and repressor roles of TBX20 regulate adult cardiac structure and function', *Hum Mol Genet*, 21: 2194-204.
- Salles, F. J., M. E. Lieberfarb, C. Wreden, J. P. Gergen, and S. Strickland. 1994. 'Coordinate initiation of *Drosophila* development by regulated polyadenylation of maternal messenger RNAs', *Science*, 266: 1996-9.
- Sander, V., B. Reversade, and E. M. De Robertis. 2007. 'The opposing homeobox genes Goosecoid and Vent1/2 self-regulate *Xenopus* patterning', *EMBO J*, 26: 2955-65.
- Scerbo, P., F. Girardot, C. Vivien, G. V. Markov, G. Luxardi, B. Demeneix, L. Kodjabachian, and L. Coen. 2012. 'Ventx factors function as Nanog-like guardians of developmental potential in *Xenopus*', *PLoS One*, 7: e36855.
- Schroeder, K. E., M. L. Condic, L. M. Eisenberg, and H. J. Yost. 1999. 'Spatially regulated translation in embryos: asymmetric expression of maternal Wnt-11 along the dorsal-ventral axis in *Xenopus*', *Dev Biol*, 214: 288-97.
- Schulte-Merker, S., and J. C. Smith. 1995. 'Mesoderm formation in response to Brachyury requires FGF signalling', *Curr Biol*, 5: 62-7.
- Seiliez, I., B. Thisse, and C. Thisse. 2006. 'FoxA3 and goosecoid promote anterior neural fate through inhibition of Wnt8a activity before the onset of gastrulation', *Dev Biol*, 290: 152-63.
- Seufert, D. W., N. L. Prescott, and H. M. El-Hodiri. 2005. '*Xenopus* aristaless-related homeobox (xARX) gene product functions as both a transcriptional activator and repressor in forebrain development', *Dev Dyn*, 232: 313-24.

- Shimuta, K., N. Nakajo, K. Uto, Y. Hayano, K. Okazaki, and N. Sagata. 2002. 'Chk1 is activated transiently and targets Cdc25A for degradation at the *Xenopus* midblastula transition', *EMBO J*, 21: 3694-703.
- Sikora-Wohlfeld, W., M. Ackermann, E. G. Christodoulou, K. Singaravelu, and A. Beyer. 2013. 'Assessing computational methods for transcription factor target gene identification based on ChIP-seq data', *PLoS Comput Biol*, 9: e1003342.
- Silva, A. C., M. Filipe, K. M. Kuerner, H. Steinbeisser, and J. A. Belo. 2003. 'Endogenous Cerberus activity is required for anterior head specification in *Xenopus*', *Development*, 130: 4943-53.
- Simon, I., J. Barnett, N. Hannett, C. T. Harbison, N. J. Rinaldi, T. L. Volkert, J. J. Wyrick, J. Zeitlinger, D. K. Gifford, T. S. Jaakkola, and R. A. Young. 2001. 'Serial regulation of transcriptional regulators in the yeast cell cycle', *Cell*, 106: 697-708.
- Simon, R., and J. D. Richter. 1994. 'Further analysis of cytoplasmic polyadenylation in *Xenopus* embryos and identification of embryonic cytoplasmic polyadenylation element-binding proteins', *Mol Cell Biol*, 14: 7867-75.
- Simon, R., J. P. Tassan, and J. D. Richter. 1992. 'Translational control by poly(A) elongation during *Xenopus* development: differential repression and enhancement by a novel cytoplasmic polyadenylation element', *Genes Dev*, 6: 2580-91.
- Sinner, D., P. Kirilenko, S. Rankin, E. Wei, L. Howard, M. Kofron, J. Heasman, H. R. Woodland, and A. M. Zorn. 2006. 'Global analysis of the transcriptional network controlling *Xenopus* endoderm formation', *Development*, 133: 1955-66.
- Skirkanich, J., G. Luxardi, J. Yang, L. Kodjabachian, and P. S. Klein. 2011. 'An essential role for transcription before the MBT in *Xenopus laevis*', *Dev Biol*, 357: 478-91.
- Sooknanan, Roy, Jim Pease, and Ken Doyle. 2010. 'Novel methods for rRNA removal and directional, ligation-free RNA-seq library preparation', *Nat Meth*, 7.
- Stancheva, I., O. El-Maarri, J. Walter, A. Niveleau, and R. R. Meehan. 2002. 'DNA methylation at promoter regions regulates the timing of gene activation in *Xenopus laevis* embryos', *Dev Biol*, 243: 155-65.
- Steinbeisser, H., A. Fainsod, C. Niehrs, Y. Sasai, and E. M. De Robertis. 1995. 'The role of *gsc* and BMP-4 in dorsal-ventral patterning of the marginal zone in *Xenopus*: a loss-of-function study using antisense RNA', *EMBO J*, 14: 5230-43.
- Stennard, F., G. Carnac, and J. B. Gurdon. 1996. 'The *Xenopus* T-box gene, Antipodean, encodes a vegetally localised maternal mRNA and can trigger mesoderm formation', *Development*, 122: 4179-88.
- Stutz, A., B. Conne, J. Huarte, P. Gubler, V. Volkel, P. Flandin, and J. D. Vassalli. 1998. 'Masking, unmasking, and regulated polyadenylation cooperate in the translational control of a dormant mRNA in mouse oocytes', *Genes Dev*, 12: 2535-48.
- Subtelny, A. O., S. W. Eichhorn, G. R. Chen, H. Sive, and D. P. Bartel. 2014. 'Poly(A)-tail profiling reveals an embryonic switch in translational control', *Nature*, 508: 66-71.
- Sudou, N., S. Yamamoto, H. Ogino, and M. Taira. 2012. 'Dynamic in vivo binding of transcription factors to cis-regulatory modules of *cer* and *gsc* in the stepwise formation of the Spemann-Mangold organizer', *Development*, 139: 1651-61.
- Summerton, J. E. 2007. 'Morpholino, siRNA, and S-DNA compared: impact of structure and mechanism of action on off-target effects and sequence specificity', *Curr Top Med Chem*, 7: 651-60.
- Summerton, J., and D. Weller. 1997. 'Morpholino antisense oligomers: design, preparation, and properties', *Antisense Nucleic Acid Drug Dev*, 7: 187-95.
- Tada, M., E. S. Casey, L. Fairclough, and J. C. Smith. 1998. 'Bix1, a direct target of *Xenopus* T-box genes, causes formation of ventral mesoderm and endoderm', *Development*, 125: 3997-4006.

- Takahashi, S., C. Yokota, K. Takano, K. Tanegashima, Y. Onuma, J. Goto, and M. Asashima. 2000. 'Two novel nodal-related genes initiate early inductive events in *Xenopus* Nieuwkoop center', *Development*, 127: 5319-29.
- Talikka, M., G. Stefani, A. H. Brivanlou, and K. Zimmerman. 2004. 'Characterization of *Xenopus* Phox2a and Phox2b defines expression domains within the embryonic nervous system and early heart field', *Gene Expr Patterns*, 4: 601-7.
- Tam, P. P., P. L. Khoo, S. L. Lewis, H. Bildsoe, N. Wong, T. E. Tsang, J. M. Gad, and L. Robb. 2007. 'Sequential allocation and global pattern of movement of the definitive endoderm in the mouse embryo during gastrulation', *Development*, 134: 251-60.
- Tan, J. T., V. Korzh, and Z. Gong. 1999. 'Expression of a zebrafish iroquois homeobox gene, *Ziro3*, in the midline axial structures and central nervous system', *Mech Dev*, 87: 165-8.
- Tan, M. H., K. F. Au, D. E. Leong, K. Foygel, W. H. Wong, and M. W. Yao. 2013. 'An Oct4-Sall4-Nanog network controls developmental progression in the pre-implantation mouse embryo', *Mol Syst Biol*, 9: 632.
- Tan, M. H., K. F. Au, A. L. Yablonovitch, A. E. Wills, J. Chuang, J. C. Baker, W. H. Wong, and J. B. Li. 2013. 'RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development', *Genome Res*, 23: 201-16.
- Tao, S., M. Witte, R. J. Bryson-Richardson, P. D. Currie, B. M. Hogan, and S. Schulte-Merker. 2011. 'Zebrafish *prox1b* mutants develop a lymphatic vasculature, and *prox1b* does not specifically mark lymphatic endothelial cells', *PLoS One*, 6: e28934.
- Trinh, L. A., D. Meyer, and D. Y. Stainier. 2003. 'The Mix family homeodomain gene *bonnie* and *clyde* functions with other components of the Nodal signaling pathway to regulate neural patterning in zebrafish', *Development*, 130: 4989-98.
- UCSC. 2015. 'UCSC Genome Browser', Accessed 23/07/2015. <https://genome.ucsc.edu/>.
- Ueno, S., and N. Sagata. 2002. 'Requirement for both EDEN and AUUUA motifs in translational arrest of *Mos* mRNA upon fertilization of *Xenopus* eggs', *Dev Biol*, 250: 156-67.
- Varnum, S. M., and W. M. Wormington. 1990. 'Deadenylation of maternal mRNAs during *Xenopus* oocyte maturation does not require specific cis-sequences: a default mechanism for translational control', *Genes Dev*, 4: 2278-86.
- Voeltz, G. K., and J. A. Steitz. 1998. 'AUUUA sequences direct mRNA deadenylation uncoupled from decay during *Xenopus* early development', *Mol Cell Biol*, 18: 7537-45.
- Wang, J., and R. E. Davis. 2014. 'Contribution of transcription to animal early development', *Transcription*, 5: e967602.
- Weeks, D. L., J. A. Walder, and J. M. Dagle. 1991. 'Cyclin B mRNA depletion only transiently inhibits the *Xenopus* embryonic cell cycle', *Development*, 111: 1173-8.
- Whitehead-Institute. 2015. 'Whitehead Institute for Biomedical Research Bioinformatics & Research Computing'. <http://jura.wi.mit.edu/bioc/tools/compare.php>.
- Wickstrom, E., K. A. Urtishak, M. Choob, X. Tian, N. Sternheim, L. M. Cross, A. Rubinstein, and S. A. Farber. 2004. 'Downregulation of gene expression with negatively charged peptide nucleic acids (PNAs) in zebrafish embryos', *Methods Cell Biol*, 77: 137-58.
- Willey, S., A. Ayuso-Sacido, H. Zhang, S. T. Fraser, K. E. Sahr, M. J. Adlam, M. Kyba, G. Q. Daley, G. Keller, and M. H. Baron. 2006. 'Acceleration of mesoderm development and expansion of hematopoietic progenitors in differentiating ES cells by the mouse Mix-like homeodomain transcription factor', *Blood*, 107: 3122-30.
- Wills, A., K. Dickinson, M. Khokha, and J. C. Baker. 2008. 'Bmp signaling is necessary and sufficient for ventrolateral endoderm specification in *Xenopus*', *Dev Dyn*, 237: 2177-86.

- Wilson, D., G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan. 1993. 'Cooperative dimerization of paired class homeo domains on DNA', *Genes Dev*, 7: 2120-34.
- Wolfe, A. D., and K. M. Downs. 2014. 'Mixl1 localizes to putative axial stem cell reservoirs and their posterior descendants in the mouse embryo', *Gene Expr Patterns*, 15: 8-20.
- Xanthos, J. B., M. Kofron, C. Wylie, and J. Heasman. 2001. 'Maternal VegT is the initiator of a molecular network specifying endoderm in *Xenopus laevis*', *Development*, 128: 167-80.
- Xie, J., H. Yin, T. D. Nichols, J. A. Yoder, and J. M. Horowitz. 2010. 'Sp2 is a maternally inherited transcription factor required for embryonic development', *J Biol Chem*, 285: 4153-64.
- Yamamoto, S., H. Hikasa, H. Ono, and M. Taira. 2003. 'Molecular link in the sequential induction of the Spemann organizer: direct activation of the cerberus gene by Xlim-1, Xotx2, Mix.1, and Siamois, immediately downstream from Nodal and Wnt signaling', *Dev Biol*, 257: 190-204.
- Yang, J., C. Tan, R. S. Darken, P. A. Wilson, and P. S. Klein. 2002. 'Beta-catenin/Tcf-regulated transcription prior to the midblastula transition', *Development*, 129: 5743-52.
- Yasuo, H., and P. Lemaire. 1999. 'A two-step model for the fate determination of presumptive endodermal blastomeres in *Xenopus* embryos', *Curr Biol*, 9: 869-79.
- Yokotal, C., T. Mukasa, M. Higashi, A. Odaka, K. Muroya, H. Uchiyama, Y. Eto, M. Asashima, and T. Momoi. 1995. 'Activin induces the expression of the *Xenopus* homologue of sonic hedgehog during mesoderm formation in *Xenopus* explants', *Biochem Biophys Res Commun*, 207: 1-7.
- Yoon, J., J. H. Kim, O. J. Lee, S. B. Yu, J. I. Kim, S. C. Kim, J. B. Park, J. Y. Lee, and J. Kim. 2011. 'xCITED2 Induces Neural Genes in Animal Cap Explants of *Xenopus* Embryos', *Exp Neurol*, 20: 123-9.
- Zhang, Y., and M. D. Sheets. 2009. 'Analyses of zebrafish and *Xenopus* oocyte maturation reveal conserved and diverged features of translational regulation of maternal cyclin B1 mRNA', *BMC Dev Biol*, 9: 7.
- Zimmerman, L. B., J. M. De Jesus-Escobar, and R. M. Harland. 1996. 'The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4', *Cell*, 86: 599-606.