

Investigating the role of individual fibroblast growth factor receptors in gene regulation during early amphibian development

Olivia Bryony Maude

Master's by Research

University of York

Biology

November 2020

Abstract

There is a good understanding of the developmental processes, including neural induction, and anteroposterior patterning, and target genes regulated by FGF signalling during early *Xenopus* development. However, several different members of the cell-surface receptor tyrosine kinases FGFR family are expressed in early development and the roles of individual FGFRs in mediating the effects of FGF signalling in *Xenopus* development are not clear. The aim of this project was to investigate the unique roles of FGFRs in mediating FGF signalling, through a meta-analysis of high-throughput transcriptomic data sets and CRISPR/Cas9 genome-modification protocol.

The development of a CRISPR/Cas9 protocol targeting *fgfr1*, *fgfr4* and *fgfr11*, involved optimisation of embryo injections and fragment analysis, to determine CRISPR/Cas9 targeting efficiency. This protocol enables gene expression analysis of FGF target genes, identified by a meta-analysis of high-throughput transcriptomic RNA-Seq and microarray data sets, and elucidates the unique roles of individual FGFRs in regulating the expression of FGF target genes and thus patterning *X. tropicalis* embryos.

This thesis presents evidence towards the unique roles of FGFRs in development through bioinformatic analyses in regulating different genes, such as *egr1*, *fos* and different members of *dusp* and *hes* gene families. The CRISPR/Cas9 protocol demonstrated phenotypic defects and mutations in *fgfr1*, *fgfr4* and *fgfr11*, however this necessitates further refinement to improve targeting efficiencies. However, this gene-editing technique could be expanded to investigate the individual roles of FGF ligands in early *Xenopus* development.

Table of Contents

Abstract.....	2
Table of Contents.....	3
List of Figures.....	8
List of Tables.....	10
List of Supplementary Figures.....	11
List of Supplementary Tables.....	12
Acknowledgements.....	15
Author’s Declaration.....	16
Chapter 1: Introduction.....	17
1.1 Fibroblast growth factor ligands.....	17
1.1.1 Ligand evolution.....	17
1.1.2 Ligand classification.....	17
1.1.3 Ligand expression.....	18
1.1.4 Ligand regulation.....	18
1.2 Fibroblast growth factor receptors.....	19
1.2.1 Receptor evolution.....	20
1.2.2 Receptor structure.....	20
1.2.3 Receptor expression.....	21
1.2.4 FGFRL1.....	21
1.3 Fibroblast growth factor signal transduction.....	22
1.3.1 Phosphoinositide-3 kinase.....	22
1.3.2 Phospholipase C gamma.....	22
1.3.3 Mitogen-activated protein kinase.....	23
1.3.4 Transcriptional regulation by FGF signal transduction.....	23
1.3.5 Regulation of FGF signal transduction.....	23
1.4 Patterning the early <i>Xenopus</i> embryo.....	25
1.4.1 Mesoderm induction.....	25
1.4.2 Dorsal-ventral patterning.....	26
1.4.3 Neural induction.....	27

1.4.4 Anteroposterior patterning.....	29
1.4.5 Left-right asymmetry.....	29
1.5 Summary.....	30
1.6 CRISPR/Cas9.....	31
1.7 Project aims.....	33
Chapter 2: Materials and Methods.....	34
2.1 Embryological methods.....	34
2.1.1 <i>X. tropicalis</i> embryo <i>in vitro</i> fertilisations.....	34
2.1.2 Microinjection.....	34
2.1.3 Photography.....	34
2.2 Molecular biology methods.....	35
2.2.1 Gel electrophoresis.....	35
2.2.2 Nanodrop.....	35
2.2.3 sgRNA design.....	35
2.2.4 sgRNA synthesis.....	35
2.2.5 Phenol chloroform extraction and ethanol precipitation.....	36
2.2.6 Genomic DNA extraction.....	37
2.2.6.1 Initial genomic DNA extraction.....	37
2.2.6.2 Optimised genomic DNA extraction.....	37
2.2.7 PCR optimisation.....	37
2.2.8 Two-step PCR for fragment analysis.....	37
2.2.9 One-step PCR for fragment analysis.....	39
2.2.10 Fragment analysis.....	39
2.2.11 Plasmid transformation.....	39
2.2.12 DNA minipreps.....	39
2.2.13 Template linearization.....	40
2.3 <i>In situ</i> hybridisation.....	40
2.3.1 Synthesis of <i>in situ</i> hybridisation probe by DIG transcription.....	40
2.3.2 <i>In situ</i> hybridisation.....	41
2.4. Bioinformatics analysis.....	42
2.4.1 Data set generation and initial processing.....	42
2.4.2 Filtering.....	42

2.4.2.1 BRB Analysis.....	42
2.4.3 Differentially expressed gene list analysis.....	42
Chapter 3: Identifying well supported FGF target genes through a meta-analysis of high-throughput transcriptomic data sets.....	44
3.1 Introduction.....	44
3.2 Results.....	46
3.2.1 Transcriptomic analysis of the effects of FGF signalling inhibition by dnFGFRs in whole embryos.....	46
3.2.1.1 dnFGFRs alters similar gene expression profiles.....	47
3.2.1.2 Genes affected by dnFGFR1 and dnFGFR4 are involved in similar biological processes in early gastrula stage whole embryos...	49
3.2.1.3 dnFGFR4 positively regulates interacting proteins in FGF, BMP and Wnt signalling feedback loops.....	52
3.2.2 Investigating genes regulated by FGFR1 and FGFR2 using iFGFR constructs in whole embryos.....	54
3.2.2.1 iFGFR signalling alters the gene expression profile in gastrula stage whole embryos.....	54
3.2.2.2 iFGFR1 and iFGFR2 affect the expression of overlapping genes in gastrula stage whole embryos.....	57
3.2.2.3 iFGFR1 up regulates genes in gastrula stages in whole embryos, which function in biological processes associated with FGF signalling.....	58
3.2.2.4 iFGFR protein-protein interaction networks contain Wnt and GTPase proteins.....	60
3.2.3 dnFGFR and iFGFR microarray data set comparison.....	61
3.2.3.1 Identification of well-supported FGF target genes, in dnFGFR4, iFGFR1 and iFGFR2 data sets.....	61
3.2.3.2 Low stringency filtering data set comparison.....	64
3.2.3.3 Microarray analysis summary.....	66
3.2.4 Investigating FGF-regulated genes using FGF4 overexpression in early neurula stage embryos.....	67
3.2.4.1 FGF4 overexpression alters gene expression profiles in early neurula stage whole embryos.....	67
3.2.4.2 Many differentially expressed genes function in FGF signalling feedback.....	68

3.2.4.3	Identification of interacting transcription factors in early neurula stage whole embryos.....	70
3.2.5	Investigating the differences in the FGFR1 and FGFR4-regulated transcriptome during neural development using iFGFR constructs in animal cap explants.....	70
3.2.5.1	A large number of genes are affected by iFGFR1 and iFGFR4 induction in neuralised animal cap explants.....	71
3.2.5.2	High stringency filtering of iFGFR1/4 gene expression data....	71
3.2.5.3	iFGFR1 and iFGFR4 affect the expression of overlapping groups of genes.....	73
3.2.5.4	Differences in gene ontologies of iFGFR1 and iFGFR4 regulated genes in neuralised animal cap explants.....	75
3.2.5.5	Similarities between PPI networks of genes affected by iFGFR1/4 signalling in neuralised animal cap explants.....	78
3.2.6	CSKA-FGF4 and iFGFR1/4 RNA-Seq data set comparison.....	81
3.2.6.1	Identification of well-supported FGF target genes, regulated by FGF4, FGFR1 and FGFR4 signalling.....	81
3.2.6.2	Heatmap comparisons.....	81
3.2.6.3	RNA-Seq analysis summary.....	85
3.2.7	In situ hybridisations of FGF target genes.....	86
3.8	Summary.....	87
3.3	Discussion.....	88
3.3.1	Unique roles of individual FGFRs.....	88
3.3.2	Unexpected contributions of FGFRs to biological processes.....	89
3.3.3	Target gene selection.....	90
3.3.3.1	Well-supported FGF target genes.....	90
3.3.3.2	Putative genes uniquely regulated by FGFR1 or FGFR4.....	91
3.3.4	Results caveats.....	92
3.3.5	Future directions.....	93
Chapter 4:	Protocol development to target FGFRs using CRISPR/Cas9.....	94
4.1	Introduction.....	94
4.2	Results.....	98
4.2.1	CRISPR/Cas9 system efficiency analysis by <i>tyrosinase</i> targeting.....	98
4.2.2	CRISPR/Cas9 target selection in FGFRs.....	102
4.2.3	Targeting <i>fgfrs</i> using CRISPR/Cas9.....	103
4.2.3.1	<i>fgfr1</i> -targeted embryos.....	106
4.2.3.2	<i>fgfr4</i> -targeted embryos.....	106

4.2.3.3 <i>fgfr1</i> -targeted embryos.....	111
4.2.4 Summary.....	111
4.3 Discussion.....	112
4.3.1 Key findings.....	112
4.3.2 Successful preliminary <i>tyrosinase</i> targeting.....	112
4.3.3 Phenotypic defects of <i>fgfr</i> knockout embryos were observed at a low rate but are consistent with known functions of FGF signalling.....	112
4.3.4 FGFR studies in other model organisms.....	115
4.3.5 Variable targeting efficiencies of sgRNAs in fragment analysis experiments.....	116
4.3.6 Recommendations for further optimisations of CRISPR/Cas9 protocol.....	117
4.3.7 Fragment analysis caveats and recommendations.....	118
4.3.8 Future work.....	120
4.3.9 Conclusions and implications.....	122
References.....	123
Abbreviations.....	131

List of Figures

Chapter 1: Introduction

Figure 1: FGFR structure.....	21
Figure 2: FGF signal transduction overview.....	24
Figure 3: Neural induction in <i>Xenopus</i>	28
Figure 4: Expression levels of <i>X. tropicalis fgfrs</i> throughout early development.....	31
Figure 5: CRISPR/Cas9 genome-modification.....	32

Chapter 3: Identifying well supported FGF target genes through a meta-analysis of high-throughput transcriptomic data sets

Figure 6: Distribution of dnFGFR Affymetrix microarray data.....	49
Figure 7: dnFGFR1 and dnFGFR4 affect similar groups of gene probes.....	52
Figure 8: The enrichment of biological processes associated with down regulated genes by dnFGFR1 and dnFGFR4.....	53
Figure 9: Protein-protein interaction networks for genes down regulated by dnFGFR4.....	54
Figure 10: Ratio of gene expression, as log ₂ values, in iFGFR single replicate Affymetrix microarray data.....	56
Figure 11: iFGFR1 and iFGFR2 affect similar groups of transcripts.....	58
Figure 12: The enrichment of biological processes associated with up regulated genes by iFGFR1.....	60
Figure 13: Protein-protein interaction networks for genes up regulated by iFGFR1, down regulated by iFGFR1 and up regulated by iFGFR2 gastrula stages in whole embryos.....	61
Figure 14: Expression of genes affected by iFGFR1 and iFGFR2 signalling.....	63
Figure 15: Expression of genes affected by dnFGFR4 and iFGFR1, and dnFGFR4 and iFGFR2 signalling.....	65
Figure 16: Expression of genes affected by dnFGFR4, iFGFR1 and iFGFR2.....	66
Figure 17: Distribution of CSKA-FGF4 data.....	68
Figure 18: The enrichment of biological processes associated with up and down regulated genes by FGF4 overexpression in early neurula stage whole embryos.....	69
Figure 19: Protein-protein interaction networks for genes up and down regulated by FGF4 signalling in early neurula stage whole embryos.....	70
Figure 20: Ratio of gene expression, as log ₂ FPKM values, in iFGFR single replicate RNA-Seq data in neuralised animal cap explants.....	72
Figure 21: iFGFR1 and iFGFR4 affect distinct groups of transcripts in neuralised animal cap explants.....	73

Figure 22: The enrichment of biological processes associated with up and down regulated genes by iFGFR1 signalling in neuralised animal cap explants.....	76
Figure 23: The enrichment of biological processes associated with up and down regulated genes by iFGFR4 signalling in neuralised animal cap explants.....	77
Figure 24: Protein-protein interaction networks for genes up and down regulated by iFGFR1 signalling in neuralised animal cap explants.....	79
Figure 25: Protein-protein interaction networks for genes up and down regulated by iFGFR4 signalling in neuralised animal cap explants.....	80
Figure 26: Expression of genes affected by iFGFR1 and iFGFR4 signalling.....	83
Figure 27: Expression of genes affected by FGF4 and iFGFR1, and FGF4 and iFGFR4 signalling.....	84
Figure 28: Expression of genes affected by FGF4, iFGFR1 and iFGFR4 signalling.....	85
Figure 29: The spatial expression pattern of <i>cdx4</i> , <i>en2</i> , <i>myod</i> , <i>n-tubulin</i> and <i>sox3</i> in mid-neurula stage 17-18 <i>X. tropicalis</i> embryos.....	86
Figure 30: The spatial and temporal expression pattern of <i>rasl11b</i> using <i>in situ</i> hybridisation in <i>X. tropicalis</i> embryos.....	87

Chapter 4: Protocol development to target FGFRs using CRISPR/Cas9

Figure 31: Preliminary CRISPR/Cas9 protocol.....	95
Figure 32: Phenotypic classification of embryos subject to <i>tyrosinase</i> gene targeting by CRISPR/Cas9.....	99
Figure 33: Fragment analysis of embryos subject to <i>tyrosinase</i> gene targeting by CRISPR/Cas9.....	100
Figure 34: CRISPR/Cas9 targets in <i>fgfr1</i> , <i>fgfr4</i> and <i>fgfr11</i>	102
Figure 35: Phenotype of <i>X. tropicalis</i> subject to CRISPR/Cas9 targeting of <i>fgfrs</i>	103
Figure 36: Targeting of <i>fgfr1</i> exon 7 by CRISPR/Cas9.....	104
Figure 37: Targeting of <i>fgfr1</i> exon 15 by CRISPR/Cas9.....	105
Figure 38: Targeting of <i>fgfr4</i> exon 3 by CRISPR/Cas9.....	107
Figure 39: Targeting of <i>fgfr4</i> exon 5 by CRISPR/Cas9.....	108
Figure 40: Targeting of <i>fgfr11</i> exon 3 by CRISPR/Cas9.....	109
Figure 41: Targeting of <i>fgfr11</i> exon 5 by CRISPR/Cas9.....	110
Figure 42: Comparison of 1 step and 2 step PCR protocols for fragment analysis.....	121

List of Tables

Chapter 1: Introduction

Table 1: FGF ligand subfamilies.....	15
--------------------------------------	----

Chapter 2: Materials and Methods

Table 2: Primer sequences for sgRNA synthesis.....	32
Table 3: Sequences of primers used in 2 step PCR reactions.....	34
Table 4: Sequences of primers used in 1 step PCR reactions.....	35
Table 5: Linearisation enzymes, buffers and polymerases used in the synthesis of <i>in situ</i> hybridisation probes.....	37

Chapter 3: Identifying well supported FGF target genes through a meta-analysis of high-throughput transcriptomic data sets

Table 6: High and low stringency filtering criteria.....	43
Table 7: Overlap of genes affected by dnFGFR1 and dnFGFR4 signalling.....	46
Table 8: Overlap of genes affected by iFGFR1 and iFGFR2 signalling.....	55
Table 9: Overlap of dnFGFR4 and iFGFR1/2 probe gene lists, which satisfied high stringency filtering criteria.....	58
Table 10: Overlap of transcripts consistently regulated by iFGFR1 and iFGFR4 signalling in neuralised animal cap explants.....	70
Table 11: Overlap of transcripts differentially regulated by iFGFR1 and iFGFR4 signalling in neuralised animal cap explants.....	71
Table 12: Overlap of CSKA-FGF4 and iFGFR1/4 gene lists, which satisfied high stringency filtering criteria.....	78

List of Supplementary Figures

Chapter 3: Identifying well supported FGF target genes through a meta-analysis of high-throughput transcriptomic data sets

Figure S1: Expected size of overlaps modelling dnFGFR microarray data using Python

Figure S2: Expected size of overlaps modelling iFGFR microarray data using Python

Figure S3: Number of gene probes present in both dnFGFR and iFGFR microarray data sets

Figure S4: Expected size of overlaps modelling dnFGFR and iFGFR microarray data using Python

Figure S5: Overlap of differentially expressed genes in embryos subject to decreased FGF signalling by dnFGFR4 and increased by iFGFR1 or iFGFR2, from low stringency filtering

Figure S6: Expected size of overlaps modelling iFGFR RNA-Seq data using Python

Figure S7: Number of genes present in both CSKA-FGF4 and iFGFR1/4 RNA-Seq data sets

Figure S8: Expected size of overlaps modelling CSKA-FGF4 and iFGFR1/4 RNA-Seq data using Python

Figure S9: Overlap of differentially expressed genes in embryos subject to increased FGF signalling by CSKA-FGF4, iFGFR1 or iFGFR4, from low stringency filtering

Chapter 4: Protocol development to target FGFRs using CRISPR/Cas9

Figure S10: Fragment analysis of embryos subject to *fgfr1* exon 7 gene targeting by CRISPR/Cas9 with corresponding phenotypes

Figure S11: Fragment analysis of embryos subject to *fgfr1* exon 15 gene targeting by CRISPR/Cas9 with corresponding phenotypes

Figure S12: Fragment analysis of embryos subject to *fgfr4* exon 3 gene targeting by CRISPR/Cas9 with corresponding phenotypes

Figure S13: Fragment analysis of embryos subject to *fgfr4* exon 5 gene targeting by CRISPR/Cas9 with corresponding phenotypes

Figure S14: Fragment analysis of embryos subject to *fgfr1* exon 3 gene targeting by CRISPR/Cas9 with corresponding phenotypes

Figure S15: Fragment analysis of embryos subject to *fgfr1* exon 5 gene targeting by CRISPR/Cas9 with corresponding phenotypes

List of Supplementary Tables

Chapter 3: Identifying well supported FGF target genes through a meta-analysis of high-throughput transcriptomic data sets

Table S1: Probes up regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR1, when filtered using high stringency filtering criteria

Table S2: Probes down regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR1, when filtered using high stringency filtering criteria

Table S3: Probes up regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR4, when filtered using high stringency filtering criteria

Table S4: Probes down regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR4, when filtered using high stringency filtering criteria

Table S5: Determining the statistical significance of gene probe list overlaps in dnFGFR microarray data using Python

Table S6: Biological processes associated with gene probes down regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR1, when filtered using high stringency filtering criteria

Table S7: Biological processes associated with gene probes up regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR4, when filtered using high stringency filtering criteria

Table S8: Biological processes associated with gene probes down regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR4, when filtered using high stringency filtering criteria

Table S9: Probes up regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR1, when filtered using high stringency criteria

Table S10: Probes down regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR1, when filtered using high stringency criteria

Table S11: Probes up regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR2, when filtered using high stringency criteria

Table S12: Probes down regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR2, when filtered using high stringency criteria

Table S13: Determining the statistical significance of gene probe list overlaps in iFGFR microarray data using Python

Table S14: Biological processes associated with gene probes up regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR1, when filtered using high stringency filtering criteria

Table S15: Determining the statistical significance of gene probe list overlaps between dnFGFR and iFGFR microarray data using Python

Table S16: Probes up regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR4, when filtered using low stringency filtering criteria

Table S17: Probes down regulated in *X. laevis* embryos due to FGF signalling inhibition by dnFGFR4, when filtered using low stringency filtering criteria

Table S18: Probes up regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR1, when filtered using low stringency criteria

Table S19: Probes down regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR1, when filtered using low stringency criteria

Table S20: Probes up regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR2, when filtered using low stringency criteria

Table S21: Probes down regulated in *X. laevis* embryos due to increased FGF signalling through iFGFR2, when filtered using low stringency criteria

Table S22: Overlap of differentially expressed genes in embryos subject to decreased FGF signalling by dnFGFR4 and increased by iFGFR1 or iFGFR2, from low stringency filtering

Table S23: Gene transcripts up regulated in *X. tropicalis* embryos overexpressing FGF4, when filtered using high stringency criteria

Table S24: Gene transcripts down regulated in *X. tropicalis* embryos overexpressing FGF4, when filtered using high stringency criteria

Table S25: Biological processes associated with gene transcripts up regulated in *X. tropicalis* embryos overexpressing FGF4, when filtered using high stringency criteria

Table S26: Biological processes associated with gene transcripts down regulated in *X. tropicalis* embryos overexpressing FGF4, when filtered using high stringency criteria

Table S27: Gene transcripts up regulated in neuralised *X. laevis* animal caps due to increased FGF signalling through iFGFR1, when filtered using high stringency criteria

Table S28: Gene transcripts down regulated in neuralised *X. laevis* animal caps due to increased FGF signalling through iFGFR1, when filtered using high stringency criteria

Table S29: Gene transcripts up regulated in neuralised *X. laevis* animal caps due to increased FGF signalling through iFGFR4, when filtered using high stringency criteria

Table S30: Gene transcripts down regulated in neuralised *X. laevis* animal caps due to increased FGF signalling through iFGFR4, when filtered using high stringency criteria

Table S31: Determining the statistical significance of gene transcript list overlaps in iFGFR RNA-Seq data using Python

Table S32: Biological processes associated with gene transcripts up regulated in neuralised *X. laevis* animal caps with increased FGF signalling through iFGFR1, when filtered using high stringency criteria

Table S33: Biological processes associated with gene transcripts down regulated in neuralised *X. laevis* animal caps with increased FGF signalling through iFGFR1, when filtered using high stringency criteria

Table S34: Biological processes associated with gene transcripts up regulated in neuralised *X. laevis* animal caps with increased FGF signalling through iFGFR4, when filtered using high stringency criteria

Table S35: Biological processes associated with gene transcripts down regulated in neuralised *X. laevis* animal caps with increased FGF signalling through iFGFR4, when filtered using high stringency criteria

Table S36: Determining the statistical significance of gene list overlaps between CSKA-FGF4 and iFGFR RNA-Seq data using Python

Table S37: Gene transcripts up regulated in *X. tropicalis* embryos overexpressing FGF4, when filtered using low stringency criteria

Table S38: Gene transcripts down regulated in *X. tropicalis* embryos overexpressing FGF4, when filtered using low stringency criteria

Table S39: Gene transcripts up regulated in neuralised *X. laevis* animal caps due to increased FGF signalling through iFGFR1, when filtered using low stringency criteria

Table S40: Gene transcripts down regulated in neuralised *X. laevis* animal caps due to increased FGF signalling through iFGFR1, when filtered using low stringency criteria

Table S41: Gene transcripts up regulated in neuralised *X. laevis* animal caps due to increased FGF signalling through iFGFR4, when filtered using low stringency criteria

Table S42: Gene transcripts down regulated in neuralised *X. laevis* animal caps due to increased FGF signalling through iFGFR4, when filtered using low stringency criteria

Table S43: Overlap of differentially expressed genes in embryos subject to increased FGF signalling by CSKA-FGF4, iFGFR1 or iFGFR4, from low stringency filtering

Chapter 4: Protocol development to target FGFRs using CRISPR/Cas9

Table S44: Phenotype of *X. tropicalis* embryos subject to CRISPR/Cas9 targeting, with water-injected controls

Acknowledgements

I would first like to thank my supervisor Dr Harv Isaacs for his unwavering dedication, guidance and patience, throughout my time in his lab but especially during the final few months of my Master's. I couldn't have asked for a more approachable, understanding and kind supervisor, whose door (or Zoom meeting room) was always open. Without Harv's support, I am certain that I wouldn't have been able to complete this thesis and I am eternally grateful for this.

I would also like to thank Dr Betsy Pownall as my honorary co-supervisor, who has always made me laugh, inspired me and had my back. Thank you to my fellow L2 frog lab members Ali, Lewis, Alex and Laura, for their endless advice, support and fun. I would like to acknowledge Professor Paul Genever as my thesis advisory panel member, and I am grateful for his time and constructive feedback. I am indebted to Dr Katherine Newling for her guidance and patience throughout the writing of my bioinformatics chapter during lockdown.

Finally, a huge thank you to my parents, brother, grandparents and Alex. Mum for cooking delicious meals to keep me fuelled and being my motivational gym partner and wellbeing life coach, Dad for our heart to hearts on long countryside bike rides, Elliot for making me laugh, Grandma for turning your spare room into my office and cheering me up with endless cups of tea and my favourite Lotus biscuits, Granda for inspiring, supporting and cheering me on and Alex for spoiling me with my favourite foods and being a huge support even though we have been apart during lockdown. Thank you all for providing me with unconditional love, unfailing support and continuous encouragement throughout my life and through the process of this Master's.

This would not have been possible without you all.

Declaration

I declare that this is an original piece of work conducted under the supervision of Dr Harv Isaacs at the University of York. All data presented is my own work unless referenced otherwise.

CSKA-FGF4 RNA-seq data was collected by Dr Michael King, a previous PhD student in the Isaacs' lab, and processed by members of staff at the University of York Biology Technology Facility. Initial quality control of the RNA-seq data and mapping of sequences to the *Xenopus tropicalis* transcriptome was undertaken by Dr John Davey at the University of York Biology Technology Facility (King 2019). Embryo injections for the iFGFR microarray screen were undertaken by Dr Harv Isaacs and processed by the University of York Biology Technology Facility and subsequently by Dr Harv Isaacs. Embryos for iFGFR RNA-Seq analysis were collected by Dr Hannah Brunsdon, a previous PhD student in the Isaacs' lab, which were then processed by staff at the Centre for Genomic Research at the University of Liverpool. Raw data processing into FPKM values was undertaken by Dr Toby Hodges of the University of York Biology Technology Facility (Brunsdon and Isaacs 2020). Embryo injections for the dnFGFR microarray screen were undertaken by Dr Peter Branney and microarrays were performed in the University of York Biology Technology Facility (Branney et al. 2009).

sgRNA targeting *tyrosinase*, *fgfr1* exon 2, *fgfr4* exon 3 and *fgfr1* exon 3 were designed by Vida Zilinskaite, a previous Master's student in the Isaacs' lab. Fragment analysis was performed by Dr Lesley Gilbert and Dr Sally James at the University of York Biology Technology Facility.

With the exceptions outlined above, work presented in this thesis has not previously been published or submitted for a qualification at this, or any other, University.

Chapter 1: General introduction

1.1 Fibroblast growth factor ligands

1.1.1 Ligand evolution

Fibroblast growth factors (FGFs) were named after their potent mitogenic ability in 3T3 fibroblast cells (Gospodarowicz 1974). Since their discovery, there has been considerable research into the evolution of these polypeptides in humans and in a variety of model organisms, namely frog, mice and zebrafish. *Drosophila melanogaster* and *Caenorhabditis elegans* have three (*branchless*, *pyramus* and *thisbe*) and two (*egl-17* and *let-756*) FGFs respectively. 22 *fgfs* are present in mice and humans, which expanded during evolution in two phases, firstly by tandem gene duplication and secondly by genome duplication (Itoh and Ornitz 2004). 20 *fgfs* have been identified in *Xenopus tropicalis* by comparing sequence homology and synteny to the mammalian orthologues (Lea et al. 2009). The focus of this introduction will be to review the knowledge of FGF function in *Xenopus* development.

1.1.2 Ligand classification

Phylogenetic analysis of the human genome has divided the ligands into 7 subfamilies, the FGF1, FGF4, FGF7, FGF8, FGF9, FGF11 and FGF19 subfamilies (Itoh and Ornitz 2004) (Table 1). Although these subfamilies are highly conserved between humans, mice and *Xenopus* (Itoh and Ornitz 2008; Lea et al. 2009), the genes vary between species. Lea et al. (2009) discovered that in the *X. tropicalis* genome *fgf23* is duplicated, however they were unable to annotate *fgf21* fully and identify orthologues for *fgfr17* and *fgf18*.

These FGF subfamilies can be further categorised depending on their signalling mechanism. Secreted FGFs, paracrine FGFs and hormone-like FGFs (hFGFs), signal through interactions with the FGF receptor (FGFR) family of receptor tyrosine kinases, and intracellular FGFs (iFGFs), which signal in an FGFR-independent manner (Itoh and Ornitz 2004; Itoh and Ornitz 2008). Paracrine FGFs have a conserved 120 amino acid core region and include the ligands in FGF1, FGF4, FGF7, FGF8 and FGF9 subfamilies. These FGFs signal in a paracrine fashion, by binding FGFRs at the receptors' extracellular domains in combination with heparan sulphate proteoglycans (HSPGs), forming a 2:2:2 interaction (Yayon et al. 1991; Schlessinger et al. 2000). HSPGs are required for the initial interaction of a single ligand and receptor and for the dimerisation of two FGF-FGFR complexes (Schlessinger et al. 2000). hFGFs include FGF15, 21 and 23 and signal in an endocrine fashion, with reduced HSPG affinity, instead requiring Klotho as cofactors to bind FGFRs, (Itoh 2010). Conversely, the ligands in FGF11

subfamily are classified as iFGFs and function in neuronal excitability by interacting with intracellular domains of voltage-gated sodium channels (Goldfarb et al. 2007).

1.1.3 Ligand expression

Expression patterns in *X. tropicalis* vary between, and sometimes within, subfamilies, temporally and spatially. For example, the spatial patterns *fgf1* and *fgf2*, of the FGF1 subfamily, are distinct in mesoderm and neural derived tissues, suggesting differing contributions to the development of these tissues. Furthermore, transient expression of *fgf14* in the lens at stage 35 and *fgf13* in the somites at stage 23 suggests ligand expression is under strict regulation (Lea et al. 2009). FGF ligand expression has since been quantitatively analysed at more frequent stages and in different tissues (Owens et al. 2016; Session et al. 2016).

1.1.4 Ligand regulation

Extracellular ligand diffusion is crucial in defining a gradient of FGF ligands, since FGFs function as morphogens by activating different cell types at different concentrations (Slack et al. 1987; Hou et al. 2007; Yu et al. 2009). Consequently, cells acquire positional information and the interaction of FGF ligands and HSPGs provides a key regulatory opportunity in defining this gradient. This can be altered at the level of FGF ligands as a single residue difference in the HSPG-binding pocket of FGF7 and FGF10 underlie their differences in HSPG binding, diffusion and biological activity (Makarenkova et al. 2009). 6-O-endosulfatase, encoded by *Sulf1*, modifies HSPGs extracellularly, and is an endogenous inhibitor of FGF signalling in *Xenopus*, by regulating the morphogen gradient and reducing the responsiveness of cells to FGF signalling. It is required for correct anteroposterior patterning and segmentation of posterior mesoderm into somites (Freeman et al. 2008).

The reversible formation of FGF9 and FGF20 homodimers conceals their receptor-binding sites and promotes their affinity for HSPG binding, restricting ligand diffusion. Together, this limits FGF signalling, creating a crucial autoregulatory mechanism (Kalinina et al. 2009). Further regulation centres around the alternative splicing of FGF ligands within the N-terminus, producing spliceforms of varying lengths and thereby different functions. Of particular interest, different FGF8 spliceforms have distinct potencies in mesoderm induction in *Xenopus*. FGF8b spliceform, which contains an additional 11 amino acids, is necessary for correct mesoderm formation and its overexpression leads to an increase in mesodermal tissue. Conversely, FGF8a spliceform posteriorizes neural tissue, with minimal effect on mesoderm induction, thereby demonstrating distinct roles of different spliceforms of the same gene (Fletcher et al. 2006). This has also been demonstrated in the context of mid-hindbrain patterning as the

Table 1: FGF ligand subfamilies, based on phylogenetic analysis of the human genome, where FGF19 is orthologous to mouse FGF15. Based on information in (Itoh and Ornitz 2004; Zhang et al. 2006).

FGF Subfamily	FGF Ligands	Receptor Preferences	Signalling
FGF1	FGF1	All FGFRs	Paracrine
	FGF2 (bFGF)	FGFR1c, FGFR2c	
FGF4	FGF4 (eFGF)	FGFR1c, FGFR2c	Paracrine
	FGF5		
	FGF6		
FGF7	FGF3	FGFR2b, FGFR1b	Paracrine
	FGF7		
	FGF10		
	FGF22		
FGF8	FGF8	FGFR3c, FGFR4, FGFR1c	Paracrine
	FGF17		
	FGF18		
FGF9	FGF9	FGFR3c, FGFR2c	Paracrine
	FGF16		
	FGF20		
FGF11	FGF11	No known FGFR activation	Intracellular
	FGF12		
	FGF13		
	FGF14		
FGF19	FGF15/19	Weak activation of FGFR1c, FGFR2c	Endocrine
	FGF21		
	FGF23		

result of altered receptor binding affinities due to an additional contact site in the additional amino acids in FGF8b (Olsen et al. 2006). This FGF ligand splicing therefore increases the functional diversity of FGF signalling, as seen in FGFR alternative splicing.

1.2 Fibroblast growth factor receptors

1.2.1 Receptor evolution

FGFRs are cell-surface receptor tyrosine kinases, which are activated by FGF ligand binding extracellularly. Two large scale genome duplications, which coincides with the FGF ligand gene family expansion, resulted in the expansion of the one invertebrate *fgfr* gene to the multiple present in vertebrates today. For example, there is one *fgfr* gene present in *C. elegans* (*egl-15*), two in *D. melanogaster* (*breathless*, *heartless*) and four canonical *fgfr* genes in humans, mice (Itoh and Ornitz 2004) and frog (*fgfr1-4*) (Lea et al. 2009).

1.2.2 Receptor structure

FGFRs contain a signal peptide, three immunoglobulin-like (Ig) loops and an acidic box in the extracellular region (Figure 1A).

FGFR1-3 are alternatively spliced in their Ig III loop generating IIIb and IIIc isoforms, containing the IIIa exon as the N-terminal half and IIIb or IIIc as the C-terminal half (Johnson et al. 1991). FGF ligands interact with this Ig III domain, along with Ig II, and the linker region between Ig II and Ig III in FGFRs (Goetz and Mohammadi 2013). Therefore, this alternative splicing provides a range of receptor isoforms, which have varying binding affinities for different FGF ligands (Zhang et al. 2006). FGFRb and FGFRc isoforms are restricted to epithelial and mesenchymal tissues respectively (Orr-Urtreger et al. 1993). However, FGFR4 lacks this alternative exon and therefore is not alternatively spliced, reviewed in (Holzmann et al. 2012).

FGFR1 and FGFR2 Ig I can also be alternatively spliced producing long and short transcripts, which include or exclude the exon encoding Ig I. Although FGFR1 and FGFR2 both exhibit splicing in this region, the resulting transcripts show different expressions temporally. Long *fgfr1* is present at higher levels early in development, however short *fgfr1* is expressed more later, whereas long *fgfr2* is the predominant isoform throughout stages up until stage 40 (Lea et al. 2009).

FGFR1 can undergo additional splicing in the juxtamembrane region, resulting in Val(423)-Thr(424) amino acid deletion and therefore production of VT- splice variant. *In vitro* VT- cannot be phosphorylated, and therefore regulated, by protein kinase C (PKC), due to loss of the Thr phosphorylation site (Gillespie et al. 1995).

Intracellularly, these canonical receptors contain a split tyrosine kinase domain, allowing the activation of various signalling pathways. This domain has been evolutionarily conserved, showing its importance functionally, as invertebrate homologues share 60% amino acid

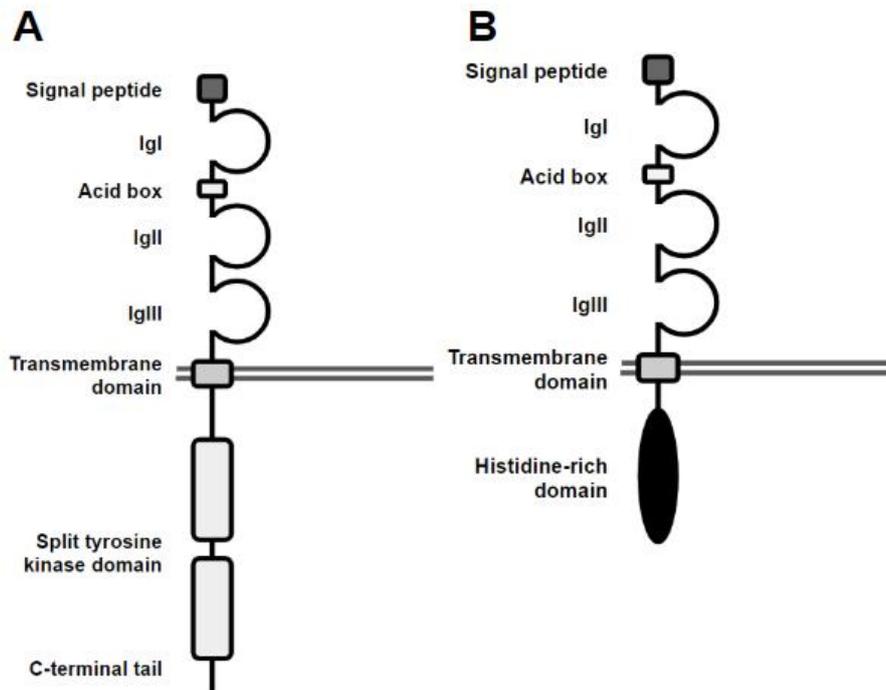


Figure 1: FGFR structure. Adapted from Zakrzewska et al. (2008). **(A)** FGFR1-4 structure containing extracellular regions (signal peptide, 3 Ig-like loops and acidic box), transmembrane region and intracellular region (split tyrosine kinase domain and C-terminal tail). **(B)** FGFR5 differs intracellularly through having a histidine-rich domain, rather than the split tyrosine kinase domain present in FGFR1-4.

identity with vertebrate FGFRs, which is twice as much as the sequence identity extracellularly (Itoh and Ornitz 2004).

1.2.3 Receptor expression

In *Xenopus*, *fgfr1*, *fgfr3* and *fgfr4* are expressed maternally, whereas only *fgfr3* is not expressed during gastrulation (Lea et al. 2009). However, this analysis did not distinguish between IIIb and IIIc isoforms, which is known to differ in different cell types, namely epithelial and mesenchymal cells respectively (Orr-Urtreger et al. 1993).

1.2.4 FGFRL1

A fifth member of the FGFR gene family was identified in mice and is referred to as FGFR like 1 (FGFRL1) or FGFR5 (Figure 1B) (Sleeman et al. 2001). It likely arose from an ancestral gene which duplicated during early metazoan evolution to give rise to FGFRL1 and FGFRs (Bertrand et al. 2009). This novel receptor differs from the canonical FGFRs due to its lack of the intracellular tyrosine kinase domain and it was therefore predicted to inhibit FGF signalling (Sleeman et al. 2001). The exon contributions and genomic synteny, for example adjacent to FGF8/17/18 and/or FGFR, is similar in metazoans. Interestingly, in the *Nematostella vectensis* genome *fgfr1* is located next to the Sprouty homologue *NvSprouty* (Bertrand et al. 2009), a negative regulator of FGF signalling (Sivak et al. 2005), suggesting it could be co-expressed.

In *Xenopus*, *fgfr1* is first expressed anteriorly in late gastrula embryos and subsequently detected in many mesodermal and neural derivatives (Hayashi et al. 2004). Similar to canonical FGFRs, FGFR1 is localised to the plasma membrane, interacts with heparin (Trueb et al. 2003) and exhibits ligand preferences (Sleeman et al. 2001; Trueb et al. 2003; Steinberg et al. 2010). In contrast to canonical FGFRs, FGFR1 does not bind to all FGF ligands from a subfamily. Its extracellular domain can be shed from cells, due to cleavage at the membrane proximal region by an unknown protease. This could provide key FGF signalling regulation, by binding and sequestering FGF ligands, thus preventing signalling through other FGFRs (Steinberg et al. 2010).

The phenotype of *Xenopus* embryos overexpressing *fgfr1* resembles that arising from blocking FGF signalling with dominant negative FGFR1 (dnFGFR1), which supports the idea that FGFR1 functions as a negative regulator of FGF signalling (Amaya et al. 1991; Steinberg et al. 2010). This can be partially rescued by *fgfr1* mRNA injection, suggesting FGFR1 and FGFR1 have overlapping ligand specificities (Steinberg et al. 2010).

1.3 Fibroblast growth factor signal transduction

Interaction of FGFR1-4 with FGF ligands and HSPG leads to receptor dimerisation and intracellular autophosphorylation on tyrosine residues. This activates three main intracellular signal transduction pathways (Figure 2): the phosphoinositide-3 kinase (PI3K), phospholipase C gamma (PLC γ) and mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) pathways, reviewed in (Böttcher and Niehrs 2005).

1.3.1 Phosphoinositide-3 kinase

PI3K pathway activation begins with the interaction of FGFR substrate 2 alpha (FRS2 α) with Gab1 via growth factor receptor bound protein (Grb2), leading to the recruitment and activation of PI3K by phosphorylated Gab1 (Ong et al. 2001). Activation of the downstream mediator, and proto-oncogene, serine/threonine kinase Akt/protein kinase B (AKT/PKB) affects cell survival and proliferation (Nicholson and Anderson 2002).

1.3.2 Phospholipase C gamma

PLC γ binds to FGFRs at the auto phosphorylated Tyr 766 (Mohammadi et al. 1991), leading to its activation and phosphatidylinositol-4,5-diphosphate (PIP $_2$) hydrolysis into inositol-1,4,5-trisphosphate (IP $_3$), which stimulates intracellular calcium release, and diacylglycerol (DAG), which activates PKC. This pathway has implications for cell migration and morphology (Sivak et al. 2005).

1.3.3 Mitogen-activated protein kinase

The lipid-anchored FRS2 α constitutively interacts with FGFRs via its phosphotyrosine-binding (PTB) domain in the juxtamembrane region (Ong et al. 2000) and the adaptor protein Grb2 (Kouhara et al. 1997). Ligand binding and therefore receptor activation leads to the binding of Grb2 to the guanine nucleotide exchange factor son of sevenless (SOS), via its Src homology 3 (SH3) domain (Kouhara et al. 1997). SOS activates Ras by promoting its dissociation from guanosine diphosphate (GDP) and association with guanosine triphosphate (GTP), which leads to a cascade of phosphorylation involving Raf, Mek and MAPK, also known as ERK (reviewed in (Tsang and Dawid 2004)). Phosphorylated, and therefore activated, serine/threonine kinase MAPK phosphorylates and activates transcription factors, thus altering cell gene expression (Raible and Brand 2001; Murphy et al. 2002; Neugebauer et al. 2009).

1.3.4 Transcriptional regulation by FGF signal transduction

A key family of transcription factors are the ETS transcription factors, which includes ETS related molecule (Erm) and Polyoma enhancer activator 3 (Pea3). Ectopic expression of these transcription factors in zebrafish embryos resulted from *fgf3* mRNA injection or implantation of FGF8-coated beads, showing direct activation by FGF signalling (Raible and Brand 2001). Conversely, *fgfr1* zebrafish morphant embryos showed reduced expression of *erm* and *pea3* (Neugebauer et al. 2009).

FGF signalling also increases the activity of activator protein 1 (AP-1) heterodimers, composed of Jun and Fos transcription factors, during mesoderm induction. Ectopic overexpression of c-jun and c-fos posteriorized *Xenopus* embryos (Kim et al. 1998), reminiscent of FGF4 overexpression (Isaacs et al. 1994), providing evidence that AP-1 is a downstream effector of FGF signalling.

1.3.5 Regulation of FGF signal transduction

Extracellular, transmembrane and intracellular components of the FGF pathway can be targeted for regulation.

The expression of transmembrane proteins FLRTs and Sef form part of FGF8-synexpression group and are induced by FGF signalling, creating positive and negative feedback loops respectively (Tsang et al. 2002; Böttcher et al. 2004; Kovalenko et al. 2006). The transmembrane protein Sef attenuates FGF signalling in a negative feedback loop

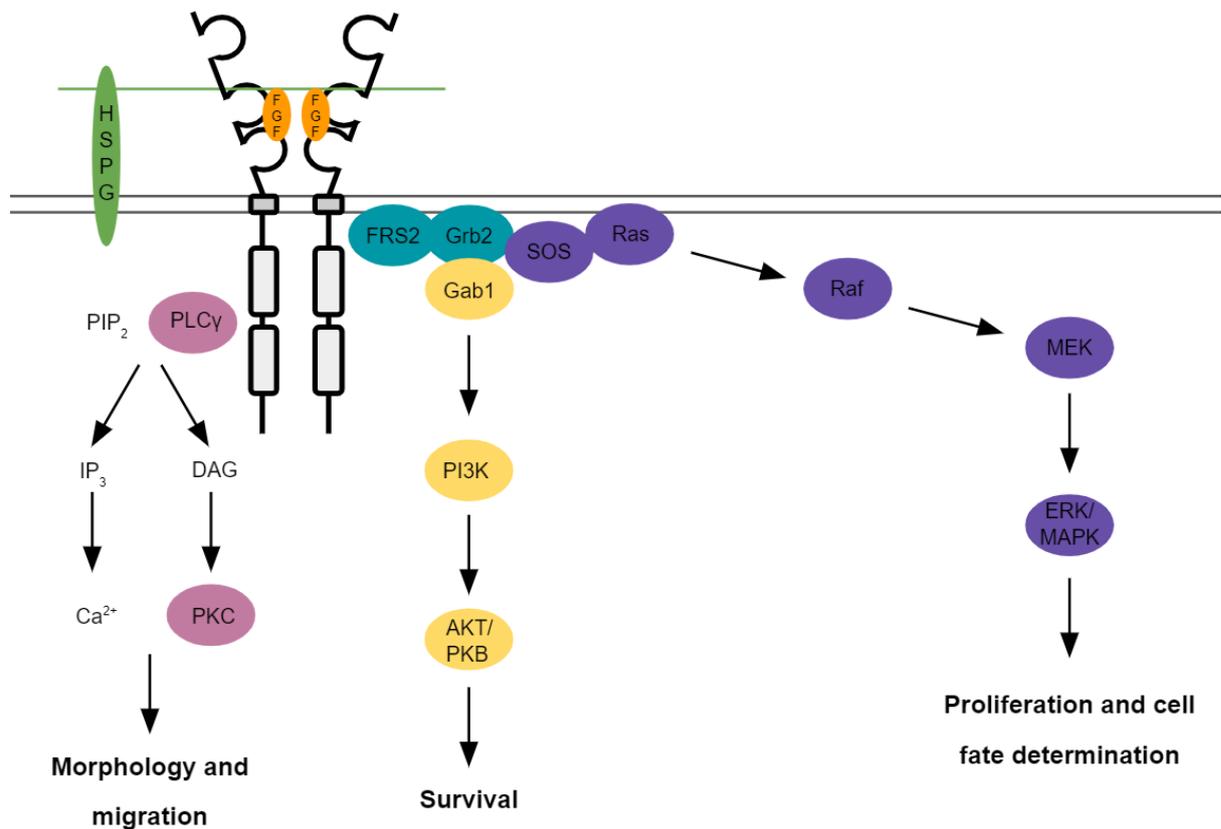


Figure 2: FGF signal transduction overview. FGF signalling begins with FGFs binding to FGFRs and heparan sulphate proteoglycans (HSPGs), in a 2:2:2 interaction. This leads to the ligand dependent dimerisation of the FGFRs and the tyrosine autophosphorylation by the receptors' intracellular tyrosine kinase domain. Various signalling proteins can now bind and activate different signal transduction pathways. Proteins which connect pathways are in blue. In the MAPK pathway (purple), FGFR kinase domains phosphorylate FRS2, which recruits Grb2 (an adaptor protein). The nucleotide exchange factor son of sevenless (SOS) can now be recruited and exchange guanosine diphosphate (GDP) for guanosine triphosphate (GTP) and activate the Ras GTPase, activating the MAPK pathway. Cell proliferation and differentiation are altered as a result. In the phosphoinositide-3 kinase (PI3K) pathway (yellow), Grb2 recruits Gab1 (an adaptor protein), which activates PI3K and thus serine/threonine kinase Akt/protein kinase B (AKT/PKB), which has implications for cell survival. The phospholipase C gamma (PLC γ) pathway (pink) begins with its recruitment by the FGFR kinase domain, which hydrolyses PIP₂ to IP₃, inducing calcium release from intracellular stores, and DAG, activating protein kinase C (PKC). This pathway alters the morphology and migration of cells. Figure adapted from (Böttcher and Niehrs 2005; Dorey and Amaya 2010).

(Tsang et al. 2002), whereby Sef interacts with the extracellular, transmembrane (Kovalenko et al. 2006) and intracellular domain of FGFRs (Tsang et al. 2002; Kovalenko et al. 2006). Sef prevents FGFR1 tyrosine phosphorylation, resulting in the inhibition of MAPK signalling and affecting cell survival (Kovalenko et al. 2006). Fibronectin leucine rich transmembrane protein 3 (FLRT3) also interacts with FGFRs, however it functions to increase MAPK signalling. Its overexpression in *Xenopus* embryos induces ectopic *fgf8* and *Brachyury* (*tbxt*) expression, along with an ectopic tail containing muscle actin and a notochord (Böttcher et al. 2004).

Sproutys, Spreds, MAP kinase phosphatase 1 (MKP1), MAP kinase phosphatase 3 (MKP3)

and dual specificity phosphatase 5 (DUSP5) are key intracellular FGF signal transduction pathway modulators. They regulate the relative signalling between pathways MAPK, PLC γ and PI3K, allowing different signal interpretation in the responding cell. Sproutys and Spreds have opposite effects, where Sproutys inhibit PLC γ and Spreds inhibit MAPK signalling (Sivak et al. 2005). *spry2*, *mkp1*, *mkp3* and *dusp5* were downregulated by dnFGFRs and therefore activated by FGF signalling. MKP1, MKP3 and DUSP5 are also negative regulators of FGF-mediated MAP kinase signalling, by inhibiting FGF-induced ERK phosphorylation (Lewis et al. 1995; Umbhauer et al. 1995; Branney et al. 2009).

1.4 Patterning the early *Xenopus* embryo

1.4.1 Mesoderm induction

Mesoderm induction is the process where mesoderm arises in the equatorial region of blastula stage embryos in response to signals produced by vegetal cells. *Vg1* is a vegetally localised maternal RNA of the transforming growth factor beta (TGF β) signalling family (Rebagliati et al. 1985; Weeks and Melton 1987). However, unlike other TGF β s and contrary to expectations, *Vg1* does not form dimers, is not processed or secreted, and synthetic mRNA injection does not induce mesoderm in animal caps (Dale et al. 1993). However, evidence for an involvement of *Vg1* in mesoderm induction included the discovery *Vg1*-depleted embryos exhibiting delayed gastrulation and reduction in organiser (dorsal mesoderm) gene expression (Birsoy et al. 2006).

Isaacs et al. (1994) proposed a model which involved *Vg1* requiring maternal FGF in responding tissues to activate the T-box transcription factor *tbxt*, which then activates zygotic *fgf4* expression. This forms an autocatalytic loop during blastula stage. By gastrula stage, *fgf4* is still required to maintain *tbxt*, although *fgf4* expression is now independent of *tbxt* and FGF signalling functions in mesoderm maintenance (Isaacs et al. 1994; Schulte-Merker and Smith 1995). This model is supported by the commencement of zygotic FGF signalling at mid-blastula stage 8.5 (Branney et al. 2009) and the detection of FGF mRNA in the marginal zone, not in the vegetal cells, of stage 10.5 early gastrula *X. tropicalis* embryos (Lea et al. 2009). Furthermore, inhibiting endogenous FGF signalling using dnFGFRs blocked mesoderm formation, resulting in the loss of most trunk and tail mesoderm, and revealed defects in gastrulation (Amaya et al. 1991; Amaya et al. 1993; Isaacs et al. 1994).

Transcriptomic analysis of FGF-regulated genes during mesoderm induction in early gastrula embryos identified novel FGF target genes, *dusp5* and *mkp1*. These genes are negative

regulators of FGF-mediated MAPK signalling, by inhibiting FGF-induced ERK phosphorylation and FGF-mediated mesodermal tissue formation. This negative feedback loop restricts FGF signalling in *Xenopus* embryos (Branney et al. 2009).

Regulation of FGF signalling allows different interpretations in responding cells and therefore coordinates different processes in gastrulation, namely Sprouty and Spred proteins control the switch from mesoderm induction to cell movements respectively. The inhibition of Ca^{2+} and PKC signalling by Sprouty proteins enables mesoderm induction, followed by morphogenesis, resulting from MAPK activation inhibition by Spreds (Sivak et al. 2005).

Contributions to mesoderm induction vary from different FGFs, even different spliceforms, with FGF8b being the predominant spliceform during mesoderm induction and can induce mesoderm more robustly than FGF8a (Fletcher et al. 2006). Furthermore, novel contributions of Pinhead (Pnhd) and R-spondin 2 (Rspo2) secreted proteins to mesoderm induction via FGF signalling have recently been identified, acting as positive and negative regulators respectively (Ossipova et al. 2020; Reis and Sokol 2020).

1.4.2 Dorsal-ventral patterning

FGF signalling promotes dorsal and inhibits ventral mesoderm specification (Lee et al. 2011), which is consistent with its expression and FGF-dependent ERK activation dorsally (Branney et al. 2009). Bone morphogenetic protein (BMP) signalling promotes ventral cell identity and zebrafish studies have shown that FGF signalling restricts BMP gene expression ventrally, independent of BMP antagonists (Fürthauer et al. 2004).

FGF signalling further inhibits BMP signalling by activating these secreted BMP antagonists (Branney et al. 2009; Lee et al. 2011), which are expressed in Spemann's organiser at the dorsal blastopore lip (Khokha et al. 2005) and block BMP signalling by binding extracellular BMP ligands and preventing their interactions with receptors (Zimmerman et al. 1996). Drug-inhibition of FGF signalling revealed that FGFs act in the dorsal marginal zone to induce dorsal mesoderm, indicated by the positive regulation of dorsally expressed genes *chordin* and *noggin* (Lee et al. 2011). This is consistent with the significant decrease in *chordin* and *noggin* expression when FGF signalling was inhibited by dnFGFR constructs (Branney et al. 2009).

Analysis of lineage-specific markers *myogenic differentiation (myod)* and *stem cell leukemia (sc)*, both encoding basic helix-loop-helix (bHLH) transcription factors, similarly revealed that FGF4 promotes skeletal muscle from dorsal mesoderm and restricts blood development from ventral mesoderm respectively during late blastula to gastrula stages (Isaacs et al. 2007).

FGF4 is necessary for initial *myod* transcriptional activation (Fisher et al. 2002). FGF signalling leads to the phosphorylation and activation of Ets transcription factor Elk1, which forms a complex with serum response factor (SRF) to bind *early growth response 1 (egr1)* promoter (Nentwich et al. 2009). *egr1* was downregulated by FGF signalling inhibition (Branney et al. 2009) and functions as a direct activator of *myod*, specifying these cells to become muscle (Nentwich et al. 2009). This is consistent with the observation of reduced *myod* expression (Isaacs et al. 1994) and muscle formation in dnFGFR embryos (Amaya et al. 1991).

FGF signalling inhibits the ventral-specific gene *gata2* and the identity of mesoderm is transformed from dorsal to ventral when FGF signalling is inhibited (Lee et al. 2011). *scl* is regulated by GATA2 (Chan et al. 2007), which stimulates blood development and is inhibited by FGF and induced by BMP4. Furthermore, FGF activates BMP4 downstream effector *PV.1*, which inhibits blood development (Xu et al. 1999).

1.4.3 Neural induction

The balance of BMP and FGF signalling is also crucial for neural induction. Gastrula ectoderm can either differentiate into epidermal or neural cells, and this decision involves the activation or inhibition of BMP signalling respectively (Delaune et al. 2005). BMP receptor (BMPR) serine/threonine kinases phosphorylate the C-terminus of the transcription factor Smad1, which promotes nuclear translocation and transcriptional activity. This leads to epidermis differentiation in ventral ectoderm (reviewed by (Massagué and Chen 2000)). Secreted BMP antagonists, such as Follistatin, Chordin and Noggin, are expressed in Spemann's organiser at the dorsal blastopore lip (Khokha et al. 2005) and block BMP signalling by binding extracellular BMP ligands and preventing their interactions with receptors (Zimmerman et al. 1996), allowing dorsal ectoderm to develop into neural tissue (reviewed by (Massagué and Chen 2000)).

Triple knockdown of BMP antagonists *follistatin*, *chordin* and *noggin* using antisense morpholino oligonucleotides (MOs) resulted in complete loss of the neural plate in whole embryos, demonstrating the requirement of BMP inhibition in neural induction (Khokha et al. 2005). FGF signalling functions in this process of neural induction by positively regulating the expression of BMP antagonists in the organiser (Delaune et al. 2005; Fletcher et al. 2006; Branney et al. 2009).

FGF signalling is required for neural induction since the inhibition of FGFR4 signalling by dnFGFR4 resulted in the loss of early neural marker *sox2* and differentiation marker *ncam* expression, which was rescued by the intact *fgfr4* overexpression (Delaune et al. 2005).

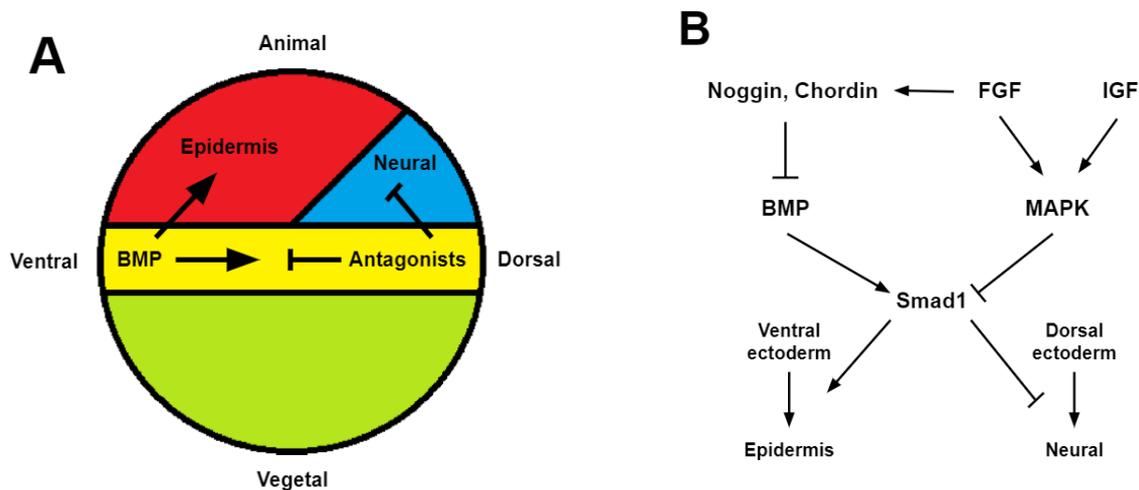


Figure 3: Neural induction in *Xenopus*. (A) The concentration of BMP is high in the ventral mesoderm (yellow), which acts on ventral ectoderm to promote epidermis formation (red). Secreted BMP antagonists, produced in the dorsal mesoderm (yellow), promote neural development in the dorsal ectoderm (blue), by inhibiting BMP signalling. (B) FGF signalling promotes the expression of BMP antagonists Noggin and Chordin, which inhibit BMP signalling. FGF and IGF activate the MAPK pathway, which inhibits Smad1 activation. In contrast BMP promotes Smad1 phosphorylation and activation, leading to epidermis formation in the ventral ectoderm, whereas inhibition of Smad1 leads to neural formation in the dorsal ectoderm. Figure adapted from Delaune et al. (2005).

BMP and FGF signalling converge at Smad1. Neural induction requires the inhibition of Smad1 via phosphorylation of its linker region (Pera et al. 2003; Kuroda et al. 2005), which is performed by FGF-activated diphosphorylated ERK (dpERK) at MAPK sites (Kuroda et al. 2005). Smurf1 recognises this inhibitory phosphorylation, which leads to either its polyubiquitination and degradation or cytoplasmic retention by preventing interaction between Smad1 and nuclear translocation factor Nup214 (Sapkota et al. 2007). This inhibitory phosphorylation is necessary for neural fate determination and can also be induced from insulin-like growth factor (IGF)-mediated MAPK activation (Pera et al. 2003).

FGF signalling is required for correct neural patterning. Treatment of *Xenopus* ectoderm explants with FGF2 induced posterior neural marker HoxB9 (Lamb and Harland 1995). Anterior neural differentiation is the result of differential Smad1 phosphorylation by MAPK and BMP (Kuroda et al. 2005). However, the ability of FGF to induce anterior neural tissue increases with increasing ectoderm age and with BMP signalling inhibition via noggin (Lamb and Harland 1995). Triple knockdown of BMP antagonists embryo phenotypes showed expansion of posterior and ventral tissues (Khokha et al. 2005). This is in accordance with the ability of Noggin to induce anterior neural tissue. FGF2 and noggin together induce a more complete pattern of markers, namely *otx2* (anterior) and *hoxb9* (posterior). *en2* (mid-hindbrain

junction) and *krox20* (third and fifth hindbrain rhombomere) were also detected, which were not induced by either molecule alone (Lamb and Harland 1995).

1.4.4 Anteroposterior patterning

Early evidence of a role for FGF signalling in anteroposterior patterning came from the observation that FGF4 is expressed posteriorly at stage 14 (Isaacs et al. 1994) and dnFGFR-mediated FGF signalling inhibition induced trunk and posterior defects, which were rescued with wild type *fgfr* overexpression (Amaya et al. 1991). Overexpression of FGF4 during gastrula resulted in embryos exhibiting diminished heads and enlarged proctodaeum. This posteriorised phenotype is the opposite of the posterior truncations observed in dnFGFR-mediated FGF signalling inhibition (Isaacs et al. 1994).

Increased expression of posteriorly expressed genes *hoxa7* and *hoxb9* was observed in FGF4-overexpressing embryos (Isaacs et al. 1994; Pownall et al. 1996) and the ectopic activation of these genes results in anterior truncations. FGF activates caudal type homeobox (Cdx) genes, namely *Xcad3* (also known as *Cdx4*). *Cdx4* regulates Hox genes and FGF regulates these genes during gastrula and neurula after mesoderm induction (Pownall et al. 1996; Faas and Isaacs 2009), leading to correct posterior neural patterning (Isaacs et al. 1994; Lamb and Harland 1995; Pownall et al. 1996; Faas and Isaacs 2009).

1.4.5 Left-right asymmetry

Internal organ asymmetry is dictated by asymmetric Nodal signalling induction in the left lateral plate mesoderm, and a ciliated epithelium left-right organiser (LRO), termed the gastrocoel roof plate (GRP) in *Xenopus*, which creates a directional fluid flow. The GRP is essential for correct left-right development, shown by incorrect organ positioning in GRP-ablated embryos (Blum et al. 2009).

FGF regulates the expression of Nodal, as the zebrafish homologue of Nodal *Southpaw* exhibits bilateral expression in *fgfr1* morphant embryos and when *fgfr1* is depleted in the Kupffer's vesicle (KV), homologous to the GRP. FGF signalling inhibition in *Xenopus* embryos revealed shorter cilia in the GRP than control embryos. Zebrafish embryos similarly displayed reduced KV cilia length, which is the result of the downregulation of ciliogenesis transcription factors. Cilia length in *dnFGFR1* mRNA injected zebrafish embryos was comparable to that observed when *fgf8* and *fgf24* were inhibited simultaneously, but not alone, suggesting these ligands function redundantly to signal through FGFR1 to regulate KV cilia length. FGF signalling manipulation in zebrafish and *Xenopus* embryos revealed a conserved role for FGF

in regulating ciliogenesis in the GRP, KV and other ciliated structures (Neugebauer et al. 2009).

FGFR4 signalling is also implicated in left-right asymmetry, as inverted cardiac and gut looping is observed in *fgfr4* knockout *Xenopus* embryos (Sempou et al. 2018). However, unlike *fgfr1*, which is expressed in the KV (Neugebauer et al. 2009), *fgfr4* is not detected in the GRP. Instead, FGFR4 functions earlier, to pattern the paraxial mesoderm in gastrula stages, which contributes to the lateral GRP domain (Sempou et al. 2018).

1.5 Summary

There is a good understanding of the developmental processes and target genes regulated by FGF signalling during early *Xenopus* development. However, several different members of the FGFR family are expressed in early development. For example the temporal expression of *fgfr1* and *fgfr4* overlap maternally and during the period of mesoderm induction, gastrulation and neural induction (Golub et al. 2000; Lea et al. 2009) (Figure 4). It is not clear the roles of individual FGFRs in mediating the effects of FGF signalling in *Xenopus* development. For this reason, previous work has aimed to discriminate between the genes activated by FGFR1 and FGFR4 signalling by overexpressing dnFGFR1 and dnFGFR4 (Delaune et al. 2005; Branney et al. 2009). Using this approach, Branney et al. (2009) found largely the same gene expression profiles in high-throughput transcriptomic analysis when FGFR1 and FGFR4 were inhibited using their respective dnFGFRs. However, there is concern that dnFGFR overexpression could lead to the construct forming non-productive dimers with and inhibiting other FGFRs than its specific FGFR (Ueno et al. 1992). This could explain why no difference was observed in the gene expression profiles regulated by FGFR1 and FGFR4, when inhibited through their respective dnFGFR. Conversely, the use of inducible FGFR1 and FGFR4 (iFGFR) to increase FGF signalling suggests FGFR1 and FGFR4 activate different gene expression profiles. This could be explained by the different abilities of FGFR1 and FGFR4 to activate ERK, where FGFR1 is a stronger activator of ERK than FGFR4. This suggests FGFR1 and FGFR4 regulate distinct processes in *Xenopus* development, by differentially activating downstream signal transduction pathways (Brunsdon and Isaacs 2020), despite their similar temporal expression pattern (Figure 4).

fgfr1 is first expressed anteriorly in last gastrula embryos (Hayashi et al. 2004) and subsequently mimics the expression patterns of *fgfr1* and *fgfr4* (Figure 4). This is interesting since FGFR1 and FGFR4 are thought to have overlapping ligand specificities (Steinberg et al. 2010). *fgfr1*, *fgfr4* and *fgfr1* transcripts are all present from neural induction and they could

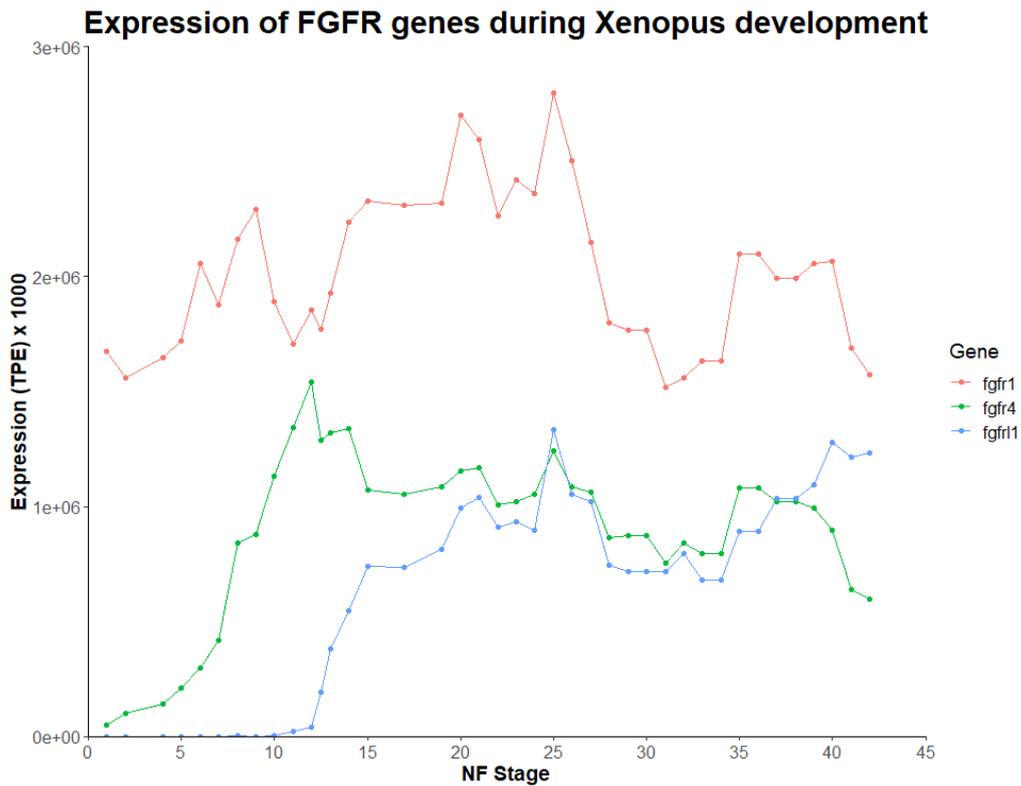


Figure 4: Expression levels of *X. tropicalis* fgfrs throughout early development. Transcripts measured in transcripts per embryo (TPE) x 1000. Data from (Owens et al. 2016).

be involved in similar processes. Therefore, this project aims to determine the effects on *Xenopus* development from inhibiting *fgfr1* and *fgfr4*. *fgfr11* will also be inhibited to investigate its putative role as a negative regulator of FGF signalling.

1.6 CRISPR/Cas9

The efficiency of the widely used MO approach, to knockdown a gene transiently, has been called into question recently. It has been suggested that MOs can induce many off-target splicing events and an innate immune response, which was not observed in transcription activator-like effector nucleases (TALEN) knockout embryos, despite morphology and mesoderm lineage markers being indistinguishable between the *tbxt* mutants and morphants. It was suggested that optimised conditions, including dose and GC content of the MO, could reduce but not eliminate these effects (Gentsch et al. 2018). These conclusions have been challenged, with evidence instead pointing to these effects being confined to the specific *tbxt* and *tbxt2* MOs used, as *foxh1* (60% GC content) and *gsc* (56%) MOs didn't induce innate immune response genes during gastrula. Furthermore, the effect of MOs on immune response genes at later stages is unclear as data sets for early stages, for example gastrula and neurula, are more abundant and therefore have been analysed here (Paraiso et al. 2019). However,

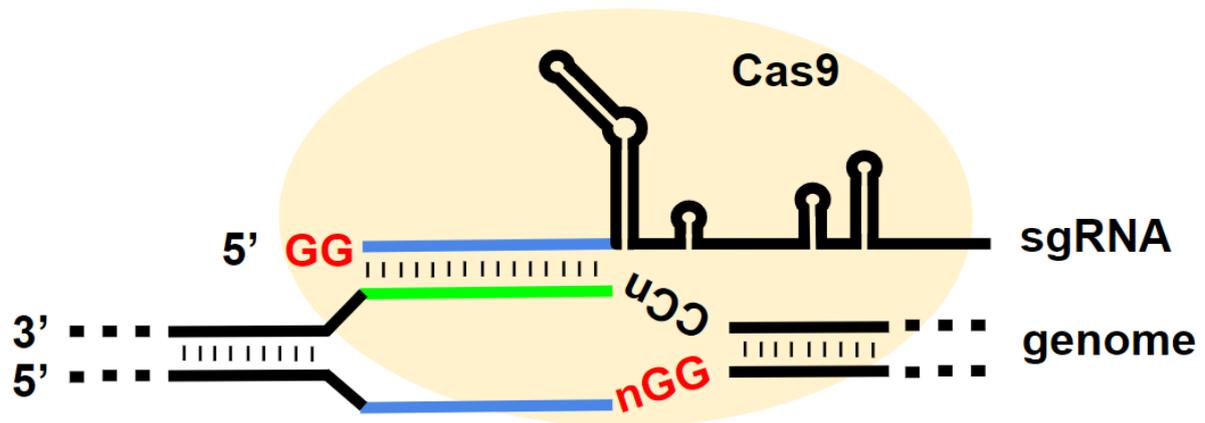


Figure 5: CRISPR/Cas9 genome-modification. sgRNA is coinjected with endonuclease Cas9 protein into 1-2 cell stage embryos, whereby the complex is recruited to the target site, via complementary base pairing, upstream of the PAM sequence. Cas9 cleaves the target site. Figure adapted from (Nakayama et al. 2013).

there was an increase in interferon-stimulated genes in MO-injected zebrafish embryos during segmentation stages (Lai et al. 2019), which is akin to *Xenopus* tailbud stages. This is particularly pertinent to this investigation, since embryo observation at late tailbud stages is crucial to determine the individual roles of FGFRs in laterality.

Together with the transient gene knockdown by MOs, without genetic manipulation, CRISPR/Cas9-mediated gene editing (henceforth referred to as CRISPR/Cas9) is an attractive system to selectively knockout *fgfrs* and investigate the developmental effects of inhibiting *fgfr1*, *fgfr4* and *fgfr11*. CRISPR/Cas9 has been chosen for genome modification as it has been successfully implemented in the widely used and researched model organism of choice, *X. tropicalis* (Blitz et al. 2013; Nakayama et al. 2013; Bhattacharya et al. 2015; McQueen and Pownall 2017), with its diploidy, fast generation time and sequenced genome, enabling F0 embryo analysis.

In order to perform genome editing using CRISPR, the single guide RNA (sgRNA) and the Cas9 endonuclease are coinjected into *X. tropicalis* embryos. The sgRNA forms a complex with and recruits the Cas9, to a ~20 bp target DNA sequence, which is complementary to the sgRNA and followed by a 5'-NGG-3' protospacer adjacent motif (PAM) site (where N is any nucleotide) in the genome for recognition. Cas9 cleaves DNA, resulting in nonhomologous end-joining (NHEJ), which can lead to insertions or deletions of bases (indels) (Figure 5). NHEJ is error-prone and frequently leads to mosaicism (Blitz et al. 2013; Nakayama et al. 2013; Bhattacharya et al. 2015; McQueen and Pownall 2017), which is the presence of more than two alleles in an individual and therefore reduces the generation of a knockout in

one step without the need of F0 founder breeding (Lamas-Toranzo et al. 2019), which is impossible in the timescale of this project, and reduces the penetrance of the expected phenotype. For these reasons, it is crucial to accurately determine the efficiency of CRISPR/Cas9 in inducing target gene mutations. To do this, indel frequency will be analysed by fragment analysis to determine CRISPR/Cas9 targeting efficiency. This process relies on the generation of fluorescently labelled DNA fragments from genomic DNA from embryos injected with CRISPR/Cas9 constructs, which can be separated at single base resolution according to size (Schuelke 2000; Yang et al. 2015).

The development of this genome-modification protocol will enable gene expression analysis of FGF target genes identified by a meta-analysis of high-throughput transcriptomic data sets. This will reveal the unique roles of individual FGFRs in regulating the expression of FGF target genes.

1.7 Project aims

The main hypothesis for this investigation was that FGFR1, FGFR4 and FGFR1 have distinct roles in patterning *X. tropicalis* embryos and regulating FGF target gene expression. The overall aims of this project to test this hypothesis were:

- To identify putative FGF target genes from a meta-analysis of high-throughput transcriptomic RNA-Seq and microarray data sets
- To develop a CRISPR/Cas9 protocol for F0 analysis of knockout embryos
- To investigate the unique roles of FGFR1, FGFR4 and FGFR1 in mediating FGF signalling, using CRISPR/Cas9 to analyse FGF target gene expression.

Chapter 2: Materials and Methods

2.1 Embryological methods

2.1.1 *X. tropicalis* embryo *in vitro* fertilisations

X. tropicalis females were primed by subcutaneous injection of 10 units of human chorionic gonadotropin (hCG) 24 hours prior to the time of required laying. On the day of laying, females were injected with 100 units of hCG. Males were sedated using 10% Benzocaine 70% Ethanol, culled and testes placed in L15 + 10% foetal calf serum (FCS). Eggs were transferred to 55mm plates coated with L15 + 10% FCS to be fertilised by testes, crushed in L15 + 10% FCS. Fertilised embryos were flooded with Modified Ringers Solution (MRS)/9 (11.11mM NaCl, 0.2mM KCl, 0.22mM CaCl₂, 0.11mM MgCl₂, 5mM HEPES/NaOH in Tindall et al. (2007)) after 10 minutes after fertilisation. After a further 30 minutes, embryos were dejellied in 3% L-cysteine (Sigma) in MRS/9 pH 7.88. Once dejellied, embryos were washed with reverse osmosis purified water and transferred to 1.5% agar-coated 55mm plates (VWR) containing MRS/9.

2.1.2 Microinjection

X. tropicalis embryo injections were performed in 3% ficoll (VWR International) in MRS/9 with pulled Narishige needles. Embryos were injected with either 2nl molecular grade water or 2nl 1.5ng EnGen® Spy Cas9 NLS protein (subsequently referred to as Cas9 - NEB) as controls or 2nl 600pg sgRNA and 1.5ng Cas9, which was allowed to incubate on ice for 30 minutes to enable the formation of the riboprotein complex. Embryos were cultured in this ficoll solution, to allow healing, before being transferred to MRS/20 (5mM NaCl, 0.09mM KCl, 0.1mM CaCl₂, 0.05mM MgCl₂, 5mM HEPES) before the onset of gastrulation. Embryos were staged according to Nieuwkoop and Faber (1994) and were then either flash frozen on dry ice and stored at -80°C (for genomic DNA extraction for PCRs and subsequently fragment analysis) or fixed in MEMFA (0.1M MOPS pH 7.4, 2mM EGTA, 1mM MgSO₄, 3.7% formaldehyde) for *in situ* hybridisation, after removal of the vitelline membrane.

2.1.3 Photography

Embryos were photographed using the SPOT 14.2 Colour Mosaic camera (Diagnostic Instruments Inc.) and SPOT Advanced software with a Leica MZ FLIII microscope.

2.2 Molecular biology methods

2.2.1 Gel electrophoresis

DNA and RNA samples were run on 1-1.5% agarose gels in Tris-Acetate-EDTA buffer (40mM Tris pH 7.6, 20mM acetic acid, 1mM EDTA) with ethidium bromide (1µl/100ml TAE). Samples were loaded with 6X gel loading dye (NEB) and fragment size was compared to 4µl 1 kb Plus DNA Ladder (NEB).

2.2.2 Nanodrop

The concentration of samples was determined using Nanodrop 1000 Spectrophotometer V3.8 (Thermo Fisher Scientific). 1µl molecular grade water was used as a blank in the appropriate setting (DNA, RNA and ssDNA for DNA, RNA and sgRNA), and 1µl product was measured to determine concentration in ng/µl.

2.2.3 sgRNA design

CRISPR/Cas9 target sites were selected using CHOPCHOP (<https://chopchop.cbu.uib.no/>) web tool, UCSC genome browser (<https://genome-euro.ucsc.edu/>) and NCBI (<https://www.ncbi.nlm.nih.gov/>), taking into account predicted knock-out efficiency, GC content, self-complementarity and targeted exon location.

2.2.4 sgRNA synthesis

PCR reactions contained 27µl molecular grade water, 10µl 5x HF buffer (NEB), 6.5µl 2µM common reverse primer (Sigma), 5µl 2µM forward primer (Sigma) and 1µl 10mM dNTPs (Invitrogen), with primer sequences detailed in Table 2. This was then subject to a hot start of 98°C for 30 seconds and 1µl Phusion Polymerase (NEB) added. PCR settings were as follows: 35 cycles of [98°C for 10 seconds, 60°C for 30 seconds, 72°C 15 seconds], 72°C for 10 minutes and 4°C hold. 1µl was checked on 2% agarose gel. This PCR product was then directly used in *in vitro* transcription reactions using MEGAshortscript™ Kit (Thermo Fisher Scientific), as described in manufacturer's protocol (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_055515.pdf). 2µl was checked on 2% agarose gel for successful sgRNA synthesis. 1µl DNase (Promega) was added and incubated at 37°C for 15 minutes. 40µl ammonium acetate stop solution, before continuing to a phenol chloroform clean up and ethanol precipitation.

Table 2: Primer sequences for sgRNA synthesis. Forward primers contain T7 promoter (orange), engineered GG sequence (bold) to aid T7 promoter function, sequence complementary to gene target sites (blue) and a common sequence (black), which overlaps with the universal reverse primer. A previous MSc student designed the primers targeting *fgfr1* exon 2, *fgfr4* exon 3, *fgfr11* exon 3 and *tyrosinase* exon 1 (Zilinskaite 2019).

Gene	Primer sequences (5' - 3')	Targeted exon	Domain
<i>fgfr1</i>	GCAGCTAATACGACTCACTATA GG (AACTGGGATGTTCTCCGGA) GTTTTAGAGCTAGAAATA	2	Extracellular
	GCAGCTAATACGACTCACTATA GG (ATGTGGCGAACGTTCTGCGA) GTTTTAGAGCTAGAAATA	7	Extracellular
	GCAGCTAATACGACTCACTATA GG (TCTCGGTGGATACACTGTGG) GTTTTAGAGCTAGAAATA	15	Tyrosine kinase
<i>fgfr4</i>	GCAGCTAATACGACTCACTATA (GGGAGGGAAGATTTCGCATGG) GTTTTAGAGCTAGAAATA	3	Extracellular
	GCAGCTAATACGACTCACTATA GG (AGTTTCTTGTCATGCGGTG) GTTTTAGAGCTAGAAATA	5	Extracellular
<i>fgfr11</i>	GCAGCTAATACGACTCACTATA GG (TGATGCACCGGCAGACTGTG) GTTTTAGAGCTAGAAATA	3	Extracellular
	GCAGCTAATACGACTCACTATA G (GTTATCCCAATCTCCTGAG) GTTTTAGAGCTAGAAATA	5	Extracellular
<i>tyrosinase</i>	GCAGCTAATACGACTCACTATA G (GAAAGGAACATGGTCCCTC) GTTTTAGAGCTAGAAATA	1	-
Universal reverse primer	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAGCTCTAAAC		

2.2.5 Phenol chloroform extraction and ethanol precipitation

Molecular grade water was added to bring the volume to 400µl, before adding 400µl phenol chloroform, vortexing for 20 seconds and centrifuging for 5 minutes. The aqueous phase was removed and an equal volume of chloroform isoamyl alcohol was added to it, before vortexing for 20 seconds and centrifuging for 5 minutes. The aqueous phase of this reaction was removed and 2X volume 100% EtOH and 1/10X volume 3M sodium acetate added. This was vortexed for 1 minute and stored at -20°C overnight. The sample was then centrifuged for 30 minutes at 4°C. The supernatant was removed, and the pellet washed using 100µl ice cold 70% EtOH, vortexed for 1 minute and centrifuged for 10 minutes. The supernatant was

removed, and the pellet dried via desiccation. Resuspension required 30µl or 20µl molecular grade water (for linear DNA or linear RNA respectively), vortexing for 20 seconds, heating to 80°C for 1 minute and vortexing for 20 seconds. Successful purification was checked on agarose gels and concentration was determined using Nanodrop.

2.2.6 Genomic DNA extraction

2.2.6.1 Initial genomic DNA extraction

100µl 50mM NaOH was added to 1.5ml Eppendorfs containing previously flash-frozen embryos (and subsequently stored at -80°C), before incubating for 7.5 minutes at 95°C, vortexed for 15 seconds, picofuged for 15 seconds and transferred back at 95°C for a further 7.5 mins. 25µl of 1M Tris-HCl pH 8.0 was added to neutralise. Genomic DNA from this extraction method was used as a 1:10 dilution in PCR reactions.

2.2.6.2 Optimised genomic DNA extraction

Digestion buffer consisted of 1ml embryo lysis buffer (1M Tris, 5M NaCl, 0.5M EDTA, SDS), 0.1g Chelex beads and 10µl 25µg/µl Proteinase K (Roche). 100µl digestion buffer was added to 1.5ml Eppendorfs containing previously flash-frozen embryos (and subsequently stored at -80°C), before incubating at 55°C for 1 hour and then 95°C for 15 minutes, vortexing and centrifuging for 10 minutes at 13000RPM at 4°C. Genomic DNA from this extraction method was used neat in PCR reactions.

2.2.7 PCR optimisation

The T_m of primers was calculated using NEB T_m Calculator (<https://tmcalculator.neb.com/>) and then annealing temperatures were optimised using a gradient of temperatures, resulting in the final temperatures stated in Table 3 and 4.

2.2.8 Two-step PCR for fragment analysis

Two-step PCR reactions were modified from Bhattacharya et al. (2015). Step 1 PCR reaction consisted of: 12.5µl PCR Master Mix X2 (Promega), 9.5µl molecular grade water, 1µl 10µM forward primer (Sigma), 1µl 10µM reverse primer (Sigma) and 1µl 1:10 dilution genomic DNA (from the initial genomic DNA extraction protocol). This 25µl reaction was subject to 94°C for 2 minutes, 35 cycles of [94°C for 10 seconds, X°C for 30 seconds (Table 3), 72°C for 30 seconds], 72°C for 30 minutes and 4°C hold. 5µl of this reaction was run on a 1% agarose gel to check for successful PCR fragment generation.

Table 3: Sequences of primers used in 2 step PCR reactions, to amplify regions targeted by CRISPR/Cas9. A previous MSc student designed the primers for *tyrosinase* exon 1 (Zilinskaite 2019).

Gene	Forward primer sequence (5' - 3')	Reverse primer sequence (5' - 3')	Targeted exon	Product length (bp)	Annealing temperature (°C)
<i>fgfr1</i>	(TCCCAGTCACGACGT) GATGCCTCACTGTCTG CCTT	TGACCAAAAC CAGTTCAATTC AG	2	289	55
	(TCCCAGTCACGACGT) TTGAGAAAAGTGCTCCA GCTACA	CATGGCACAT AGGACAACAG AT	7	290	52
	(TCCCAGTCACGACGT) AGCAACCTGTTGCCTTT ACTTC	GTCATGTGAT GGATATCACG G	15	290	50
<i>fgfr4</i>	(TCCCAGTCACGACGT) AGCAGCAGCATCAACT GGTA	TGGGGTATCC TTGCAGACTTT	3	292	53
	(TCCCAGTCACGACGT) ATTCGTACATTCCAAAG ATGGC	CATCGAATAGT GGGAAGAGGA C	5	298	50
<i>fgfr11</i>	(TCCCAGTCACGACGT) AAAGGTACTGGATGCT GTGGTT	CAGCTTCTACT GGGCACAAAA	3	270	56
	(TCCCAGTCACGACGT) ACACAGCCAGCGAAAA TGAG	AAGCCGCCGT TGGTACTTA	5	306	55
<i>tyr</i>	(TCCCAGTCACGACGT) AGTGAGGAGCAGCATG GAAA	TCTCTCTATCG TCAACCCAG T	1	260	56
Universal FAM primer	[6FAM]TCCCAGTCACGACGT				

Step 2 PCR reaction consisted of: 12.5µl PCR Master Mix X2, 9.5µl molecular grade water, 1µl 10µM FAM primer (Sigma), 1µl 10µM reverse primer (Sigma) and 1µl 1:40 dilution of Step 1 PCR product. This 25µl reaction was subject to 94°C for 2 minutes, 35 cycles of [94°C for 10 seconds, X°C for 30 seconds (Table 3), 72°C for 30 seconds], 72°C for 30 minutes and 4°C hold. 5µl of this reaction was run on a 1% agarose gel to check for successful PCR fragment generation.

Table 4: Sequences of primers used in 1 step PCR reactions, to amplify regions targeted by CRISPR/Cas9.

Gene	Forward primer sequence (5' - 3')	Reverse primer sequence (5' - 3')	Targeted exon	Product length (bp)	Annealing temperature (°C)
<i>fgfr1</i>	[6FAM]GATGCCTCACT GTCTGCCTT	TGACCAAACC AGTTCAATTCAG	2	274	55
<i>fgfr1</i>	[6FAM]ACACAGCCAG CGAAAATGAG	AAGCCGCCGTT GGTACTTA	5	291	55

2.2.9 One-step PCR for fragment analysis

One-step PCR reactions were modified from Bhattacharya et al. (2015). PCR reaction consisted of: 12.5µl PCR Master Mix X2 (Promega), 9.5µl molecular grade water, 1µl 10µM forward primer (Sigma), 1µl 10µM reverse primer (Sigma) and 1µl genomic DNA (from the optimised genomic DNA extraction protocol). This 25µl reaction was subject to 94°C for 2 minutes, 30 cycles of [94°C for 10 seconds, X°C for 30 seconds (Table 4), 72°C for 30 seconds], 72°C for 30 minutes and 4°C hold. 5µl of this reaction was run on a 1% agarose gel to check for successful PCR fragment generation.

2.2.10 Fragment analysis

1:20 dilution of PCR products was used for fragment analysis, which was performed in the Technology Facility at the University of York by Dr Lesley Gilbert and Dr Sally James to determine CRISPR/Cas9 construct efficiency. Thermo Fisher Data Connect™ software (<https://apps.thermofisher.com/apps/spa/#/dataconnect>) was used to construct and analyse output graphs.

2.2.11 Plasmid transformation

dam-/dcm- competent *E. coli* cells were used for transformations, whereby 25µl cells and 5µl plasmid were combined on ice for 30 minutes, heat shocked at 42°C for 45 seconds, incubated on ice for 2 minutes and incubated at 37°C for 1 hour. 1ml LB was added to the reaction before incubating for 1 hour at 37°C with agitation, before plating out bacteria onto LB-agar plates containing 100µg/ml ampicillin for overnight growth at 37°C. Colonies were isolated and grown overnight in 3ml LB-broth with 100µg/ml ampicillin at 37°C with agitation.

2.2.12 DNA minipreps

Plasmid DNA was isolated the following day according to the QIAprep Spin Miniprep Kit (Qiagen) using 1.5ml culture medium. DNA was eluted in 25µl molecular grade water. Concentration was determined using Nanodrop.

2.2.13 Template linearisation

100µl plasmid restriction digestion reactions (Table 5), containing 1µg DNA, 2µl enzyme, 10µl 10X buffer and molecular grade water to bring the total reaction volume to 100µl, were incubated at 37°C for 90 minutes. 10µl digest was loaded onto a 1% agarose gel, alongside 1µl circular plasmid, to check for complete linearisation. Concentration was determined using Nanodrop. The remaining volume was used in phenol chloroform extraction and ethanol precipitation.

2.3 *In situ* hybridisation

2.3.1 Synthesis of *in situ* hybridisation probe by DIG transcription

Dioxygenin labelled RNA *in situ* probes were synthesised by *in vitro* transcription, comprising of 4µl 5X transcription buffer (NEB), 2µl 10X DIG NTP mix (Roche), 2µl 100mM DTT (Invitrogen), 1µl polymerase, 1µg DNA, and molecular grade water to bring the total reaction volume to 20µl. This transcription reaction was incubated overnight at 37°C. 2µl was checked on a 1.5% agarose gel. 50µl molecular grade water, 50µl 5M ammonium acetate, 1µl GlycoBlue (Invitrogen) and 300µl 100% EtOH were added to the reaction and stored at -20°C overnight. The reaction was centrifuged at 13000RPM for 15 minutes at 4°C, before the supernatant was removed and the pellet was washed with 100µl 70% EtOH, vortexed for 1 minute and centrifuged for 15 minutes at 4°C. The supernatant was removed, and the pellet dried via desiccation. Probe was resuspended in 50µl molecular grade water by vortexing for 1 minute, heating to 80°C for 2 minutes and vortexing for 1 minute. 3µl probe was loaded onto a 1.5% agarose gel and then stored at -80°C.

The *n-tubulin* probe was hydrolysed to enable embryo penetration. 25µl probe was incubated at 60°C for 12.5 minutes in hydrolysis solution (80mM NaHCO₃ and 120mM Na₂CO₃ (Silva et al. 2006)). 50µl 5M ammonium acetate (NH₄OAc) and 312.5µl 100% EtOH were added, before incubating the solution overnight at -80°C. The solution was then centrifuged for 15 minutes at 13000RPM. The supernatant was removed and 100µl 70% EtOH was added, before vortexing and centrifuging for 15 minutes at 13000RPM. The supernatant was removed, and the pellet was dried via desiccation under vacuum. The probe was resuspended in 25µl water and 2µl was checked on a 2% agarose gel to confirm successful probe purification.

Table 5: Linearisation enzymes, buffers and polymerases used in the synthesis of *in situ* hybridisation probes. Enzymes and buffers used in the linearisation reaction, whereby plasmids, containing cDNA for genes of interest, were linearised. Polymerases were used in DIG transcription reactions to generate RNA probes.

Insert	Restriction enzyme	Buffer	Polymerase	Vector
<i>cdx4</i>	PvuII	M	T7	pBlueScript II SK (+) (Reece-Hoyes et al. 2002)
<i>en2</i>	XbaI	H	T3	(Hemmati-Brivanlou et al. 1991)
<i>myod</i>	EcoRI	H	T3	pCS2+, Pownall lab
<i>n-tubulin</i>	ApaI + hydrolysis	A	SP6	pGEM-5Zf (-), gift from Richard Harland
<i>rasl11b</i>	NcoI	H	SP6	pGEM T-Easy, (Cowell 2019)
<i>sox3</i>	SmaI	A	T7	pBSK (+), (Zygar et al. 1998)

2.3.2 *In situ* hybridisation

In situ hybridisations were carried out according to Harland (1991), with modifications described in Pownall et al. (1996). Embryos, stored at -20°C, were brought to room temperature with 100% EtOH, before washing once in 75% EtOH/PBSAT for 10 minutes, once in 50%/PBSAT for 10 minutes and three times in PBSAT for 5 minutes. Embryos were then treated with Proteinase K (Roche) at 10µg/ml for 7-10 minutes depending on the stage, before two washes with 0.1M Triethanolamine pH 7.8, with two consecutive additions of 12.5µl acetic anhydride to the second wash. This was followed by two PBSAT washes for 5 minutes, refixing in 10% Formalin/PBSAT for 20 minutes and five washes in PBSAT for 5 minutes each. Incubation in pre-hybridisation buffer (50% formamide, 1mg/ml total yeast RNA, 5x SSC pH7, 100µg/ml heparin, 1x Denhart's, 0.1% Tween, 0.1% CHAPS, 10mM EDTA) at 60°C on a horizontal tube rocker for 2 hours was performed before hybridisation with 3µl/ml DIG labelled antisense RNA probes at 60°C overnight. Embryos were maintained at temperatures above 60°C during two washes with hybridisation buffer for 10 minutes, three washes with 2x SSC + 0.1% Tween for 20 minutes and three washes with 0.2x SSC + 0.1% Tween for 30 minutes. Following this, embryos were washed at room temperature for two washes with Maleic acid buffer (MAB) (100mM maleic acid, 150mM NaCl, pH 7.8, 0.1% Tween) for 15 minutes and the pre-incubated in 2ml MAB + 2% BMB + 20% heat treated lamb serum (at 60°C) for 2 hours on a horizontal rocker, before replacing with fresh solution, including 1/2000 dilution of anti-DIG antibody coupled to alkaline phosphatase (Roche) and rolled at 4°C overnight. Embryos were washed three times with MAB for 5 minutes, three times with MAB for 1 hour and before two washes in alkaline phosphatase buffer (100mM Tris pH 9.5, 50mM MgCl₂, 100mM NaCl, 0.1% Tween), the first was for 3 minutes and the second for 10 minutes. BM purple (Roche)

was added to visualise the staining, before two washes in PBSAT for 15 minutes and fixing overnight in 10% Formalin. Embryos were bleached in 5% H₂O₂/PBSAT to remove pigment, therefore improving gene expression observation.

2.4 Bioinformatics analysis

2.4.1 Data set generation and initial processing

Experimental design and initial processing of RNA-Seq and microarray data sets are described in (Branney et al. 2009; King 2019; Brunsdon and Isaacs 2020).

2.4.2 Filtering

Data sets were filtered in Excel, using high or low stringency criteria (Table 6). Xenbase (<http://www.xenbase.org>) was used to update gene names in data sets and to identify gene functions (Karimi et al. 2018).

2.4.2.1 BRB Analysis

Microarray Affymetrix CEL files were normalised using the MAS5.0 algorithm and imported into BRB-Array Tools (<http://linus.nci.nih.gov/BRB-ArrayTools.html>) (Simon et al. 2007). Imported data were filtered according to spot filter, normalisation and gene filter settings. A spot filter involved thresholding the intensity at the minimum value if the intensity is below 5 and averaging the replicate spots within an array. Each array was normalised using quantiles. Genes were excluded if the percentage missing exceeded 50% and/or less than 20% of expression data values have at least a two-fold change in either direction from the genes median value.

Class comparison tool between groups of arrays enabled the generation of gene lists, using unpaired, two sample T-test with random variance model and nominal significance level $p=0.05$ (high stringency filtering) or $p=0.1$ (low stringency filtering). Volcano plots for dnFGFR data were generated using the output of class comparison tool in BRB-Array Tools.

2.4.3 Differentially expressed gene list analysis

High stringency filtered gene lists were analysed by Protein ANalysis Through Evolutionary Relationships (PANTHER, <http://www.pantherdb.org>) classification tool (Mi et al. 2019) and Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, <https://string-db.org/>) (Szklarczyk et al. 2019), using the *Mus musculus* genome, due to its better annotated genome than *X. tropicalis*. PANTHER gene ontology (GO) analysis reveals biological processes

associated with the protein products from differentially expressed genes. Genes are classified according to their functional characteristics, using Fisher's exact statistical overrepresentation tests and false discovery rate (FDR) values. STRING generates protein-protein interaction (PPI) networks based on known and predicted interactions, using a high confidence interaction score of 0.7.

Python version 3.8.3 via Spyder application (<https://www.spyder-ide.org/>) was used to determine the significance of the overlap between gene lists, by simulating with 10,000 iterations. Python script was written by Dr Katherine Newling (Appendix 1).

Figures were generated in RStudio version 1.3.959 (<https://rstudio.com/products/rstudio/download/>), requiring R version 4.0.0 (<https://www.r-project.org/>). The with function was used to construct the CSKA-FGF4 data set volcano plot and the preliminary script was written by Alastair Pizzey. The ggplot2 function, within the ggplot library, was used to construct bar charts to visualise PANTHER GO biological processes and phenotypes of CRISPR embryos. Scatterplots were generated using the ggplot function, within the tidyverse library to visualise the distribution of single replicate iFGFR RNA-Seq and microarray data sets. The heatmap.2 function, within the gplots library, was used to construct heatmaps.

Multiple List Comparator was used to compare filtered lists of genes, probes or transcripts and to construct Venn diagrams (<http://www.molbiotools.com/listcompare.html>). *Xenopus* adult and embryo illustrations were taken from BioRender (<https://biorender.com/>).

Chapter 3: Identifying well supported FGF target genes through a meta-analysis of high-throughput transcriptomic data sets

3.1 Introduction

In this chapter, existing CSKA-FGF4 (King 2019), dnFGFR (Branney et al. 2009) and drug-inducible FGFR (iFGFR) (Brunsdon and Isaacs 2020) high-throughput transcriptomic data sets will be reanalysed to investigate the specificity of transcriptional events downstream of different FGFRs. These approaches to manipulate FGF signalling can be categorised into those which increase, namely CSKA-FGF4 and iFGFR, and those which inhibit FGF signalling, namely dnFGFR. Comparisons within and between these data sets will provide a list of well supported FGF target genes for analysis in *fgfr* knockout embryos resulting from CRISPR/Cas9 targeting.

dnFGFRs lack the intracellular kinase domain and were intended to target individual receptors by forming non-productive dimers with their respective wild-type receptors. Embryo microinjection and subsequent translation of mRNA encoding dnFGFRs aimed to inhibit FGF through specific FGFRs (Amaya et al. 1991). Branney et al. (2009) utilised dnFGFR1 and dnFGFR4 in an Affymetrix microarray data set, which aimed to identify FGF targets activated immediately after the commencement of FGF signalling at the midblastula transition (MBT) and the different contributions of FGFR1 and FGFR4 in this.

Conversely, CSKA-FGF4 plasmids (previously referred to as eFGF) increase FGF signalling as FGF4 expression is driven by a β -actin promoter. Injection of this construct into embryos aimed to enable FGF4 overexpression in every cell, allowing the study of the role of FGF4 after the MBT (Isaacs et al. 1994). This data set contains FGF target genes activated between the MBT to embryo collection at neurula stage for RNA-Seq analysis (King 2019).

iFGFRs target individual receptors, however they increase FGF signalling from a precisely timed activation. Translated iFGFR proteins, from mRNA injection, are identical to their endogenous FGFR in the transmembrane and intracellular tyrosine kinase domains but differ extracellularly. The ligand binding domain, which is activated by FGF ligands, is replaced with a ligand binding domain of the synthetic dimerisation agent AP20187 (Pownall et al. 2003).

Therefore, dimerisation of and FGF signalling through iFGFRs is only activated by this agent, not endogenous FGF ligands (Brunsdon and Isaacs 2020). Two iFGFR data sets will be reanalysed in this chapter. Firstly, Affymetrix microarray data of iFGFR1 and iFGFR2 signalling reveals FGF-regulated genes in gastrula stages and secondly, the different contributions of FGFR1 and FGFR4 in gene transcription during neural induction can be investigated in neuralised animal caps subject to RNA-Seq analysis (Brunsdon and Isaacs 2020).

Reanalysis of these transcriptomic data sets will highlight well-supported FGF target genes, aiding the clustering of similarly regulated genes. The spatial expression of FGF target genes in wild type embryos will also be presented in this chapter and the effect of these various FGF signalling manipulation approaches on the expression of these genes will be analysed.

The aims of this chapter are:

- Analyse high-throughput sequencing data sets to identify genes up and down regulated by FGF signalling
- Perform gene enrichment analysis to identify the functions of and biological processes associated with differentially expressed genes
- Analyse the interactions of proteins translated from differentially expressed genes *in silico*
- Determine the statistical significance of the overlap of differentially expressed gene lists between data sets
- Cluster similarly regulated expressed genes
- Select well-supported FGF regulated target genes for analysis in FGFR-targeted CRISPR embryos
- Present the spatial expression of previously characterised FGF target genes by *in situ* hybridisation.

3.2 Results

A meta-analysis of previously generated high-throughput transcriptomic RNA-seq (CSKA-FGF4 (King 2019) and iFGFR (Brunsdon and Isaacs 2020)) and microarray (dnFGFR (Branney et al. 2009) and iFGFR (Brunsdon and Isaacs 2020)) data sets will determine the specificity of different FGF signalling manipulation approaches and identify well-supported FGF target genes for analysis in *fgfr* knockout embryos. RNA-Seq and microarray experiments will be compared separately due to differences in annotation.

Data sets will be filtered twice to produce gene lists for different subsequent analysis. High stringency filtering criteria will reveal a small number of genes to be analysed using PANTHER and STRING to investigate the biological processes associated with FGF-regulated genes and the interactions between the protein products from these genes respectively (Mi et al. 2019; Szklarczyk et al. 2019). Comparisons between genes satisfying high stringency criteria from different data sets will enable well-supported FGF target gene selection. Whereas, comparisons between the larger gene lists resulting from low stringency criteria will reveal patterns of gene expression, to be visualised in heatmaps. Filtering criteria, the number of transcripts which satisfied these and the corresponding number of genes, are outlined in Table 6.

3.2.1 Transcriptomic analysis of the effects of FGF signalling inhibition by dnFGFRs in whole embryos

dnFGFRs lack the intracellular kinase domain and were intended to target individual receptors by forming non-productive dimers with their respective wild type receptors. Embryo microinjection and subsequent translation of mRNA encoding dnFGFRs aimed to inhibit FGF signalling through specific FGFRs (Amaya et al. 1991).

In this data set, *X. laevis* embryos were injected with 4ng of *dnfgfr1* or *dnfgfr4* mRNA and collected at early gastrula stage 10.5, along with sibling control embryos for triplicate Affymetrix microarray experiments. This experimental design enables the identification of FGF targets which are activated immediately after the commencement of FGF signalling at the MBT, shown by the increase in dpERK levels, a key signal transducer in MAPK signalling (Branney et al. 2009).

Table 6: High and low stringency filtering criteria. The number of positively (red) and negatively FGF-regulated (blue) transcripts (RNA-Seq) or probe sets (microarray) which satisfy the criteria, and the number of genes they correspond to are shown in brackets.

	High stringency filtering		Low stringency filtering	
Data set	Criteria	Number of transcripts (genes)	Criteria	Number of transcripts (genes)
FGF4-CSKA (RNA-Seq)	$q \leq 0.05$ Effect ≥ 1.5	81 (75)	$q \leq 0.1$ Effect ≥ 1.5	117 (109)
	$q \leq 0.05$ Effect ≤ 0.75	92 (87)	$q \leq 0.1$ Effect ≤ 0.75	157 (149)
iFGFR1 (RNA-Seq)	FPKM ≥ 20 Fold ≥ 2	257 (189)	FPKM ≥ 10 Fold ≥ 2	479 (332)
	FPKM ≥ 20 Fold ≤ 0.5	108 (91)	FPKM ≥ 10 Fold ≤ 0.5	145 (123)
iFGFR4 (RNA-Seq)	FPKM ≥ 20 Fold ≥ 2	368 (290)	FPKM ≥ 10 Fold ≥ 2	688 (516)
	FPKM ≥ 20 Fold ≤ 0.5	186 (147)	FPKM ≥ 10 Fold ≤ 0.5	324 (244)
dnFGFR1 (microarray)	$p \leq 0.05$ Fold ≥ 2	18		
	$p \leq 0.05$ Fold ≤ 0.5	59		
dnFGFR4 (microarray)	$p \leq 0.05$ Fold ≥ 2	30	$p \leq 0.1$ Fold ≥ 2	45
	$p \leq 0.05$ Fold ≤ 0.5	75	$p \leq 0.1$ Fold ≤ 0.5	91
iFGFR1 (microarray)	Unit ≥ 20 Fold ≥ 2	45	Unit ≥ 10 Fold ≥ 2	45
	Unit ≥ 20 Fold ≤ 0.5	154	Unit ≥ 10 Fold ≤ 0.5	154
iFGFR2 (microarray)	Unit ≥ 20 Fold ≥ 2	41	Unit ≥ 10 Fold ≥ 2	41
	Unit ≥ 20 Fold ≤ 0.5	48	Unit ≥ 10 Fold ≤ 0.5	48

3.2.1.1 dnFGFRs alters similar gene expression profiles

Figure 6 shows the distribution of the microarray data, with each data point representing the $\log_2(\text{fold change})$ and $-\log_{10}(\text{p-value})$ of a single probe. The fold change is the ratio of

expression between two groups, for example a value of 1 represents no change, > 1 represents an increase relative to controls, and < 1 represents a decrease.

Data was filtered based on p-value and fold change. High stringency filtering criteria was used to classify probes with a significant change in expression satisfying a p-value ≤ 0.05 (blue, Figure 6), showing that dnFGFR1 and dnFGFR4 significantly altered the expression of a large number of genes, relative to control embryos. Probes were classed as significantly up or down regulated if they exhibited a ≥ 2 or ≤ 0.5 -fold change respectively.

dnFGFR1 and dnFGFR4 significantly up regulated 18 and 30 probes respectively (Table S1, S3), which are negatively regulated by FGF signalling in normal development. Some probes such as *hesx1-b*, encoding a homeodomain transcription factor, and *ATPase, Na+/K+ transporting, beta 2 polypeptide (atp1b2)*, encoding Na⁺/K⁺ ATPase, beta subunit, increased in both dnFGFR1 and dnFGFR4 embryos. The expression of *mylc2a* (myosin light chain 7) and *tubb2b (n-tubulin)* (fold change = 2.18) was increased in only dnFGFR1 and dnFGFR4 embryos respectively. dnFGFR1 and dnFGFR4 significantly down regulated 59 and 75 probes respectively (Table S2, S4), indicating that FGF signalling positively regulates these genes in normal development. Well-characterised FGF target genes are present in these lists, such as *spry2*, *chrd* (chordin) and *cdx4* (fold change = 0.37 (dnFGFR1), 0.15 (dnFGFR4)) (Pownall et al. 1996; Branney et al. 2009; Faas and Isaacs 2009; Lee et al. 2011), suggesting that other genes in these high stringency filtered lists are also FGF-regulated.

As alluded to, there is considerable overlap between dnFGFR1 and dnFGFR4 high stringency filtered gene lists (Table 7, Figure 7). Up regulated lists share 14 genes and there are 57 genes in common to down regulated lists.

In order to investigate the probability of these observed overlaps, Python was used to randomly sample sets of the number of up regulated probes (18 by dnFGFR1 and 30 by dnFGFR4), which represent genes negatively regulated by FGF signalling in normal development, from numbers between 1 and 15,611 (number of probes in the microarray). The highest number of overlaps in 10,000 iterations was 2 so the probability of getting an overlap of size 3 or greater is $p < 0.0001$ (Figure S1A; Table S5). Therefore, an overlap of 14 is statistically significant. dnFGFR1 and dnFGFR4 down regulated 59 and 75 probes respectively, which represent genes positively regulated by FGF signalling, and these numbers were randomly sampled. The highest number of overlaps in 10,000 iterations was 4 so the probability of getting an overlap of size 5 or greater is $p < 0.0001$ (Figure S1B; Table S5). Consequently, an overlap of 57 significantly down regulated probes is statistically

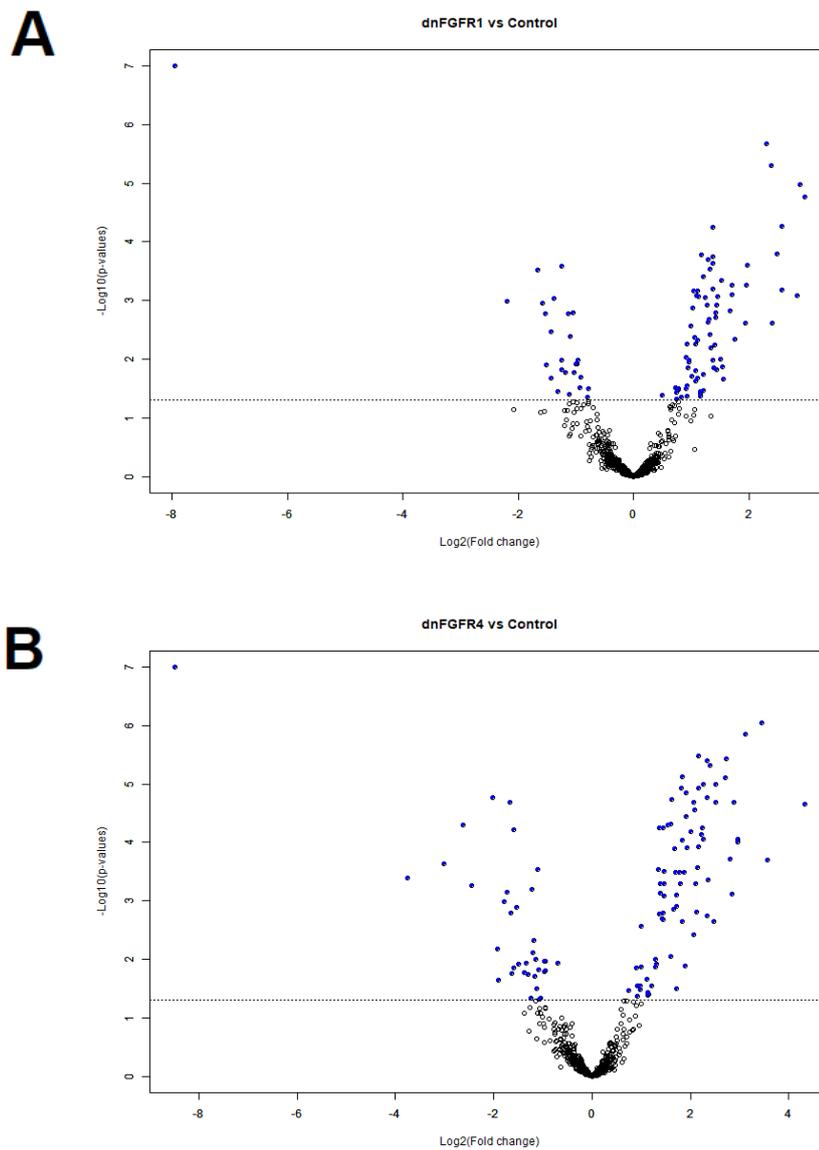


Figure 6: Distribution of Affymetrix microarray data from dnFGFR FGF signalling inhibition in early gastrula whole embryos. (A) dnFGFR1 vs Control. (B) dnFGFR4 vs Control. Data points represent the $\log_2(\text{Fold change})$ and $-\log_{10}(\text{p-value})$ values of a single probe. The horizontal line ($-\log_{10}(\text{q-value}) = 1.3$) represents $\text{p-value} = 0.05$. Below this line are probes with a $\text{p-value} \geq 0.05$ (black), which are not significant. Data points above this line represent probes which satisfy $\text{p-value} \leq 0.05$ and are significant. Volcano plots were generated using BRB-Array Tools.

significant. Taken together, dnFGFR1 and dnFGFR4 affect the expression of similar gene probes.

3.2.1.2 Genes affected by dnFGFR1 and dnFGFR4 are involved in similar biological processes in early gastrula stage whole embryos

PANTHER gene ontology (GO) analysis reveals the biological processes associated with genes in these high stringency filtered gene lists to investigate the enrichment of functions

Table 7: Overlap of genes affected by dnFGFR1 and dnFGFR4 signalling in early gastrula stage whole embryos. Gene lists satisfied the high stringency filtering criteria (p-value ≤ 0.05 , ≥ 2 fold change). Probes are distinguished by their gene name (bold text) or accession number (plain text), if the probe is unannotated. List analysis was performed using Multiple List Comparator (<http://www.molbiotools.com/listcompare.html>).

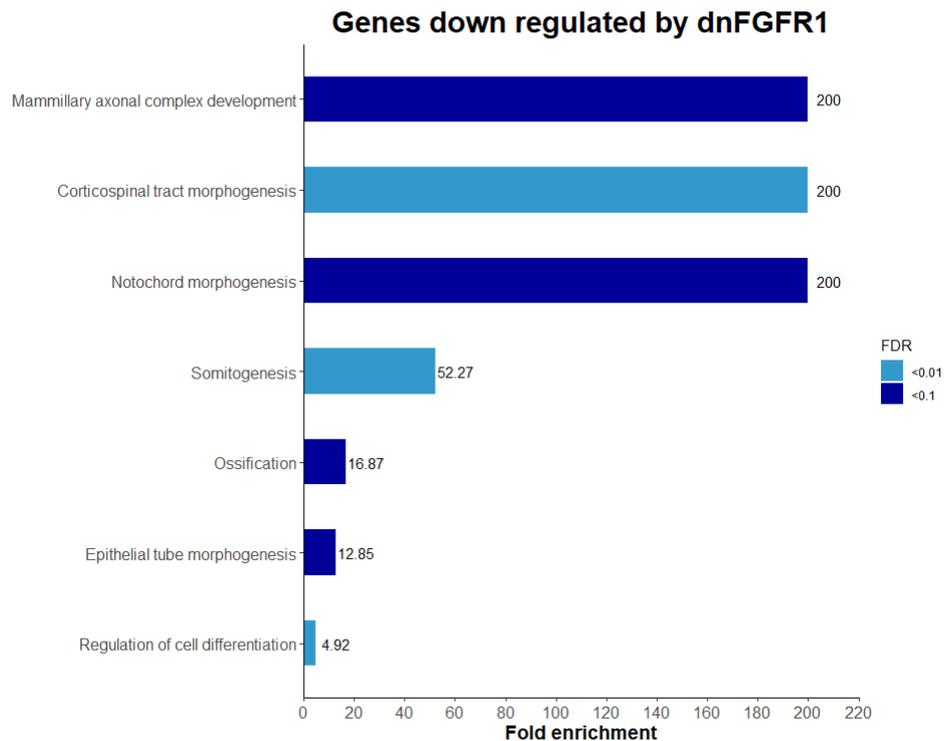
dnFGFR1 up regulated dnFGFR4 up regulated		dnFGFR1 down regulated dnFGFR4 down regulated	
Affymetrix probe set	Gene name/accession number	Affymetrix probe set	Gene name/accession number
XI.11598.1.A1_at XI.12126.1.S1_at XI.12378.1.S1_at XI.131.1.S1_at XI.15572.1.A1_at XI.16094.1.A1_at XI.1685.1.S1_at XI.23988.1.S1_at XI.2565.4.S1_x_at XI.509.1.S1_at XI.5501.1.A1_at XI.6024.1.S1_at XI.841.3.S1_a_at XI.8949.1.S1_at	AW460608 hes7.1 tsc22d3 hesx1-b BJ088128 grhl2 LOC398260 BJ044287 BG810694 atp1b2 BE027102 darmin pdgfa adc	XI.10269.1.S1_at XI.1082.1.S1_at XI.1108.1.S1_at XI.11129.1.A1_at XI.11594.1.A1_at XI.11619.1.S1_at XI.11964.1.S2_at XI.11965.1.S1_at XI.11965.1.S1_s_at XI.1299.1.S1_at XI.12993.1.A1_at XI.13.1.S1_at XI.13.2.A1_at XI.14524.1.S1_at XI.146.1.S1_at XI.15270.1.A1_at XI.15374.1.A1_at XI.15623.1.A1_at XI.1607.1.S1_at XI.16206.1.A1_at XI.16733.1.A1_at XI.18179.1.S1_at XI.1929.1.A1_at XI.19933.1.S1_at XI.19933.2.A1_at XI.212.2.S1_a_at XI.23638.1.S1_at XI.25136.1.A1_at XI.2755.1.S1_at XI.2755.2.A1_at XI.3005.1.S1_at XI.3352.1.S1_at XI.3370.1.S1_at XI.3468.1.S1_at XI.3529.1.A1_at XI.3540.1.S1_at XI.3549.1.S1_at XI.403.1.S1_at XI.4522.1.S1_at XI.514.1.S1_at XI.523.1.S1_at XI.5454.1.S1_at XI.5454.1.S2_at XI.5479.1.A1_at XI.5876.1.A1_at XI.5908.2.A1_at XI.6173.1.A1_at XI.637.1.A1_at	cdx4 foxa4-b pcdh8.2 AW766736 AW460550 AW148258 spry2 LOC398232 AF331825 alpl BJ051675 epha4-b epha4 LOC398356 myf5 BI447679 BJ077463 pfkfb3 BG347479 pnp BJ054400 BI312705 frzb BQ401062 BQ400802 frzb-1 kcnk6 cnrip1 gli1.2 sp5l sall1-a L11263 hoxd1 gl pnhd LOC398134 chrd foxb1 irx3 t-a foxd3-b xmc xmc BJ092401 apobec2 BJ051206 MGC81522 egr1-a

		Xl.642.1.S1_at Xl.644.1.S1_at Xl.7713.1.A1_at Xl.7720.1.A1_at Xl.7815.1.A1_at Xl.7969.1.S1_at Xl.802.1.S1_at Xl.933.1.S1_at Xl.958.1.S2_at	foxd411.1-a AF223426 BF231796 BF615090 MGC80198 zic3 LOC397753 t2 zeb2
--	--	--	--

associated with FGF-regulated genes (Mi et al. 2019). This provides an insight into the diverse roles of FGF signalling in development, with potentially unique contributions by different ligands and receptors. The *M. musculus* genome was selected, due to its better annotated genome than *X. tropicalis*. Genes are classified according to their functional characteristics, using Fisher's exact statistical overrepresentation tests and false discovery rate (FDR) values.

Genes up regulated by dnFGFRs, and therefore negatively regulated by FGF signalling, show minimal enrichment for GO terms, with no GO terms enriched in dnFGFR1 up regulated genes and only regulation of multicellular organismal development in dnFGFR4 up regulated genes (6.84x enrichment, FDR = 4.5×10^{-2} ; Table S7). Contrastingly, GO analysis of dnFGFR1 down regulated genes, which are positively regulated by FGF signalling, revealed enrichment in notochord morphogenesis (200x, 4.42×10^{-2}), somitogenesis (52.27x, 1.5×10^{-3}) and regulation of cell differentiation (4.92x, 7.26×10^{-3}) (Figure 8A; Table S6). These biological processes are similarly enriched in genes down regulated by dnFGFR4 (Figure 8B; Table S8), which also exhibit an enrichment for anteroposterior axis specification (54.21x, 7.0×10^{-4}), mesoderm formation (34.75x, 2.84×10^{-2}) and positive regulation of JUN kinase activity (29.12x, 4.18×10^{-2}). Genes associated with canonical Wnt signalling are enriched (27.98x, 4.34×10^{-2}), in accordance with the presence of secreted Wnt antagonist *frizzled* (*frzb*) in the genes down regulated by dnFGFR4 (Table S4). Taken together, genes regulated by different dnFGFRs show similar GO enrichment, which is expected due to the large overlap between high stringency filtered gene lists (Figure 7).

A



B

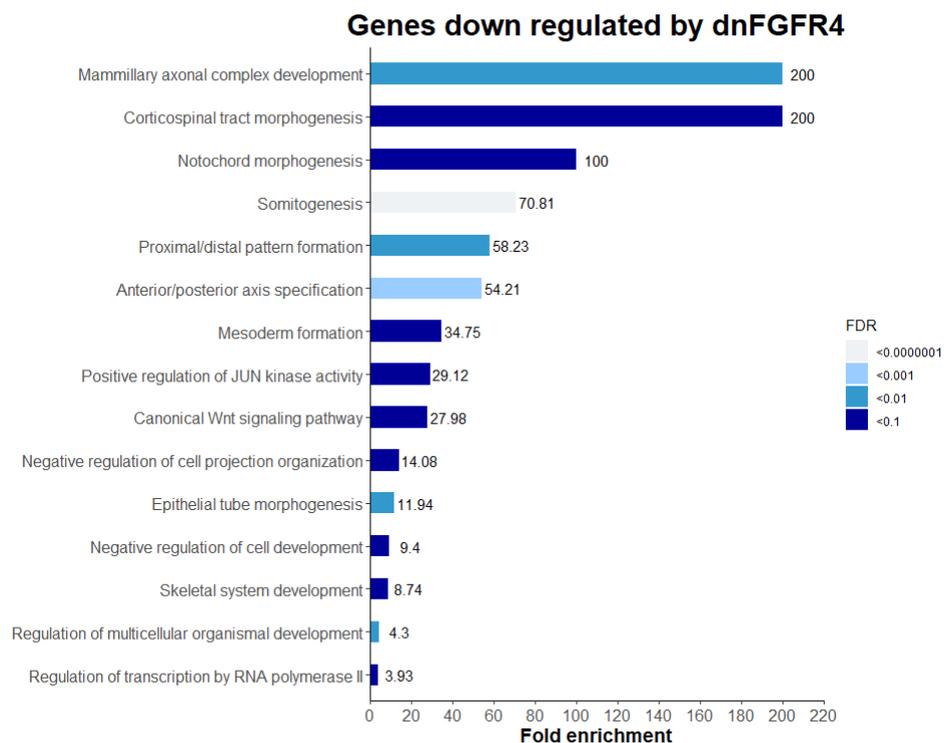


Figure 8: The enrichment of biological processes associated with down regulated genes by (A) dnFGFR1 and (B) dnFGFR4 in early gastrula stage whole embryos. Numbers to the right of bars represent fold enrichment and bar colours represent the FDR. The 15 biological processes with the highest fold enrichment are shown, with the complete list present in the corresponding table.

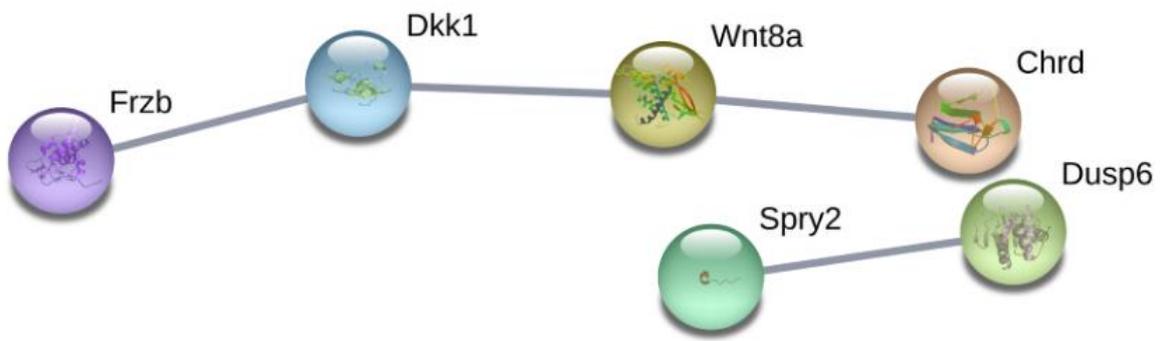


Figure 9: Protein-protein interaction networks for genes down regulated by dnFGFR4 in early gastrula stage whole embryos. The interactions shown have passed the high confidence minimum required interaction score of 0.7 and the disconnected nodes have been removed.

Spry2 and Dusp6 (Sivak et al. 2005; Branney et al. 2009). Furthermore, there are associations between Wnt signalling components Frzb, Dkk1 and Wnt8a and BMP antagonist Chrd (Figure 9). Taken together, this suggests that dnFGFR4 negatively regulates the expression of genes whose protein products interact in signal transduction and feedback loops.

3.2.2 Investigating genes regulated by FGFR1 and FGFR2 using iFGFR constructs in whole embryos

iFGFRs enable a precisely timed activation of FGF signalling, when dimerising agent AP20187 is introduced, enabling analysis of the unique roles of FGFRs in FGF signalling in early development at specific stages (Brunsdon and Isaacs 2020).

X. laevis embryos were injected at the two-cell stage with 20pg *ifgr* (either *ifgr1* or *ifgr2*) and iFGFR signalling was induced at early gastrula stage 10.5 until embryo collection at late gastrula/early neurula stage 13. Single replicate Affymetrix microarray was carried out to elucidate FGF-regulated genes in gastrula stages. This experimental design enables the investigation of iFGFR1 and iFGFR2 signalling effects on the early embryo transcriptome, since FGFR1 and FGFR2 demonstrate similar abilities to activate ERK and similar phenotypes from ectopic signalling (Brunsdon and Isaacs 2020). In this microarray data set, probe expression is measured in normalised arbitrary units.

3.2.2.1 iFGFR signalling alters the gene expression profile in gastrula stage whole embryos

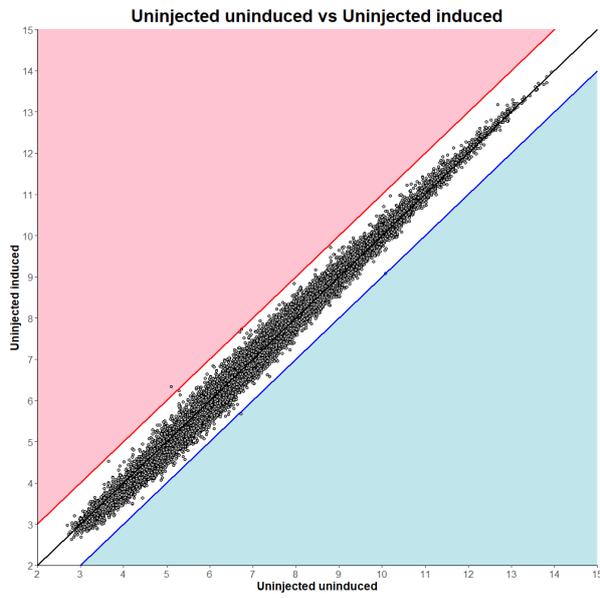
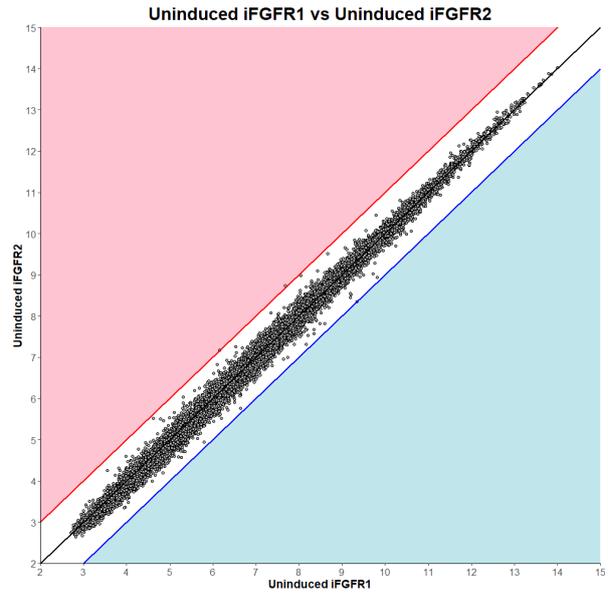
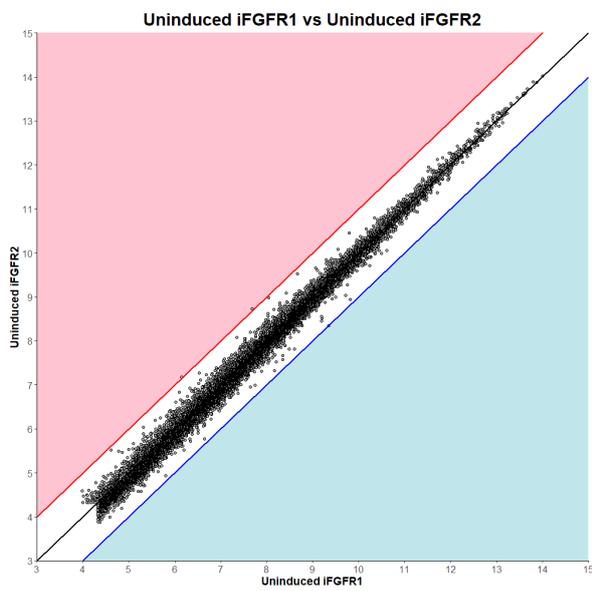
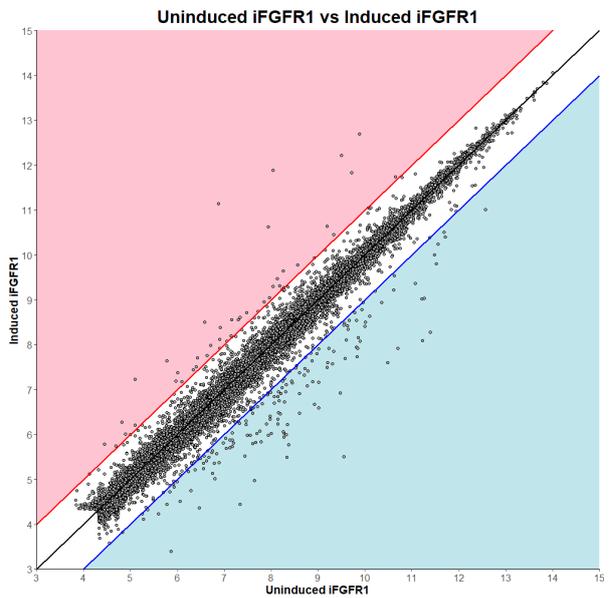
In order to determine the effect of AP20187 on gene expression, a scatterplot was generated to analyse the expression levels in uninjected control embryos cultured in the presence or absence of AP20187 (Figure 10A). Of the 15,476 probe sets passing the initial quality control,

the expression of two probe sets changed ≥ 2 fold (1 up and 1 down) due to the addition of AP20187, demonstrating that AP20187 had little effect on gene expression in the absence of iFGFRs.

In order to visualise the expression of probe sets and choose appropriate filtering conditions, a scatterplot was generated, comparing the \log_2 values in uninduced iFGFR1 and uninduced iFGFR2 embryos (Figure 10B). The majority of data points are situated along $y=x$ showing no fold expression change between the two uninduced control samples, except only two probe sets showed a ≥ 2 fold increase in expression, out of 15,476 probe sets. This is expected because although the iFGFRs are expressed, they are inactive and therefore should not increase FGF signalling and alter gene expression.

Since the data set contains only a single biological replicate, a strict filtering criteria is required. The data set was filtered using high stringency criteria, whereby genes are classed as being differentially expressed if ≥ 20 arbitrary unit is satisfied in either or both samples and if they exhibit an expression change of ≥ 2 -fold. Of the 13,461 probe sets which satisfy arbitrary unit ≥ 20 in either or both uninduced samples, only 2 probe sets showed at least a two-fold increase in expression (Figure 10C). This shows that the expression of probe sets in the absence of AP20187 in uninduced iFGFR1 and uninduced iFGFR2 embryos was highly similar.

45 and 154 probes were identified as being up and down regulated from iFGFR1 activation (Figure 10D; Table S9, S10), and 41 and 48 probes as up and down regulated from iFGFR2 activation (Figure 10E; Table S11, S12), when compared with respective uninduced controls. This shows that there is a considerable number of genes altered in response to increased FGF signalling through iFGFR1 and iFGFR2. iFGFR1 up regulated probe list contains *hoxa7*, *hoxa10* and *cdx1* (Table S9), which are associated with known FGF functions in regulating Cdx and Hox genes (Isaacs et al. 1994; Pownall et al. 1996; Faas and Isaacs 2009). iFGFR1 down regulated *tubb2b* (*n-tubulin*), along with *anterior gradient 1* (*agr1*), *agr2* and *agr3* genes (Table S10). iFGFR2 up regulated *wnt8a* and FGF target gene *spry2* (Table S11). iFGFR2 down regulated *frizzled class receptor 4* (*fzd4*), a transmembrane receptor in the Wnt signalling pathway (Karimi et al. 2018) (Table S12).

A**B****C****D**

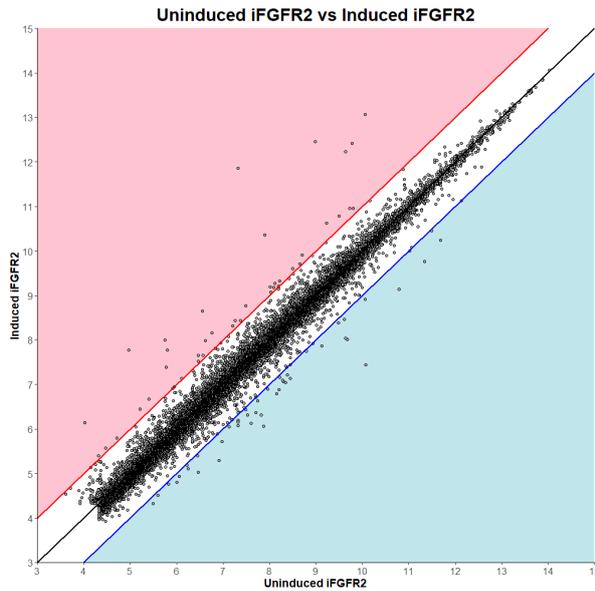
E

Figure 10: Ratio of gene expression, as \log_2 values, in iFGFR single replicate Affymetrix microarray data in gastrula stage whole embryos. (A) unfiltered uninjected uninduced plotted against uninjected induced. (B) unfiltered uninduced iFGFR1 plotted against uninduced iFGFR2. (C) filtered uninduced iFGFR1 plotted against uninduced iFGFR2. (D) filtered uninduced iFGFR1 plotted against induced iFGFR1. (E) filtered uninduced iFGFR2 plotted against induced iFGFR2. Data was filtered according to arbitrary unit ≥ 20 . The black line is $y=x$. Data points in the red area (to the left of $y=x+1$, red line) are transcripts which exhibit a ≥ 2 fold up regulation by iFGFR induction and those in the blue area (to the right of $y=x-1$, blue line) a ≥ 2 fold down regulation.

3.2.2.2 iFGFR1 and iFGFR2 affect the expression of overlapping genes in gastrula stage whole embryos

In order to compare the genes regulated by iFGFR1 and iFGFR2 signalling, a scatterplot was generated to analyse the expression of probes in embryos subject to increased FGF signalling through these inducible receptors. The expression of 49 probes show a greater than two-fold difference between iFGFR1 and iFGFR2 induced embryos (Figure 11A), indicating that at this stage of development, the genes affected by increased FGF signalling through iFGFR1 and iFGFR2 are similar.

Comparisons of these filtered probe lists from iFGFR1 and iFGFR2 signalling, according to the high stringency criteria (Table 6), revealed probes common to both (Figure 11B; Table 8). These shared probes are only consistently regulated, for example both either up or down regulated by iFGFR1 and iFGFR2 signalling. 25 probe sets are shared between iFGFR1 and iFGFR2 up regulated lists, including *hoxa7*, *egr1-a* and *junb*. 41 probe sets are shared between iFGFR1 and iFGFR2 down regulated lists, including homeodomain transcription factor *pitx2-a*.

Table 8: Overlap of genes affected by iFGFR1 and iFGFR2 signalling in gastrula stages in whole embryos. Probe set lists satisfied the high stringency filtering criteria (arbitrary unit ≥ 20 , fold change ≥ 2). List analysis was performed using Multiple List Comparator (<http://www.molbiotools.com/listcompare.html>). Codes in the left hand column are Affymetrix probe sets, right hand column contains gene names (bold text) or accession number (plain text), if the probe is unannotated.

iFGFR1 upregulated iFGFR2 upregulated		iFGFR1 downregulated iFGFR2 downregulated	
Affymetrix probe set	Gene name/accession number	Affymetrix probe set	Gene name/accession number
XI.12130.1.S1_at	hoxa7	XI.10362.1.A1_at	ca12
XI.12993.1.A1_at	BJ051675	XI.104.1.S1_at	pitx2-a
XI.13967.1.A1_at	BJ089550	XI.10583.1.S1_at	slc3a2
XI.14397.1.S2_at	nek6	XI.11128.1.S1_at	AW766729
XI.15202.1.A1_at	AW766492	XI.11145.1.A1_at	AW766955
XI.15920.1.A1_at	BJ048594	XI.12727.1.A1_at	fa2h
XI.16457.1.A1_at	junb	XI.1317.1.A1_at	BI443530
XI.18073.1.A1_at	BG885063	XI.13575.1.A1_at	b3gnt1
XI.20488.1.S1_at	BQ731489	XI.14452.1.A1_at	bcat1
XI.22857.1.A1_at	BJ088428	XI.15163.1.S1_at	BG408248
XI.23897.1.S1_at	cnfn-a	XI.15702.1.A1_at	BJ076178
XI.23988.1.S1_at	BJ044287	XI.1589.1.S1_at	agr3
XI.24218.1.S1_at	dynll1-a	XI.1616.1.A1_at	fam115a
XI.24294.1.S1_at	BJ098811	XI.16262.1.A1_at	BJ051781
XI.24337.1.A1_at	CB564601	XI.16466.1.A1_at	BJ082483
XI.24793.1.S1_at	tspan1	XI.16543.1.S1_at	LOC398404
XI.4789.1.S1_at	MGC52875	XI.16589.1.A1_at	BJ080730
XI.4965.1.S1_at	irg1	XI.1683.1.S1_at	MGC68910
XI.5082.1.A1_at	MGC68521	XI.1685.1.S1_at	LOC398260
XI.637.1.A1_at	egr1-a	XI.18858.1.A1_at	BI446930
XI.708.1.S1_at	lefty-a	XI.20089.1.S1_at	foxi1
XI.736.1.S1_at	LOC398207	XI.22874.1.A1_at	ubp1
XI.8124.1.S1_at	MGC115642	XI.24091.1.A1_at	CB565543
XI.880.1.S1_at	wnt3a	XI.24199.1.A1_at	CB756654
XI.9671.1.S1_at	capn8-a	XI.24199.3.A1_at	BJ049353
		XI.24565.1.A1_at	BG485946
		XI.25847.1.S1_at	agr2
		XI.26141.1.S1_at	MGC80993
		XI.2659.1.S1_at	atp12a-b
		XI.2924.1.S1_at	MGC53311
		XI.522.1.S1_at	pitx1
		XI.5296.1.A1_at	BJ080084
		XI.5324.1.S1_at	otog
		XI.5846.19.S1_at	CB560320
		XI.5912.1.A1_at	eppk1
		XI.6266.1.S1_at	itln1
		XI.7213.3.S1_a_at	LOC100158288
		XI.8935.1.A1_at	BJ078657
		XI.909.1.S1_at	xepsin
		XI.9576.1.S1_at	ca2
		XI.9974.1.A1_at	BJ088045

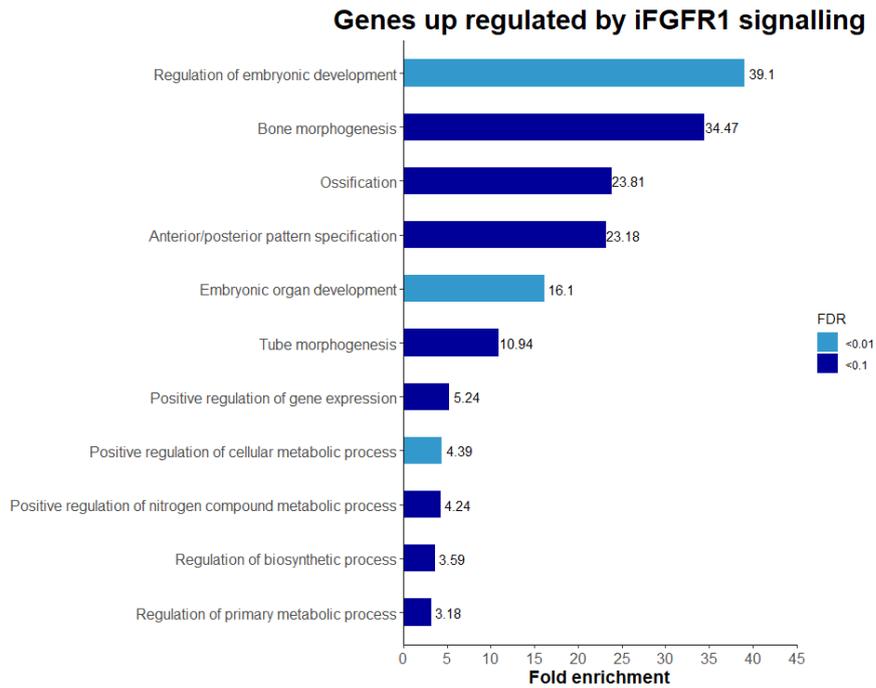


Figure 12: The enrichment of biological processes associated with up regulated genes by iFGFR1 in gastrula stage whole embryos. Numbers to the right of bars represent fold enrichment and bar colours represent the FDR. The 15 biological processes with the highest fold enrichment are shown, with the complete list present in the corresponding table.

(FDR = 1.88×10^{-2}), in line with the role of FGF signalling in anteroposterior patterning (Amaya et al. 1991; Isaacs et al. 1994; Pownall et al. 1996; Faas and Isaacs 2009) (Table S14; Figure 12) and suggesting FGF target genes are present in this list.

3.2.2.4 iFGFR protein-protein interaction networks contain Wnt and GTPase proteins

PPI networks for iFGFR1 and iFGFR2 up regulated genes both show interactions involving Wnt3a with either Cdx1 (iFGFR1; Figure 13A) or Wnt8a (iFGFR2; Figure 13C), demonstrating the role of FGF signalling in regulating other pathways. However, these networks do not show a significant enrichment for interactions (PPI enrichment p-value = 8.08×10^{-2} ; 3.18×10^{-1}). iFGFR1 down regulates genes whose protein products show an association between thioredoxin domain containing Agr1 and small GTPase Rabs (Figure 13B) and this network shows a significant enrichment for PPIs (4.77×10^{-2}). Proteins from iFGFR2 down regulated genes do not show any interactions. Taken together, this suggests that iFGFR1 and iFGFR2 positively regulate the expression of genes whose protein products interact in the Wnt signalling pathway, and the expression of interacting GTPases is negatively regulated by iFGFR1 signalling.

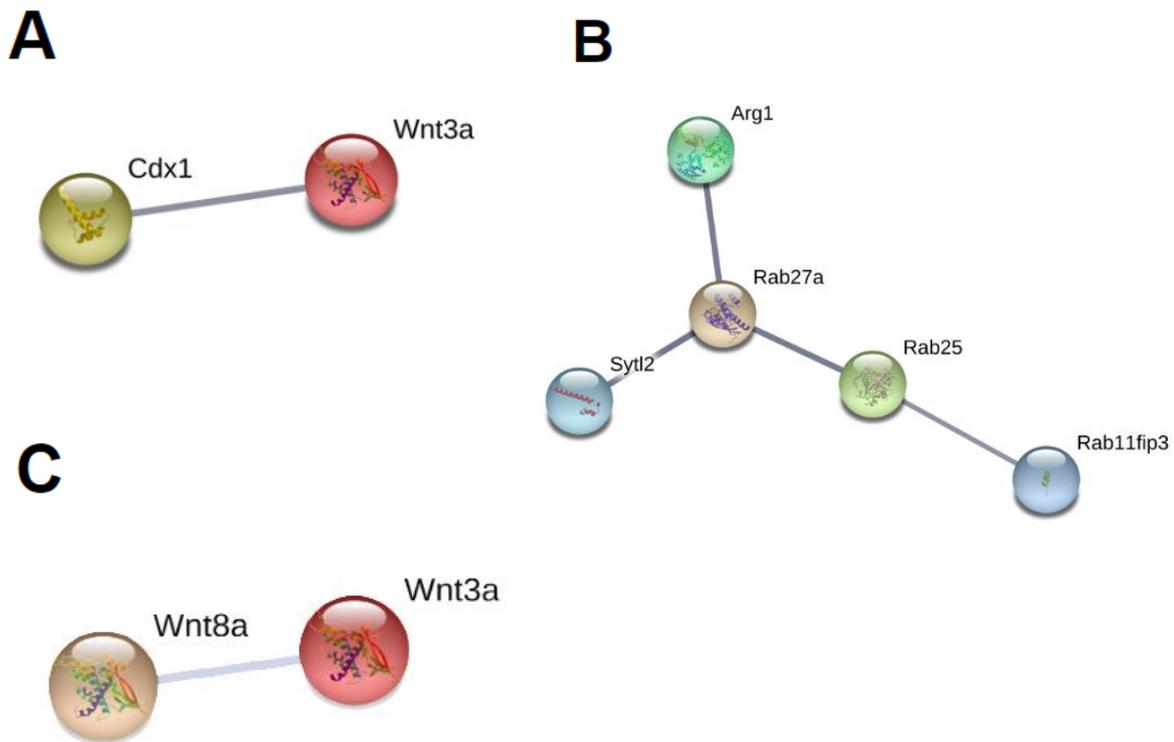


Figure 13: Protein-protein interaction networks for genes up regulated by iFGFR1 (A), down regulated by iFGFR1 (B) and up regulated by iFGFR2 (C) gastrula stages in whole embryos. The interactions shown have passed the high confidence minimum required interaction score of 0.7 and the disconnected nodes have been removed.

3.2.3 dnFGFR and iFGFR microarray data set comparison

Comparisons between data sets using high and low stringency filtered gene lists will identify consistently regulated FGF target genes and patterns of gene expression respectively. Due to the considerable overlap between the genes affected by dnFGFR1 and dnFGFR4, dnFGFR4 will be used in the following data set comparison analysis as it was found to be more potent than dnFGFR1, when compared per mass of injected mRNA (Branney et al. 2009).

3.2.3.1 Identification of well-supported FGF target genes, in dnFGFR4, iFGFR1 and iFGFR2 data sets

Comparisons between differentially expressed genes in dnFGFR and iFGFR data sets will enable the identification of well-supported FGF target genes, to be analysed in CRISPR/Cas9 embryos. These microarrays both utilise *X. laevis* and the same probe annotation, meaning that comparisons can be performed at the probe level. Lists of probe sets were compared between data sets to determine which probes are present in both analyses, which revealed 15,476 shared probes present in both data sets (Figure S3). High stringency filtered gene lists (Table S3, S4; S9, S10, S11, S12) were then overlapped (Table 9) to identify genes, whose expression is affected by FGF signalling in both data sets.

Table 9: Overlap of dnFGFR4 and iFGFR1/2 probe gene lists, which satisfied high stringency filtering criteria. Numbers in brackets define the number of probes present in each filtered probe list. Light grey box colour represents a statistically significant overlap between probe lists. White box colour represents an insignificant overlap between probe lists. Probe names are stated and in brackets is either the corresponding gene name if it is annotated (bold) or the accession number if unannotated (plain text). List analysis was performed using Multiple List Comparator (<http://www.molbiotools.com/listcompare.html>).

	iFGFR1 up regulated (45)	iFGFR1 down regulated (154)	iFGFR2 up regulated (41)	iFGFR2 down regulated (48)
dnFGFR4 up regulated (30)	XI.23988.1.S1_at (BJ044287) XI.4965.1.S1_at (irg1) XI.736.1.S1_at (LOC398207) XI.8124.1.S1_at (MGC115642)	XI.1685.1.S1_at (LOC398260) XI.509.1.S1_at (atp1b2) XI.1604.1.A1_at (MGC78986)	XI.23988.1.S1_at (BJ044287) XI.4965.1.S1_at (irg1) XI.736.1.S1_at (LOC398207) XI.8124.1.S1_at (MGC115642)	XI.1685.1.S1_at (LOC398260) XI.841.3.S1_a_at (pdgfa)
dnFGFR4 down regulated (75)	XI.12993.1.A1_at (BJ051675) XI.637.1.A1_at (egr1-a) XI.11965.1.S1_at (LOC398232) XI.7720.1.A1_at (BF615090) XI.1465.1.S1_s_at (BC046253)	-	XI.12993.1.A1_at (BJ051675) XI.637.1.A1_at (egr1-a) XI.1082.1.S1_at (foxa4-b) XI.11964.1.S2_at (spry2) XI.15623.1.A1_at (pfkfb3) XI.49.1.S1_at (wnt8a)	-

Python was used to investigate the statistical significance of the observed overlaps between high stringency filtered gene lists from dnFGFR4 and iFGFR1/2 data sets (Table 9; Table S15; Figure S4). Probes up regulated in response to dnFGFR4 and iFGFR1 or iFGFR2 are identical and both exhibit a significant overlap (Figure S4A, S4C). These probes exhibit a differential response to FGF signalling manipulation because they are up regulated in response to FGF signalling inhibition by dnFGFR4 and up regulated to increased FGF signalling by iFGFR1 and iFGFR2.

Conversely, probes present in the overlap of probes up regulated by dnFGFR4 and down regulated by iFGFR1 and iFGFR2 are down regulated by FGF signalling (Table 9), however they do not exhibit a significant overlap (Table S15; Figure S4A, S4D). Probes present in the overlap of probes down regulated by dnFGFR4 and up regulated by iFGFR1 or iFGFR2 are consistently positively regulated by FGF signalling (Table 9). Therefore, along with both overlaps being statistically significant (Table S15; Figure S4E, S4F), the probes present are likely to represent biologically relevant genes, in the context of being regulated by FGF signalling. The expression of well-characterised FGF target genes *spry2* and *egr1* are decreased by dnFGFR4 and increased by iFGFR2, whereas, *egr1* is also increased by iFGFR1. The overlap between probes down regulated by dnFGFR4 and up regulated by

iFGFR2 signalling highlights two genes, namely *forkhead box A4 (fox4a-b, fox4a)* and *wnt8a (wnt8)*. The initial publication of dnFGFR data clustered genes positively regulated by FGF signalling by patterns of temporal expression, which grouped *fox4a* with *egr1* and *goosecoid (gsc)*, and *wnt8* with *dusp5*, *noggin* and *spry2* (Branney et al. 2009).

In conclusion, there is a high degree of similarity between probes present in separate iFGFR1 and iFGFR2 overlaps with dnFGFR4, providing further indication that FGFR1 and FGFR2 regulate the expression of similar genes. There are significant overlaps between dnFGFR4, iFGFR1 and iFGFR2 affected probe lists, resulting in the selection of known FGF target genes *egr1* and *spry2* (Branney et al. 2009). The expression of these genes will be analysed in CRISPR/Cas9 embryos to validate successful FGF signalling inhibition and to investigate if FGFR targeting differentially affects the expression of these selected genes.

3.2.3.2 Low stringency filtering data set comparison

Filtering these data sets according to low stringency criteria (Table 6) reveals larger probe lists (Table S16, S17; S18, S19, S20, S21). Up and down regulated probe lists were compiled into differentially expressed probe lists, containing 199 genes for iFGFR1, 89 for iFGFR2, and 136 for dnFGFR4. These were overlapped (Figure S5) and results are stated in Table S22.

The expression of genes in iFGFR1 and iFGFR2 microarray experiments can be compared quantitatively, as they use the same experimental design. The normalised expression of overlapping genes, using \log_{10} of arbitrary units, is visualised in Figure 14. The gene expression profiles of probes satisfying the low stringency filtering criteria in iFGFR1 and iFGFR2 embryos appear almost identical, providing further evidence that FGFR1 and FGFR2 have similar roles in development (Figure 14A). Genes present in clusters with known FGF targets may be biologically relevant since they exhibit a highly similar expression pattern to known FGF target genes, suggesting a similar degree of regulation (Figure 14B). *NIMA-related kinase 6 (nek6)*, encoding a serine/threonine kinase, is clustered with *egr1-a* and *hoxa7* is clustered with microtubule motor component *dynein light chain LC8-type 1 (dynll1)*.

Although the iFGFR1/2 and dnFGFR4 microarrays use the same method of transcriptomic analysis and *X. laevis* as the model organism, they differ in their protocols and developmental stages. Therefore, the expression of probes cannot be compared quantitatively in heatmaps, even using transcript expression relative to respective controls. Instead, the expression of these overlapping probes can be compared qualitatively in heatmaps, by scoring genes +1 if their expression increases (red) and -1 if their expression decreases (blue) relative to controls.

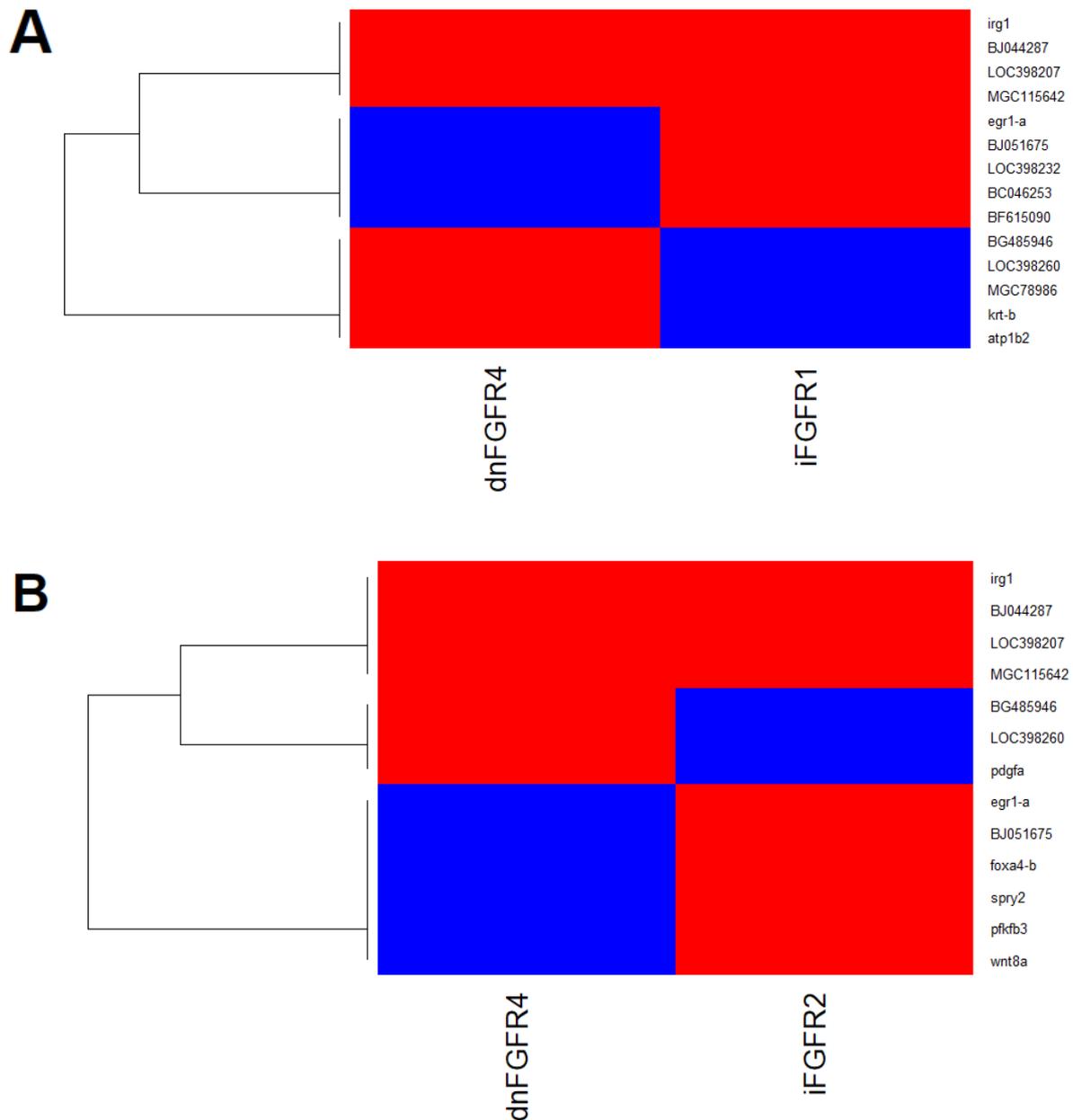


Figure 15: Expression of genes affected by dnFGFR4 and iFGFR1 (A), and dnFGFR4 and iFGFR2 signalling (B). The expression of overlapping genes is scored +1 if their expression increases (red) and -1 if their expression decreases (blue) relative to controls. Genes satisfied low stringency filtering criteria and are Euclidean clustered.

Comparisons between dnFGFR4 and iFGFR1 or iFGFR2 reveal genes which are negatively regulated by FGF signalling. This includes *keratin 70* (*krt-b*, *krt70*), an intermediate filament, and *atp1b2*, encoding Na⁺/K⁺ ATPase, beta subunit (Karimi et al. 2018), which are affected by dnFGFR4 and iFGFR1 (Figure 15A; Table S22), whereas *foxa4-b*, *wnt8a* and *6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3* (*pfkfb3*) are affected by dnFGFR4 and iFGFR2 (Figure 15B; Table S22). *platelet derived growth factor subunit A* (*pdgfa*) is up regulated in dnFGFR4 embryos and down regulated in iFGFR2 embryos, suggesting negative regulation by FGF signalling.

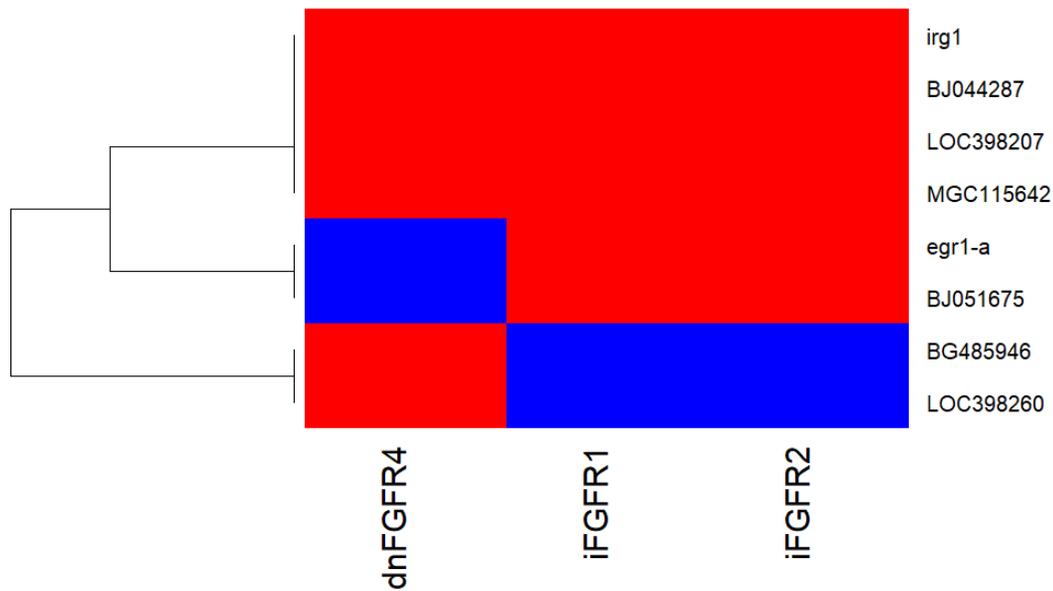


Figure 16: Expression of genes affected by dnFGFR4, iFGFR1 and iFGFR2. The expression of overlapping genes is scored +1 if their expression increases (red) and -1 if their expression decreases (blue) relative to controls. Genes satisfied low stringency filtering criteria and are Euclidean clustered.

There are 8 probes overlapping in dnFGFR4, iFGFR1 and iFGFR2 data sets (Figure 16; Table S22). *egr1-a* exhibits a decrease in embryos subject to FGF signalling inhibition by dnFGFR4 and an increase in expression from iFGFR1/2 signalling, in accordance with positive regulation by FGF signalling (Branney et al. 2009). Contrastingly, there are differences in the regulation of 4 genes, which are all upregulated in response to dnFGFR4 and iFGFR1/2, including *aconitate decarboxylase 1 like gene b (irg1, acod1lb)*.

3.2.3.3 Microarray analysis summary

Comparisons within data sets revealed that dnFGFR1/4 and iFGFR1/2 affected the expression of the similar groups of genes. Comparisons between high stringency filtered gene lists identified well-characterised FGF target genes *egr1* and *spry2*, which were significantly down regulated in embryos subject to FGF signalling inhibition dnFGFR4 and up regulated from iFGFR signalling. Low stringency filtering enabled the analysis of the degree of gene inhibition or activation by iFGFR1 and iFGFR2, which is highly similar, suggesting that FGFR1 and FGFR2 have similar roles in gene regulation in early development. Furthermore, *nek6* and *dynll1* were identified due to clustering with *egr1-a* and *hoxa7* respectively in iFGFR embryos, however further experiments are required to investigate this.

3.2.4 Investigating FGF-regulated genes using FGF4 overexpression in early neurula stage embryos

A key consideration of gene expression manipulation approaches is the MBT. Cytoplasmic maternal factors control very early *Xenopus* embryo development until the MBT, when the previously quiescent zygotic nuclear genome becomes activated and transcription commences (Lee et al. 2014). It is therefore crucial that mechanisms to alter FGF signalling do not disrupt this fundamental process. This is the case for the CSKA plasmid approach. Injection of this construct into embryos aimed to enable FGF4 overexpression in every cell, allowing the study of the role of FGF4 after the MBT (Isaacs et al. 1994), which is more favourable than the rapid accumulation of FGF protein resulting from synthetic mRNA injection.

In this data set, 50pg CSKA-FGF4 was injected into two-cell stage *X. tropicalis* embryos. Triplicate embryos were collected at neurula stage 14, along with 4nl water-injected sibling controls, for RNA-Seq using Illumina sequencing, which identified 43,588 transcripts for 23,635 genes (King 2019).

3.2.4.1 FGF4 overexpression alters gene expression profiles in early neurula stage whole embryos

Figure 17 shows the distribution of the data, with each data point representing the $\log_2(\text{effect size})$ and $-\log_{10}(\text{q-value})$ of a single transcript. The q-value is a FDR adjusted p-value; a value of 0.05 means 5% of significant genes will be false positives. The effect size is the magnitude difference between two groups, for example a value of 1 represents no change, > 1 represents an increase relative to controls, and < 1 represents a decrease.

Data was filtered based on q-value and effect size. Gene transcripts with a significant change in expression satisfied a q-value ≤ 0.05 (purple, Figure 17), and up and down regulated genes having effect sizes of ≥ 1.5 and ≤ 0.75 respectively. This resulted in the identification of 81 significantly up regulated transcripts (blue, Figure 17; Table S23), which are positively regulated by FGF signalling in normal development. *fgf4* exhibits the highest effect size of 27.9, confirming its overexpression. The most significant increase was in the expression of *fos* transcription factor (7.15×10^{-17}) and its $-\log_{10}(\text{q-value})$ is 16.1 (Figure 17). Among this narrowed gene list were genes previously characterised as positively regulated FGF targets, including *egr1* and *spry2* (Branney et al. 2009), suggesting that the RNA-Seq data set contains FGF target genes. Putative FGF target gene *rasl11b* was also among this list (q-value = 0.001, effect size = 1.99). 92 transcripts were significantly down regulated (red, Figure 17; Table

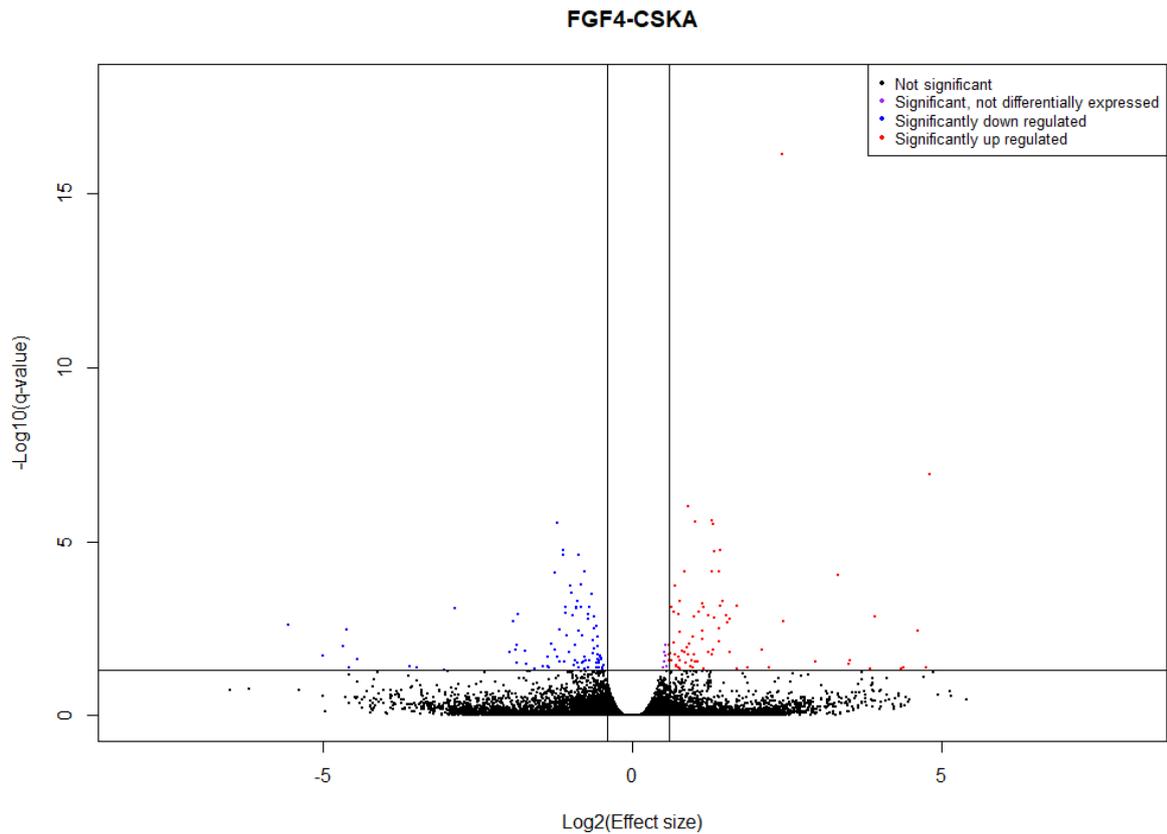


Figure 17: Distribution of CSKA-FGF4 data. Data points represent the $\log_2(\text{Effect size})$ and $-\log_{10}(\text{q-value})$ values of a single transcript. The horizontal line ($-\log_{10}(\text{q-value}) = 1.3$) represents $\text{q-value} = 0.05$. Below this line are transcripts with a $\text{q-value} \geq 0.05$ (black), which are not significant. Data points above this line represent transcripts which satisfy $\text{q-value} \geq 0.05$. These are either not differentially expressed (purple), up regulated (red, to the right of the vertical line at $\log_2(\text{Effect size}) = 0.585$, representing effect size = 1.5) or down regulated (blue, to the left of the vertical line at $\log_2(\text{Effect size}) = 0.415$, representing effect size = 0.75).

S24), indicating that FGF signalling negatively regulates these genes, which include *pax6* and *fgr4*. Transcripts from the same gene are not present in the up and down regulated transcript lists, suggesting FGF4 signalling doesn't differentially regulate splice isoforms.

3.2.4.2 Many differentially expressed genes function in FGF signalling feedback

GO analysis of up regulated genes revealed enrichment in transmembrane receptor protein serine/threonine kinase signalling pathway (10.49x enrichment, $\text{FDR} = 4.5 \times 10^{-2}$), regulation of cellular response to growth factor stimulus (8.5x, 3.68×10^{-2}), cellular response to growth factor stimulus (7.85x, 8.98×10^{-3}), regulation of protein kinase activity (5.13x, 3.17×10^{-2}) and negative regulation of signal transduction (5.13x, 3.22×10^{-3}) (Table S25, Figure 18A). This is in accordance with FGF functioning as a growth factor and stimulating downstream signalling pathways. Furthermore, this suggests that the expression of these genes is FGF-regulated as FGF4 interacts with tyrosine kinase receptors to activate intracellular kinases (Cowell 2019).

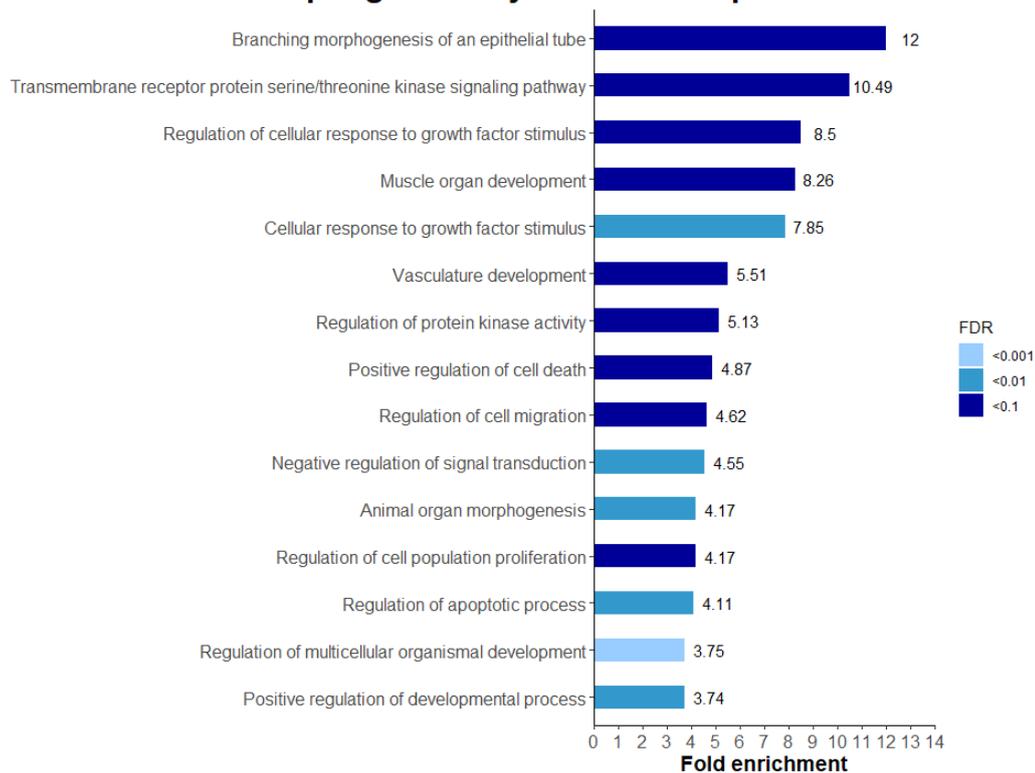
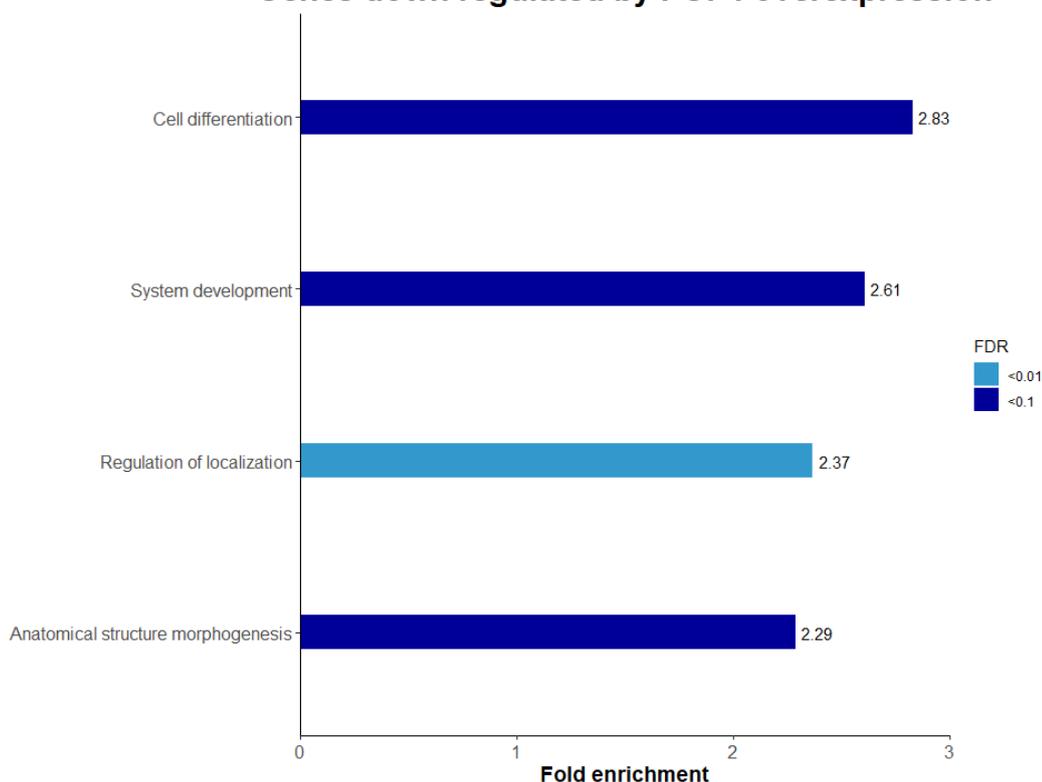
A**Genes up regulated by FGF4 overexpression****B****Genes down regulated by FGF4 overexpression**

Figure 18: The enrichment of biological processes associated with up (A) and down (B) regulated genes by FGF4 overexpression in early neurula stage whole embryos. Numbers to the right of bars represent fold enrichment and bar colours represent the FDR. The 15 biological processes with the highest fold enrichment are shown, with the complete list present in the corresponding table.

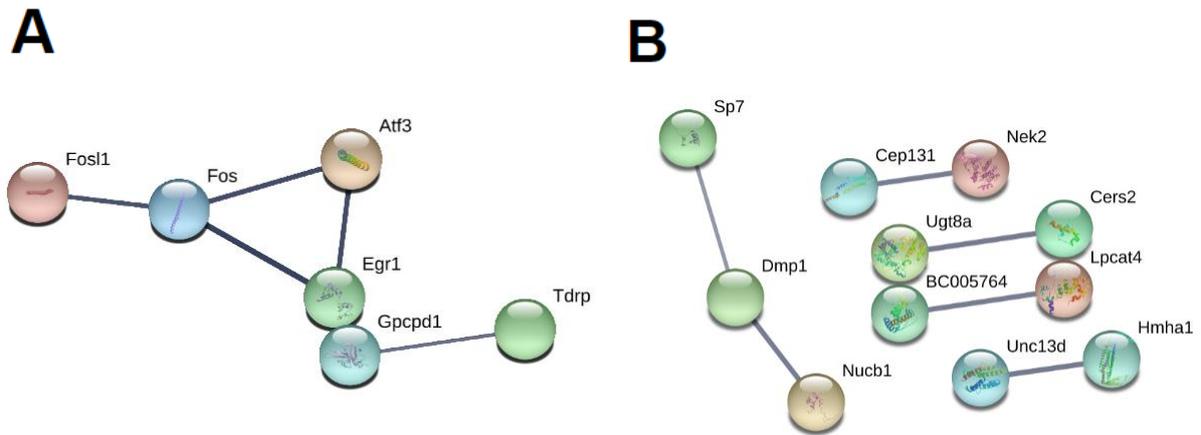


Figure 19: Protein-protein interaction networks for genes up (A) and down (B) regulated by FGF4 signalling in early neurula stage whole embryos. The interactions shown have passed the high confidence minimum required interaction score of 0.7 and the disconnected nodes have been removed.

Interestingly, genes positively regulated by FGF4 exhibit 8.26x enrichment for genes involved in muscle organ development (3.76×10^{-2}). GO analysis of genes down regulated by FGF4 showed a 2.29x enrichment in genes associated with cell differentiation (3.05×10^{-2}) (Table S26, Figure 18B), in accordance with the role of the MAPK pathway in cell fate determination (Figure 1). In conclusion, GO analysis highlights expected biological processes regulated by FGF signalling in development, suggesting the up and down regulated gene lists contain FGF target genes.

3.2.4.3 Identification of interacting transcription factors in early neurula stage whole embryos

The PPI network generated from FGF4 up regulated genes has significantly more interactions than expected (PPI enrichment p-value = 6.74×10^{-3} ; Figure 19A). There is a triangle of interactions between transcription factors Fos, Atf3 and Egr1. Contrastingly, the down regulated gene list does not show a significant enrichment in PPIs (0.119), although there are interactions between transcription factor Sp7, Dmp1 and DNA-binding Nucb1 (Figure 19B). Taken together, this suggests that FGF4 up regulates the transcription of genes whose protein products are transcription factors and interact with each other.

3.2.5 Investigating the differences in the FGFR1 and FGFR4-regulated transcriptome during neural development using iFGFR constructs in animal cap explants

X. laevis embryos were co-injected at the two-cell stage with 20pg *ifgfr* (either *ifgfr1* or *ifgfr4*) and 50pg *noggin* mRNA, before being cultured to mid-blastula stage 8, at which point animal caps were explanted. These caps were cultured until stage-matched control embryos reached early gastrula stage 10.5, when iFGFR signalling was induced for 3 hours. This experimental

design allows FGF signalling to be activated in neuralised animal caps, via the BMP inhibitor Noggin, enabling the investigation of the unique roles of FGFR1 and FGFR4 in regulating gene transcription during neural development (Brunsdon and Isaacs 2020). Contrasting with triplicate data in the CSKA-FGF4 study (King 2019), RNA-Seq analysis of neuralised caps with a single replicate was undertaken (Brunsdon and Isaacs 2020). Expression is measured in fragments per kilobase of transcript per million mapped reads (FPKM), which represents the relative expression of a transcript proportional to the number of cDNA fragments which originate from it. These normalised values prevent biases towards longer genes and enable comparisons between uninduced and induced samples to calculate fold change in expression.

3.2.5.1 A large number of genes are affected by iFGFR1 and iFGFR4 induction in neuralised animal cap explants

In order to visualise the data points and choose appropriate filtering conditions, an unfiltered scatterplot was generated, comparing the \log_2 FPKM of fragments in uninduced iFGFR1 and uninduced iFGFR4 caps (Figure 20A). 0.0001 was added to all FPKM values, in order to visualise fragments with FPKM expression values of 0, because $\log_2(0)$ is undefined. The majority of data points are situated along $y=x$ showing no fold expression change between the two uninduced control samples. This is expected because although the iFGFRs are expressed, they are inactive and therefore should not increase FGF signalling and alter gene expression. However, there is variability in the expression of transcripts with low FPKM values, towards the bottom left of Figure 20A, with many showing a ≥ 2 -fold increase or decrease in expression, which are present in the red and blue domains respectively. Therefore, a strict threshold is required to filter out transcripts with low FPKM values and minimise false positives.

3.2.5.2 High stringency filtering of iFGFR1/4 gene expression data

Transcripts are classed as differentially expressed if they satisfy $\text{FPKM} \geq 20$ for one or both of the uninduced or induced samples, and if they exhibit an expression change of ≥ 2 -fold. Of the 9,195 genes which satisfy $\text{FPKM} \geq 20$ in either or both uninduced samples, only 4 and 5 genes showed at least a two-fold increase or decrease in expression respectively (Figure 20B). This shows that the expression of transcripts in the absence of AP20187 in uninduced iFGFR1 and uninduced iFGFR4 caps were very similar. 257 and 108 transcripts were identified as being up and down regulated from iFGFR1 activation (Figure 20C; Table S27, S28), and 368 and 186 transcripts as up and down regulated from iFGFR4 activation (Figure 20D; Table S29, S30), when compared with respective uninduced controls. This shows that there is a considerable number of genes altered in response to increased FGF signalling

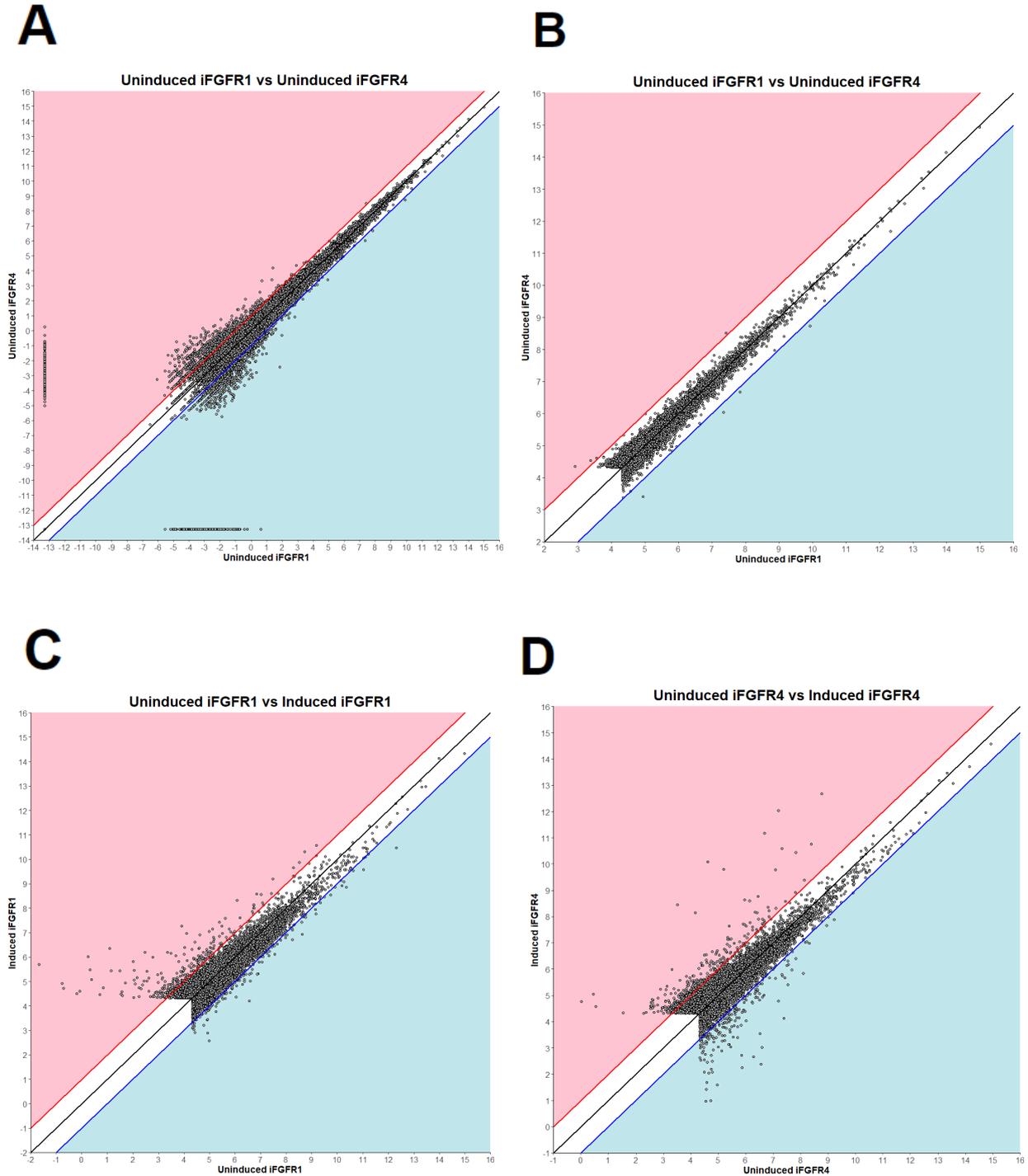


Figure 20: Ratio of gene expression, as \log_2 FPKM values, in iFGFR single replicate RNA-Seq data in neuralised animal cap explants. (A) unfiltered uninduced iFGFR1 plotted against uninduced iFGFR4. (B) filtered uninduced iFGFR1 plotted against uninduced iFGFR4. (C) filtered uninduced iFGFR1 plotted against induced iFGFR1. (D) filtered uninduced iFGFR4 plotted against induced iFGFR4. Data was filtered according to $\text{FPKM} \geq 20$. The black line is $y=x$. Data points in the red area (to the left of $y=x+1$, red line) are transcripts which exhibit a ≥ 2 fold up regulation by iFGFR induction and those in the blue area (to the right of $y=x-1$, blue line) a ≥ 2 fold down regulation.

through iFGFR1 and iFGFR4, and there are differences between the genes regulated by iFGFR1 and iFGFR4.

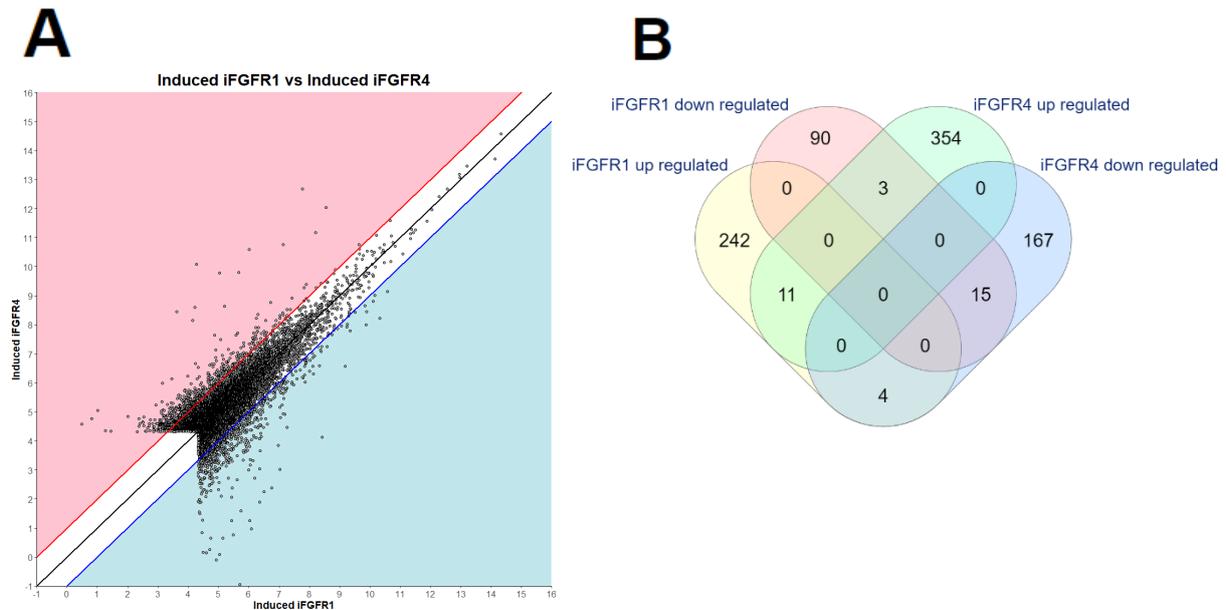


Figure 21: iFGFR1 and iFGFR4 affect distinct groups of transcripts in neuralised animal cap explants. (A) filtered induced iFGFR1 plotted against induced iFGFR4. Data was filtered according to FPKM ≥ 20 . The black line is $y=x$. Data points in the red area (to the left of $y=x+1$, red line) are transcripts which exhibit a ≥ 2 fold up regulation by iFGFR induction and those in the blue area (to the right of $y=x-1$, blue line) a ≥ 2 fold down regulation. **(B)** overlap of transcript lists which satisfy the high stringency criteria (FPKM ≥ 20 , fold change ≥ 2), in uninduced and induced groups. Venn diagram was generated using Multiple List Comparator (<http://www.molbiotools.com/listcompare.html>).

It is encouraging to see many characterised FGF target genes (Branney et al. 2009) in these high stringency filtered gene lists, for example negative regulators of FGF signalling up regulated by iFGFR signalling. *spry1*, *spry2*, *dusp1*, *dusp5* and *dusp6* are up regulated by iFGFR1 signalling (Table S27), while only *spry1* and *dusp22* are present in the genes up regulated by iFGFR4 signaling (Table S29). This suggests that FGFR1 could play a larger role in FGF signalling feedback inhibition than FGFR4. *tubb2b* (*n-tubulin*) (fold change = 0.46) and *otx2* were down regulated by iFGFR4 signalling. Both iFGFRs down regulate genes encoding subunits of RNA polymerase II and III, for example *polr2l.1* (iFGFR1; Table S28), and *polr3gl* and *polr2k* (iFGFR4; Table S30).

3.2.5.3 iFGFR1 and iFGFR4 affect the expression of overlapping groups of genes

In order to compare the transcriptomes regulated by iFGFR1 and iFGFR4 signalling, a scatterplot was generated to analyse the expression of transcripts in embryos subject to increased FGF signalling through these inducible receptors. The expression of 1300 transcripts show a greater than 2-fold difference between iFGFR1 and iFGFR4 induced embryos (Figure 21A), indicating that at this stage of development, the transcripts affected by increased FGF signalling through iFGFR1 and iFGFR4 are different.

Table 10: Overlap of transcripts consistently regulated by iFGFR1 and iFGFR4 signalling in neuralised animal cap explants. Transcript lists satisfied the high stringency filtering criteria (FPKM \geq 20, fold change \geq 2). Transcripts are distinguished by their “Align to source” reference code. List analysis was performed using Multiple List Comparator (<http://www.molbiotools.com/listcompare.html>).

iFGFR1 up regulated iFGFR4 up regulated		iFGFR1 down regulated iFGFR4 down regulated	
Gene name	Align to source	Gene name	Align to source
dact1	(c.Taira201203egg_X006008)	arg1	(c.Audic201207_X034514)
insm1	(c.Taira201203egg_X002601)	arpc3	(c.Quigley201112_X005662)
lmbd2	(c.Quigley201212_X023738)	fam55d	(c.Ismailoglu201203_X007979)
mcmbp	(c.Amin201106_X029621)	gnb3	(c.Quigley201112_X013169)
morn2	(c.Chung201110_X007967)	hesx1	(c.JGIL6RMv1_XeXenL6RMv10033507m, c.UniGene_XI_S13589749)
mrrf	(c.TeperekTkacz201206_X001915)	krt12	(c.Ueno201210st35_X000016, c.XenBase_27696404)
not-b	(c.Quigley201212_X014344)	lrat	(c.Taira201203eye_X009949)
plk3	(c.Taira201203kidney_X014170)	mdk	(c.Chung201110_X002883)
ppp1r3c.2	(c.TXGP201107_X005057)	pkdcc.2	(c.Ueno201210brain_X000869)
sgk1	(c.Taira201203ovary_X003188)	rpl27a	(c.Chang2013_X035887)
spry1	(c.Taira201203st09_X003581)	tuba1a-b	(c.mgEST_1013155827)
		Unnamed	(c.Chang2013_X033037)
		Unnamed	(c.Taira201203st08_X004257)

Comparisons of these filtered transcript lists from iFGFR1 and iFGFR4 signalling, according to the high stringency criteria (Table 6), revealed transcripts common to both transcriptomes (Figure 21B; Table 10, 11). This identified genes which are consistently regulated, such as *tuba1a-b*, which encodes alpha tubulin and is present in iFGFR1 and iFGFR4 down regulated gene lists (Table 10). Contrastingly, other genes are differentially regulated by iFGFR1 and iFGFR4 signalling, such as *id3*, which encodes a transcriptional regulator and is up regulated by iFGFR1 and down regulated by iFGFR4 (Table 11). Furthermore, some genes were regulated exclusively by either iFGFR1 or iFGFR4, including iFGFR1 up regulating *egr1*, *fos*, *junb*, *lefty-a*, *notch3* and *smurf2*. Interestingly, iFGFR1 and iFGFR4 signalling significantly altered the expression of different members of the same gene families. For example, *dusp1*, *dusp5*, *dusp6* and *hes1* were up regulated by iFGFR1, while iFGFR4 up regulated *dusp22* and *hes2*. This suggests that these genes are uniquely regulated by either FGFR1 or FGFR4 in early *Xenopus* development.

In order to investigate the probability of the observed overlaps, Python was used to randomly sample sets of the number of up regulated transcripts (257 by iFGFR1 and 368 by iFGFR4), which represent genes positively regulated by FGF signalling in normal development, from numbers between 1 and 35,532 (number of transcripts in the RNA-Seq). The highest number of overlaps in 10,000 iterations was 10 so the probability of getting an overlap of size 11 or greater is $p < 0.0001$. Therefore, an overlap of 11 is statistically significant (Table 10; Figure S6A; Table S31). iFGFR1 and iFGFR4 down regulated 108 and 186 transcripts respectively

Table 11: Overlap of transcripts differentially regulated by iFGFR1 and iFGFR4 signalling in neuralised animal cap explants. Transcript lists satisfied the high stringency filtering criteria (FPKM \geq 20, fold change \geq 2). Transcripts are distinguished by their “Align to source” reference code. List analysis was performed using Multiple List Comparator (<http://www.molbiotools.com/listcompare.html>).

iFGFR1 up regulated iFGFR4 down regulated		iFGFR1 down regulated iFGFR4 up regulated	
Gene name	Align to source	Gene name	Align to source
crabp2	(c.Taira201203egg_X003235)	appl1	(c.TeperekTkacz201206_X006093)
dynll1-a	(c.Audic201207_X053894)	ift172	(c.Chung201110_X004537)
id3	(c.Audic201207_X054642)	slc12a3	(c.TeperekTkacz201206_X004009)
Unnamed	(c.mgEST_1013119916)		

and these numbers were randomly sampled. The highest number of overlaps in 10,000 iterations was 5 so the probability of getting an overlap of size 6 or greater is $p < 0.0001$. Consequently, an overlap of 15 significantly down regulated transcripts is statistically significant (Table 10; Figure S6B; Table S31). Comparisons between transcripts up regulated by iFGFR1 but down regulated by iFGFR4, and vice versa, do not exhibit a significant overlap (Table 11; Figure S6C, S6D; Table S31). In summary, iFGFR1 and iFGFR4 regulate groups of transcripts with a statistically significant overlap, suggesting FGFR1 and FGFR4 regulate overlapping groups of transcripts in normal development.

3.2.5.4 Differences in gene ontologies of iFGFR1 and iFGFR4 regulated genes in neuralised animal cap explants

Genes up regulated by iFGFR1 signalling were enriched 27.5x and 24.6x above expected values for genes involved in inactivation of MAPK signalling (FDR = 1.78×10^{-2}) and negative regulation of FGFR signalling (FDR = 2.21×10^{-2}) respectively (Figure 22A, Table S32). Enrichment of genes involved in splicing in genes down regulated by iFGFR1 signalling is a recurring theme (Figure 22B, Table S33). This could be mediated by *eif4a3*, which is down regulated by iFGFR1 and encodes an ATP-dependent RNA helicase and is present in the exon junction complex at splice junctions on mRNAs (Karimi et al. 2018). Genes up regulated by iFGFR4 signalling showed enrichment for regulation of protein kinase activity, feedback loops, cell cycle and metabolism (Figure 23A, Table S34), however the enrichment is lower than iFGFR1. Inner ear morphogenesis, neurogenesis, tube, and migration biological processes for genes down regulated by iFGFR4 signalling (Figure 23B, Table S35).

iFGFR1 and iFGFR4 up regulated gene lists both showed enrichment of genes involved in the stress response, which could be due to increased mRNA translation, protein levels and FGF signalling as a result of this iFGFR approach. Aside from the expected enrichment in functions associated with FGF signalling, this analysis reveals distinct biological processes of

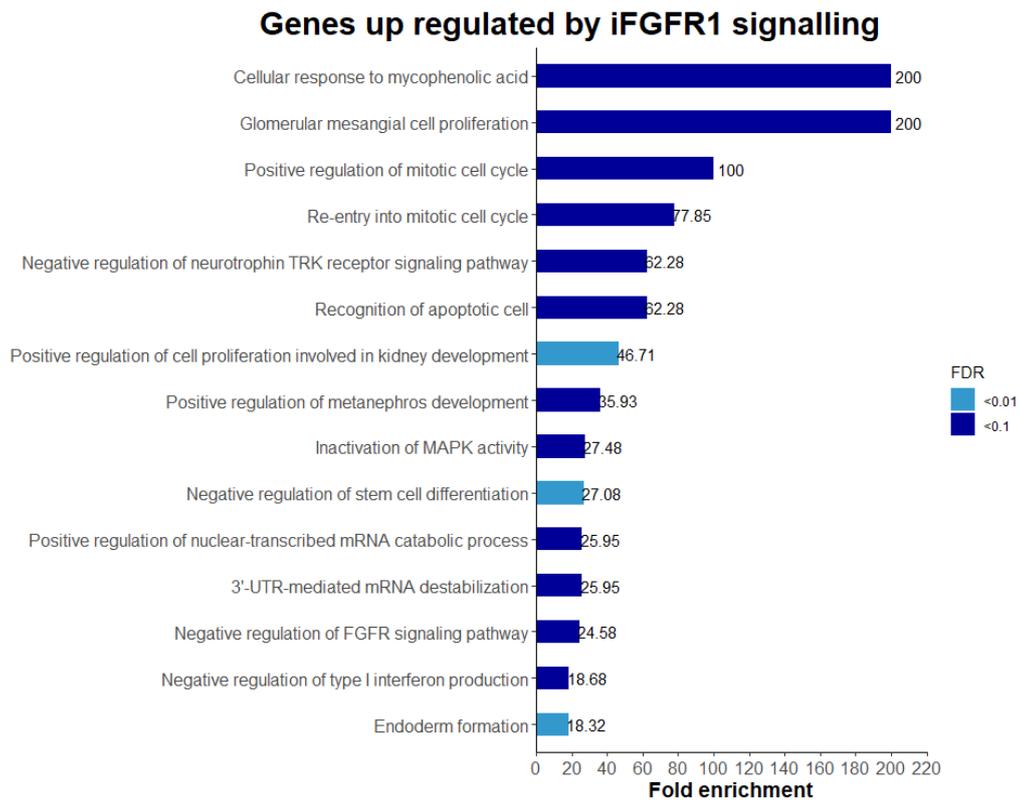
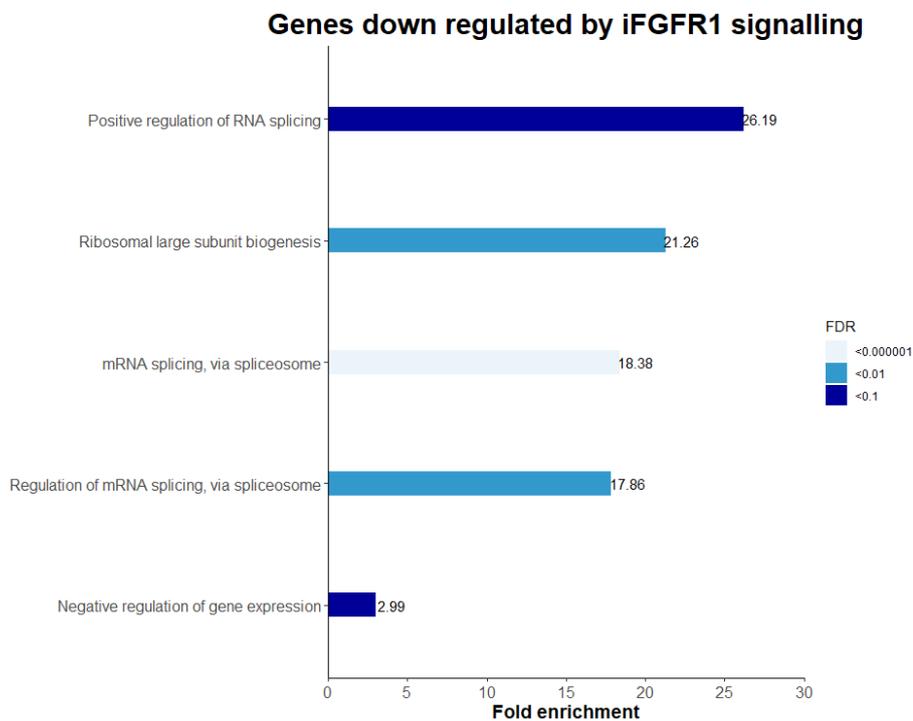
A**B**

Figure 22: The enrichment of biological processes associated with up (A) and down (B) regulated genes by iFGFR1 signalling in neuralised animal cap explants. Numbers to the right of bars represent fold enrichment and bar colours represent the FDR. The 15 biological processes with the highest fold enrichment are shown, with the complete list present in the corresponding table.

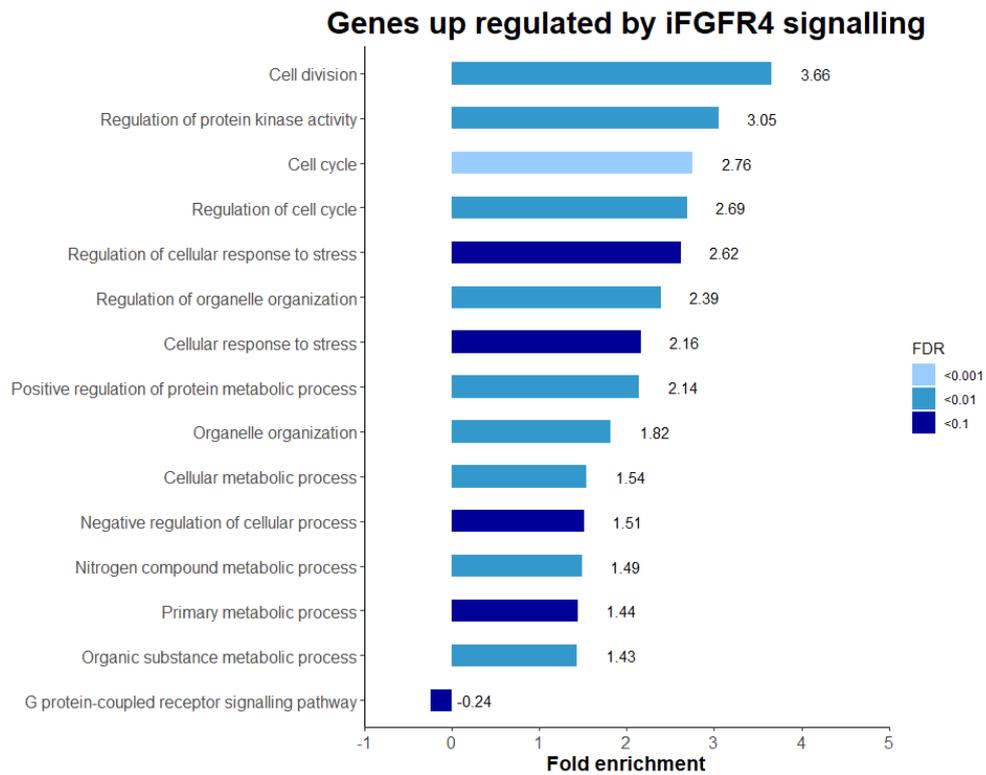
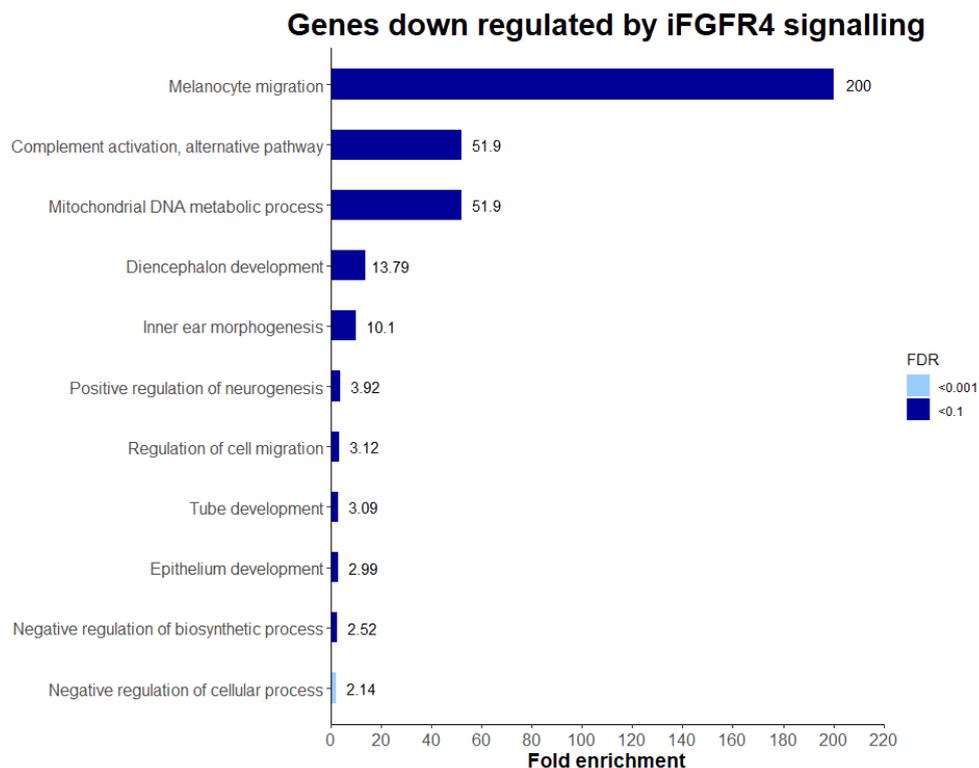
A**B**

Figure 23: The enrichment of biological processes associated with up (A) and down (B) regulated genes by iFGFR4 signalling in neuralised animal cap explants. Numbers to the right of bars represent fold enrichment and bar colours represent the FDR. The 15 biological processes with the highest fold enrichment are shown, with the complete list present in the corresponding table.

differentially expressed genes by iFGFR1 and iFGFR4 signalling suggesting that FGFR1 and FGFR4 have unique roles in development.

3.2.5.5 Similarities between PPI networks of genes affected by iFGFR1/4 signalling in neuralised animal cap explants

In the PPI network generated from genes up regulated by iFGFR1 signalling, there are interactions between Spry1, Spry2 and Dusp6, in keeping with their position as previously characterised FGF target genes and regulating MAPK signalling in a negative feedback loop (Branney et al. 2009). There are also interactions between Fos, Fosl1, Egr1, Junb and Myc transcription factors (Figure 24A).

There are similarities between the PPI networks of up regulated genes from iFGFR1 and iFGFR4 signalling, which both contain significantly more PPI interactions than expected (PPI enrichment p-value = 7.26×10^{-5} (iFGFR1), 1.53×10^{-4} (iFGFR4)). Both networks contain small GTPases as central nodes, for example Rhob (Figure 24A) and Hras (Figure 25A). There are also groups of interacting ubiquitin ligases, including Smurf2 (Figure 24A) and Pja2 (Figure 25A).

In accordance with the enrichment of splicing processes in GO analysis (Figure 22B), the ATP-dependent RNA helicase Eif4a3 interacts with a multitude of proteins. These interactions are present in the iFGFR1 down regulated PPI network (Figure 24B), which contains significantly more interactions than expected (1.55×10^{-15}).

Similarly, the ribosomal protein Rps6 is a central node in the iFGFR4 down regulated gene PPI network (Figure 25B), which is not significantly enriched for interactions (0.061). Taken together, PPI networks constructed from up regulated gene lists contain interactions within transcription factors and small GTPases, whereas down regulated PPI networks contain interactions between proteins involved in post transcriptional processes, such as splicing and translation.

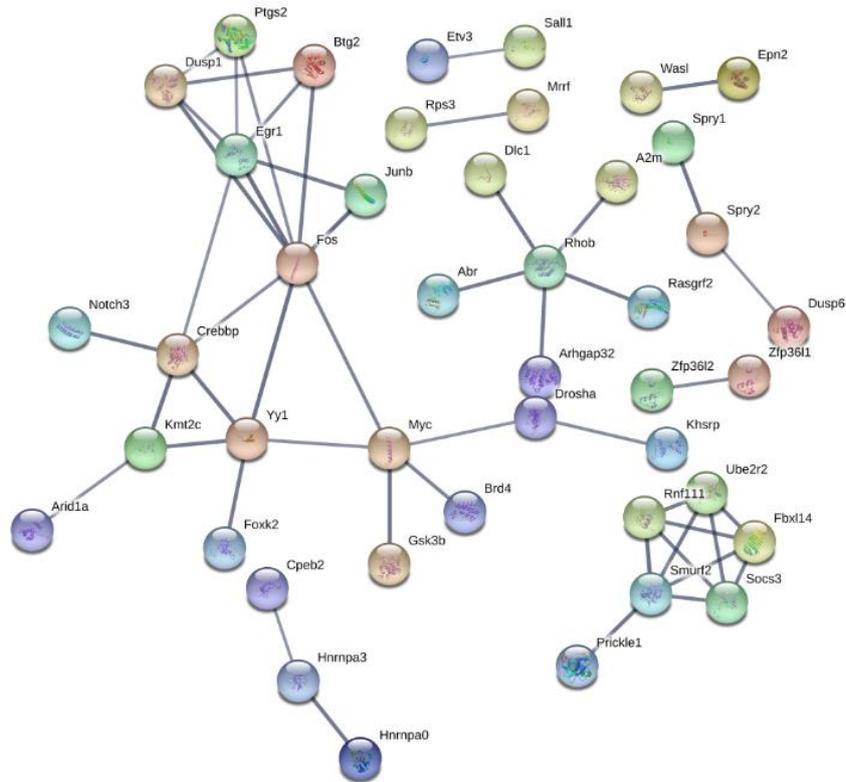
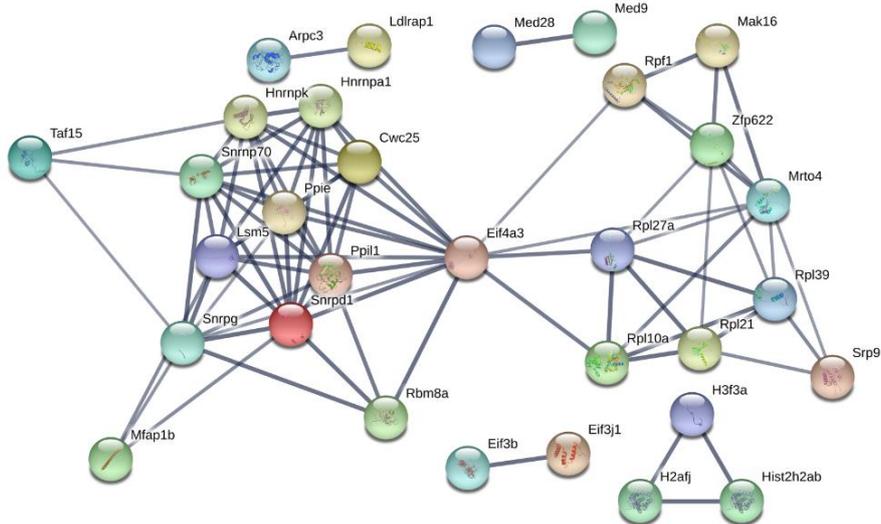
A**B**

Figure 24: Protein-protein interaction networks for genes up (A) and down (B) regulated by iFGFR1 signalling in neuralised animal cap explants. The interactions shown have passed the high confidence minimum required interaction score of 0.7 and the disconnected nodes have been removed.

3.2.6 CSKA-FGF4 and iFGFR1/4 RNA-Seq data set comparison

3.2.6.1 Identification of well-supported FGF target genes, regulated by FGF4, FGFR1 and FGFR4 signalling

Comparisons between differentially expressed genes in CSKA-FGF4 and iFGFR1/4 data sets will enable the identification of well-supported FGF target genes, regulated by FGF4, FGFR1 and FGFR4 signalling. These data sets both employ RNA-Seq technology, however they utilise highly related but different *Xenopus* species, namely diploid *X. tropicalis* and tetraploid *X. laevis*. Although these species contain similar genes, there may be differences in the gene annotation and the number of the gene copies. For these reasons, lists of gene names were compared between data sets to determine which genes are present in both analyses. Genes were counted once regardless of the number of transcripts identified. This revealed 12,398 shared genes present in both data sets (Figure S7). High stringency filtered gene lists were then overlapped (Table 12) to identify genes, which are present in both data sets and whose expression is affected by FGF signalling in both data sets.

Python was used to investigate the statistical significance of the observed overlaps between high stringency filtered gene lists from FGF4 and iFGFR1/4 data sets (Table S36; Figure S8). The overlap of genes up regulated by both FGF4 overexpression and iFGFR1 signalling was the only statistically significant overlap (Table 12; Table S36; Figure S8A), therefore genes in this list are likely to be biologically relevant, in the context of regulation by FGF signalling. *apold1* is present in this list and ontological analysis reveals that it is involved in the regulation of endothelial cell differentiation (Karimi et al. 2018), which is a mesoderm-derived tissue. This overlap also contains known FGF target genes *egr1*, *fos*, and *spry2* (Branney et al. 2009), whose expression will be analysed in CRISPR/Cas9 embryos to validate successful FGF signalling inhibition.

Three genes are consistently regulated by all three treatments, namely *pkdcc.2* (down regulated), *sgk1* and *plk3* (up regulated). These genes all encode serine/threonine kinases, which are involved in stress responses (Karimi et al. 2018) and therefore are likely not relevant to FGF signalling.

3.2.6.2 Heatmap comparisons

Comparisons between CSKA-FGF4 and iFGFR1/4 RNA-Seq data will visualise patterns of gene expression. Filtering these data sets according to low stringency criteria (Table 6) reveals larger gene lists (Table S37, S38, S39, S40, S41, S42). Up and down regulated gene lists were compiled into differentially expressed gene lists, containing 258 genes for FGF4, 453 for

Table 12: Overlap of CSKA-FGF4 and iFGFR1/4 gene lists, which satisfied high stringency filtering criteria. Numbers in brackets define the number of genes present in each filtered gene list. Light grey box colour represents a statistically significant overlap between gene lists. White box colour represents an insignificant overlap between gene lists. List analysis was performed using Multiple List Comparator (<http://www.molbiotools.com/listcompare.html>).

	iFGFR1 up regulated (137)	iFGFR1 down regulated (68)	iFGFR4 up regulated (250)	iFGFR4 down regulated (117)
FGF4 up regulated (53)	apold1, egr1, fos, plk3, sgk1, spry2	-	nuak2, plk3, sgk1	fth1
FGF4 down regulated (57)	notch3	pkdcc.2	-	cygb, eppk1, grebl1, pkdcc.2

iFGFR1 and 759 for iFGFR4. These were overlapped (Figure S9) and results are stated in Table S19.

The expression of genes in iFGFR1 and iFGFR4 experiments can be compared quantitatively because they were performed simultaneously using the same *Xenopus* species and experimental rationale. The normalised expression of overlapping genes, using \log_{10} of FPKM values, is visualised in Figure 26. When there are multiple transcripts of the same gene present with the same direction of expression change (increase or decrease relative to controls), the transcript with the highest expression in the control was selected, as this will likely be the most biologically relevant.

Genes present in clusters with known FGF targets may be biologically relevant in the context of FGF signalling since they exhibit a highly similar expression pattern to known FGF target genes, suggesting a similar degree of regulation. For example, *spry1* is clustered with *abhydrolase domain containing 15 (abhd15)*, which has hydrolase activity (Karimi et al. 2018). *dusp6* is clustered with *nemo-like kinase (nlk)* (Figure 26B). This MAPK is activated by CamKII in the noncanonical Wnt signalling pathway to antagonise canonical Wnt signaling (Ishitani et al. 2003). *abhd15* and *nlk* are not present in the high stringency filtered gene lists but these putative FGF target genes are both positively regulated by iFGFR1/4 signalling.

Although these RNA-Seqs use the same method of transcriptomic analysis, they differ in their protocols, developmental stages and organisms, namely CSKA-FGF4 utilised *X. tropicalis* and iFGFR1/4 utilised *X. laevis*. Therefore, the expression of transcripts cannot be compared quantitatively in heatmaps, even using transcript expression relative to respective controls. Instead, the expression of these overlapping genes can be compared qualitatively in heatmaps, by scoring genes +1 if their expression increases (red) and -1 if their expression decreases (blue) relative to controls.

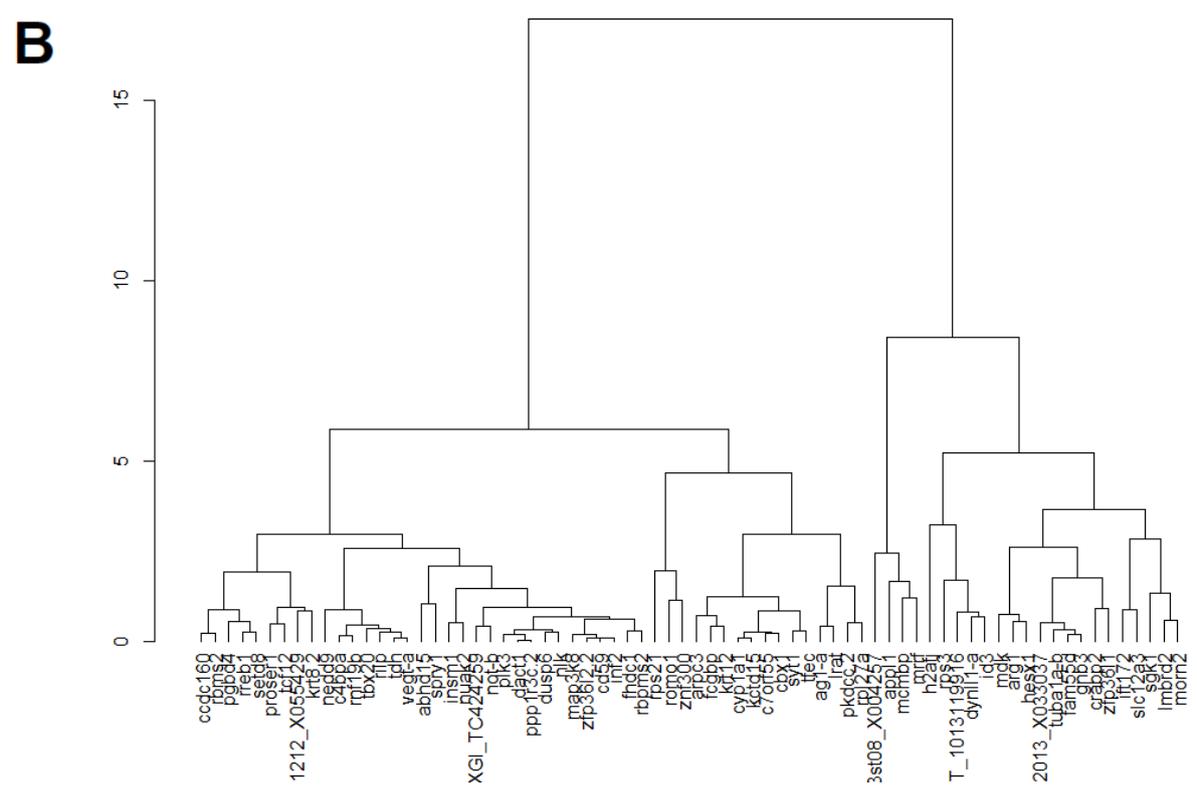
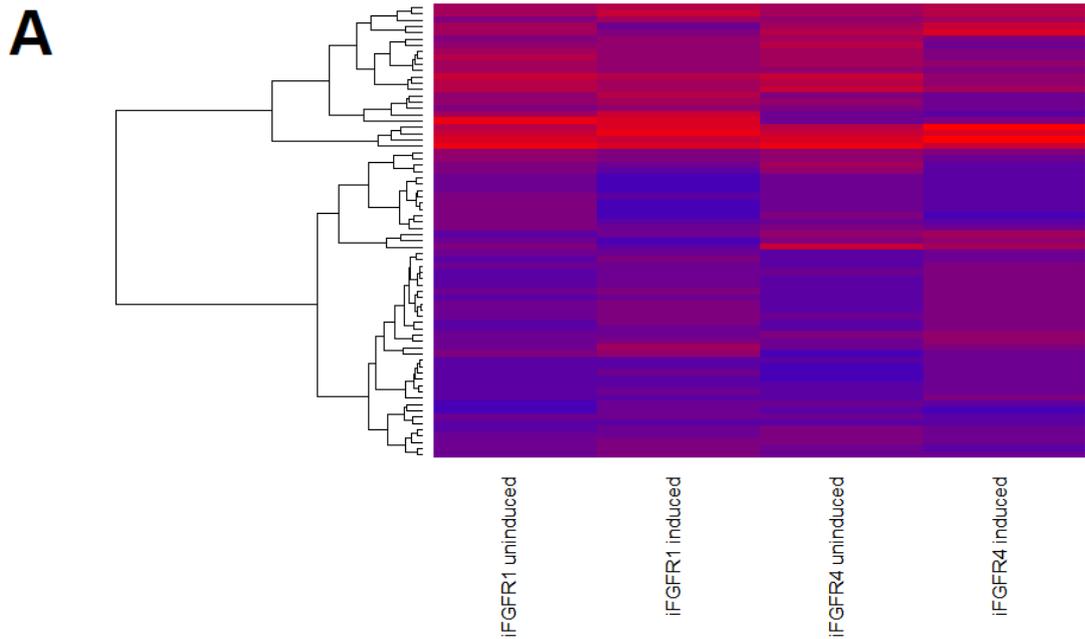


Figure 26: Expression of genes affected by iFGFR1 and iFGFR4 signalling. (A) Colours represent increases (red) and decreases (blue) in expression. Genes satisfied low stringency filtering criteria and are Euclidean clustered. **(B)** Clustering output with gene names. 0 means highly related clusters.

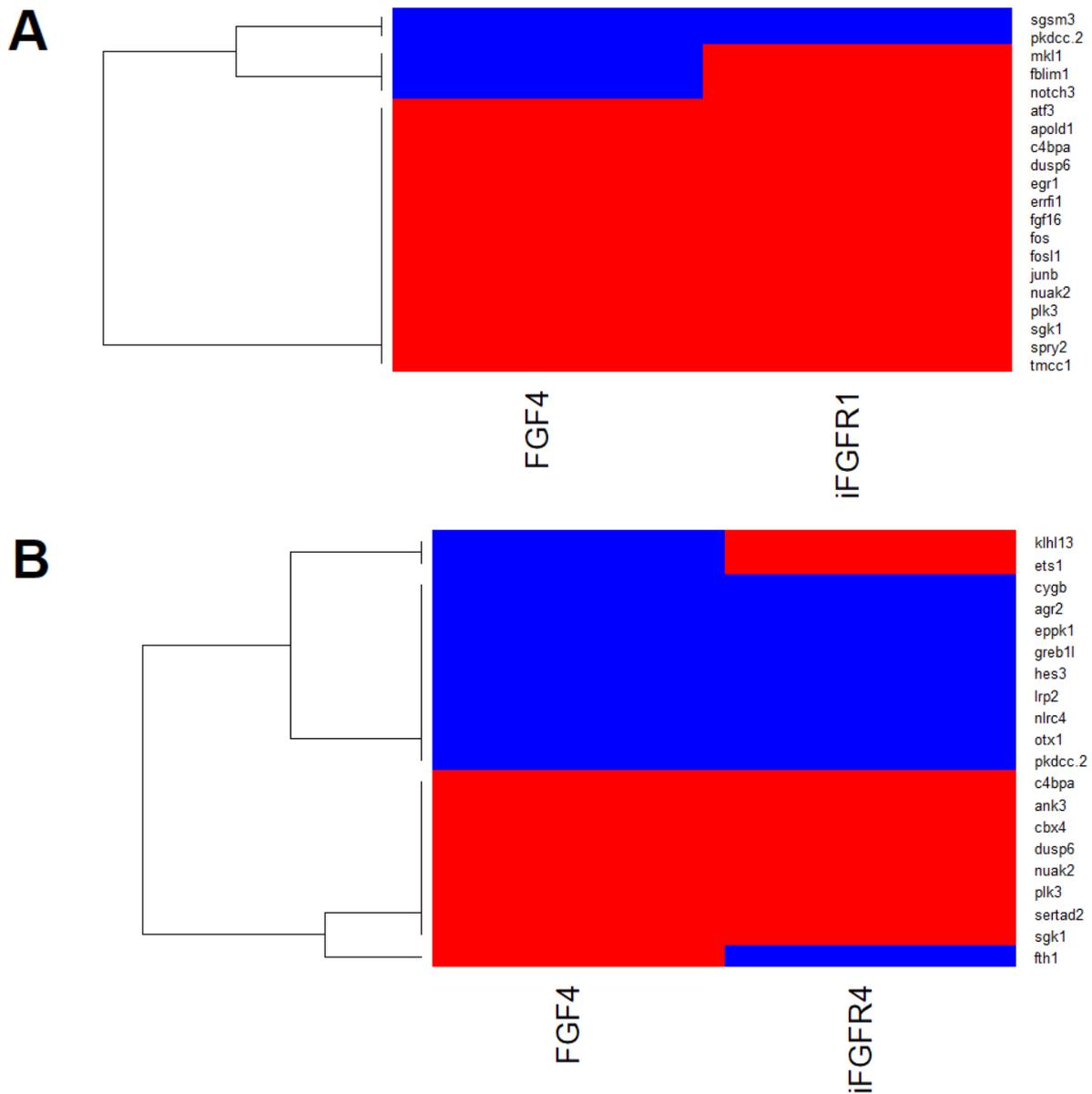


Figure 27: Expression of genes affected by FGF4 and iFGFR1 (A), and FGF4 and iFGFR4 signalling (B). The expression of overlapping genes is scored +1 if their expression increases (red) and -1 if their expression decreases (blue) relative to controls. Genes satisfied low stringency filtering criteria and are Euclidean clustered.

There are differences in the expression of three genes when comparing FGF4 with iFGFR1 (Figure 27A) and iFGFR4 (Figure 27B). The expression of *mlk1* (transcriptional co activator), *fbim1* (filamin binding protein) and *notch3* (member of the notch receptor family) are all decreased by FGF4 overexpression but increased by iFGFR1 signalling, suggesting differing regulation by FGF4 and FGFR1. *notch3* is of particular importance as notch receptors function in notch-delta signalling in lateral inhibition to regulate the transcription of proneural genes (Hori et al. 2013).

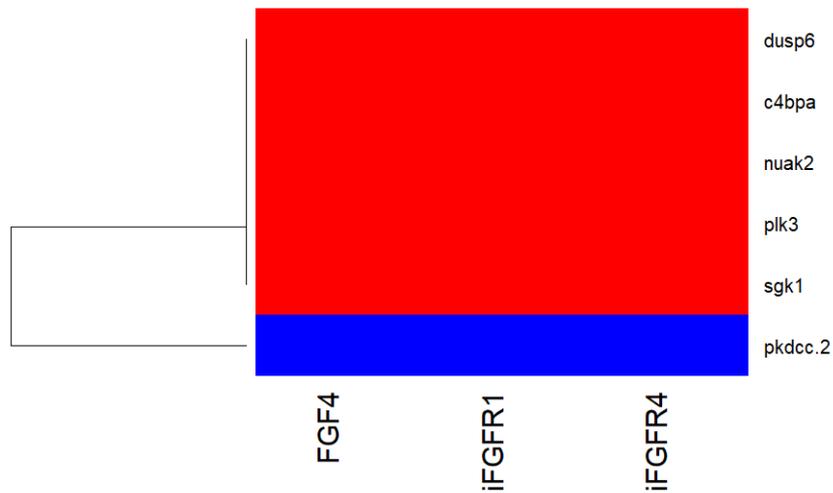


Figure 28: Expression of genes affected by FGF4, iFGFR1 and iFGFR4 signalling. The expression of overlapping genes is scored +1 if their expression increases (red) and -1 if their expression decreases (blue) relative to controls. Genes satisfied low stringency filtering criteria and are Euclidean clustered.

The expression of many transcription factors was consistently up regulated by FGF4 and iFGFR1 signalling, including *egr1*, *fos*, *fosl1* and *junb*. There is also evidence for feedback loops, including *fgf16* (to increase FGF signalling) *dusp6* and *spry2* (to decrease FGF signalling) (Figure 27A).

FGF4 overexpression increases *fth1* expression, which is involved in cellular ion homeostasis (Karimi et al. 2018), but is decreased by iFGFR4 signalling (Figure 27B). This is the opposite for *ets1* transcription factor and ubiquitination associated *klhl13*, which are increased by iFGFR4 but decreased by FGF4 overexpression. Consistently up regulated genes include chromatin-binding *cbx4* and predicted serine/threonine kinase *nuak2*.

Genes overlapping in FGF4, iFGFR1 and iFGFR4 data sets all exhibit a consistent change in expression (Figure 28), suggesting that these genes are similarly regulated by FGF4, FGFR1 and FGFR4 signalling.

3.2.6.3 RNA-Seq analysis summary

Comparisons between high stringency filtered gene lists revealed well-characterised FGF target genes *egr1*, *fos*, and *spry2*, which were significantly up regulated in embryos subject to increased FGF4 or iFGFR1 signalling. Their expression will be analysed in CRISPR/Cas9 embryos to validate successful FGF signalling inhibition and to determine potential unique roles of different FGFRs in regulating their expression. Low stringency filtering enabled the identification of *abhd15* and *nlk*, whose expression closely resembled *spry1* and *dusp6* respectively in iFGFR embryos, however further experiments are required to investigate this.

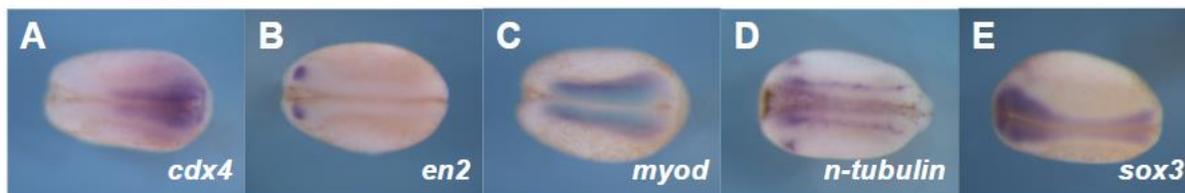


Figure 29: The spatial expression pattern of *cdx4* (A), *en2* (B), *myod* (C), *n-tubulin* (D) and *sox3* (E) in mid-neurula stage 17-18 *X. tropicalis* embryos. Dorsal view, anterior left. Embryos are staged according to Nieuwkoop and Faber (1994) stages of *Xenopus* development.

3.2.7 In situ hybridisations of FGF target genes

This investigation initially had aimed to utilise *in situ* hybridisation and RT-PCR experiments to determine the contributions of different FGFRs in regulating the expression of mesodermal and neuronal cell development markers, and well-supported and putative FGF target genes in embryos. However, the pandemic led to the lab closure before I was able to perform these experiments in CRISPR/Cas9 embryos to investigate the unique roles of FGFR1, FGFR4 and FGFR1 in regulating the expression of these genes. The expression of previously characterised FGF target genes *cdx4*, *en2*, *myod* and *sox3*, marker of neural development *n-tubulin* and putative FGF target gene *rasl11b* are visualised by *in situ* hybridisations, as markers of mesodermal and neural tissues (Isaacs et al. 1994; Lamb and Harland 1995; Pownall et al. 1996; Hardcastle et al. 2000; Fisher et al. 2002; Zhao et al. 2006; Isaacs et al. 2007; Faas and Isaacs 2009; Nentwich et al. 2009; Yamagishi and Okamoto 2010; Cowell 2019; Maude 2019). This is a preliminary demonstration of the ability to use this as an assay as marker analysis in embryos.

Neurula stage embryos show expression of homeodomain transcription factors *cdx4* and *en2* in the paraxial mesoderm and posterior neural tube (Figure 29A), and presumptive midbrain-hindbrain boundary (Figure 29B). Myogenic helix-loop-helix transcription factor *myod* is an FGF-regulated mesodermal tissue marker (Figure 29C). *n-tubulin* is a marker of differentiating neurons and expressed throughout the nervous system in early *Xenopus* development, particularly in differentiating neurons in neurogenesis (Figure 29D). HMG-box transcription factor *sox3* marks the neural plate (Figure 29E).

rasl11b (Ras-like protein, family 11, member B) was identified as a putative FGF target due to being significantly upregulated in the CSKA-FGF4 RNA-Seq data set (Cowell 2019; Maude 2019), whereby embryos were subject to increased FGF signalling from CSKA-FGF4 plasmid injection (King 2019). The expression of this gene, encoding a poorly characterized GTPase with a high degree of similarity to Ras proteins, closely resembles that of *fgf8* in early *X. tropicalis* development (Maude 2019). *rasl11b* expression increased in FGF4 injected

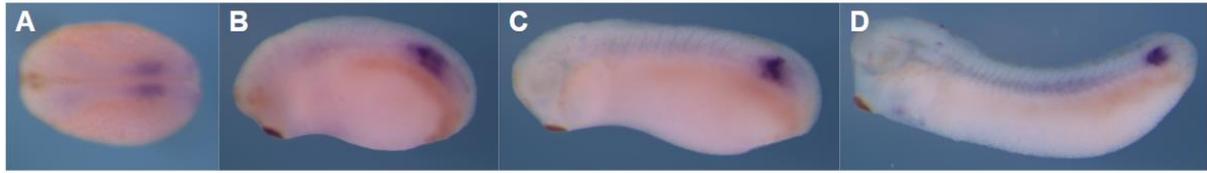


Figure 30: The spatial and temporal expression pattern of *rasl11b* using *in situ* hybridisation in *X. tropicalis* embryos. (A) Mid-neurula stage 17, view anterior left. (B) Stage 24, lateral view. (C) Stage 26, lateral view. (D) Stage 28, lateral view. Embryos are staged according to Nieuwkoop and Faber (1994) stages of *Xenopus* development.

embryos, when validating this RNA-Seq data set (Cowell 2019). Stage series of *in situ* hybridisations for *rasl11b* reveals its spatial and temporal expression in the presomitic mesoderm throughout early *X. tropicalis* development (Figure 30).

Analysis of these FGF-regulated tissue markers could also suggest unique contributions of different receptors in regulating their expression, and potentially the development of the tissues they are expressed in.

3.8 Summary

Within data sets, there was a significant overlap between transcripts consistently up and down regulated by iFGFR1 and iFGFR4 and gene probes consistently up and down regulated by iFGFR1 and iFGFR2, and dnFGFR1 and dnFGFR4. RNA-Seq data set comparison revealed that only the genes up regulated by both FGF4 and iFGFR1 exhibited a significant overlap. Whereas between the microarray data sets, there was a significant overlap between probes up regulated by iFGFR1 and iFGFR2 signalling with those which increased and decreased as a result of FGF signalling inhibition by dnFGFR4.

egr1, *fos* and *spry2* have been identified in the presented meta-analysis of previously generated high-throughput transcriptomic data sets (Branney et al. 2009; King 2019; Brunson and Isaacs 2020). The expression of these well-supported and previously characterised FGF target genes will be analysed in *fgfr* knockout embryos, as a result of CRISPR/Cas9 targeting, to validate successful FGF signalling inhibition. This analysis, along with the genes whose expression differed by either iFGFR1 or iFGFR4, could highlight unique roles of individual FGFRs in regulating the expression of FGF target genes.

3.3 Discussion

3.3.1 Unique roles of individual FGFRs

There is evidence that individual FGFRs have different ligand preferences and degrees of activation of downstream signalling pathways. The complex relationship between ligands and receptors has been observed in neural development, whereby marker gene expression is controlled by ligand competition between FGFRs with different potencies, namely FGFR1 and FGFR4 (Yamagishi and Okamoto 2010). This idea of ligand-receptor preference has been demonstrated in the RNA-Seq data set comparison, which revealed that only the genes up regulated by both FGF4 and iFGFR1 exhibited a significant overlap. This is in accordance with studies showing FGF4 has receptor preference for FGFR1c and FGFR2c (Itoh and Ornitz 2004; Zhang et al. 2006). However, it should be noted the CSKA-FGF4 plasmid-based approach results in mosaic expression throughout the embryo (Pownall et al. 1996).

FGFR1 and FGFR2 activate MAPK more strongly than FGFR4 (Umbhauer et al. 2000; Brunson and Isaacs 2020). The differential activation of downstream pathways by different FGFRs therefore suggests that activation of different FGFRs would result in different gene expression profiles. Contrasting this idea, data presented in this chapter and the initial publication of dnFGFR data observed only a small number of genes whose expression differed when FGFR1 and FGFR4 were targeted using dnFGFRs. The authors concluded that FGFR1 and FGFR4 regulate the same sets of genes from the commencement of FGF signalling at the MBT to the time of analysis at early gastrula (Branney et al. 2009). However, there is concern that dnFGFR overexpression could lead to the construct promiscuously forming non-productive dimers with and inhibiting other FGFRs than its specific FGFR (Ueno et al. 1992). This could explain the considerable overlap in genes regulated by FGFR1 and FGFR4, when inhibited through their respective dnFGFR (Figure 7). Nevertheless, FGF signalling inhibition by dnFGFRs has been confirmed by a reduction in the level of dpERK and the expression known FGF target genes, namely *tbxt*, *cdx4* and *myod* (Branney et al. 2009). Consequently, although dnFGFRs remain a useful method to understand the role of FGF signalling broadly, it is likely that they do not inhibit signalling through individual FGFRs. Therefore, caution should be exhibited with their use in understanding the unique roles of FGFRs.

Similar to dnFGFR analysis, iFGFR1 and iFGFR4 transcript expression profiles exhibited a statistically significant overlap from Python simulations (Figure S6; Table S31), suggesting that FGFR1 and FGFR4 regulate the expression of overlapping groups of transcripts in normal development. Despite this statistically significant overlap, there is a large number of transcripts

whose expression differed greater than two-fold between iFGFR1 and iFGFR4 induced embryos (Figure 21). This suggests that FGFR1 and FGFR4 regulate the expression of different genes in normal development and individual genes are discussed in Section 3.3.3.2. This conclusion is in accordance with that of the initial publication, which used a higher threshold for filtering (> 30 FPKM) and only employed analysis using scatterplots and Venn diagram, without statistical simulation (Brunsdon and Isaacs 2020). Furthermore, the initial publication was unable to find a greater involvement of FGFR4 in neural induction (Brunsdon and Isaacs 2020), which had been previously investigated using dnFGFRs (Hongo et al. 1999; Hardcastle et al. 2000).

Data presented in this investigation revealed a significant overlap in the probes affected by iFGFR1 and iFGFR2 signalling (Figure S2; Table S13). Furthermore, there was a small number of genes whose expression differed two-fold or greater between iFGFR1 and iFGFR2 induced embryos (Figure 11). Taken together, this suggests that FGFR1 and FGFR2 regulate the expression of overlapping groups of genes in late gastrula embryos and therefore, this conclusion is consistent with that of the initial publication (Brunsdon and Isaacs 2020). However, the analyses presented in this investigation expanded on that of the initial publication by exploiting Python statistical analysis, STRING PPI networks and comprehensive PANTHER GO enrichment analysis. Furthermore, data sets were subject to multiple filtering criteria for different subsequent analysis, facilitating extensive comparisons between data sets.

3.3.2 Unexpected contributions of FGFRs to biological processes

Since FGF positively regulates cell proliferation (Figure 2), which requires increased gene transcription, it would be expected that genes encoding subunits RNA polymerases would be up regulated. Contrastingly, iFGFR1 and iFGFR4 down regulate genes encoding subunits of RNA polymerase II and III, for example *polr2l.1* (iFGFR1; Table S28), and *polr3gl* and *polr2k* (iFGFR4; Table S30). However, downregulation of these subunits may not decrease gene transcription globally but instead selectively alter the gene expression of the cell. This idea is supported by the common theme of FGF signalling up regulating the transcription of genes whose protein products are transcription factors and interact with each other, for example the up regulation of *etv3*, *egr1*, *fos* and *junb* by increased iFGFR1 signalling (Table S27). This could result in the selective alteration of transcription, which is likely to be dynamic and linked to the developmental stage, for example in the CSKA-FGF4 data set at neurula stage 14 to selectively increase the expression of neural genes.

Neural tissue markers *otx2* and *tubb2b* (*n-tubulin*) were both down regulated by iFGFR4 signalling (Table S30), while *tubb2b* (*n-tubulin*) was up regulated in dnFGFR4 (Table S3) and down regulated in iFGFR1 embryos (Table S10). The negative regulation of these genes in this meta-analysis contradicts previous research which suggests that these genes are positively regulated by FGF signalling (Lamb and Harland 1995; Pera et al. 2003; Fletcher et al. 2006) and that FGFR4 is more heavily involved in neural development than FGFR1 (Hardcastle et al. 2000). The experimental design of the dnFGFR data set could explain this result, since embryos were collected at early gastrula stage 10.5 for microarray analysis, with the aim of identifying FGF targets activated immediately after the commencement of FGF signalling at the MBT and the different contributions of FGFR1 and FGFR4 in this (Branney et al. 2009). However, the RNA-Seq and microarray iFGFR experiments aimed to investigate gene transcription in neurula stages (Brunsdon and Isaacs 2020) and therefore the negative regulation of these neural tissue genes was unexpected.

Of the previously characterised FGF target genes, whose expression in wild type neurula stage embryos was presented, only *cdx4* (decreased by dnFGFR1/4), *n-tubulin* (increased by dnFGFR4) and *rasl11b* (increased by FGF4 overexpression) were identified as being differentially expressed in this meta-analysis. This low occurrence of significant changes in gene expression of these FGF target genes is surprising. Furthermore, there were no significant differences in expression in the iFGFR RNA-Seq and microarray data, which could be the result of only performing one biological replicate. Nevertheless, *rasl11b*, encoding a member of the Ras-like subfamily in the small monomeric GTPase Ras protein family, expression increased in FGF4 injected embryos, when validating this RNA-Seq data set (Cowell 2019). This provides an avenue for further investigation, due to its overlapping expression with FGFs in *X. tropicalis* embryos (Cowell 2019), along with its putative role as a putative tumour suppressor (Louro et al. 2004; He et al. 2018) and negative regulator of MAPK signalling (Emerson et al. 2017).

3.3.3 Target gene selection

3.3.3.1 Well-supported FGF target genes

egr1, *fos* and *spry2* were consistently positively regulated by FGF signalling in RNA-Seq and microarray data sets. *egr1* encodes a zinc finger transcription factor, which is expressed in the dorsal marginal zone (Nentwich et al. 2009). *egr1* was identified in iFGFR1, iFGFR2 and dnFGFR4 microarrays and FGF4 and iFGFR1 RNA-Seq data sets as being positively regulated by FGF signalling. It is a direct target of FGF signalling at gastrula stage (Branney et al. 2009) and its expression is also induced indirectly via phosphorylation and activation of

Elk1 by FGF signalling, which forms a complex with SRF to bind the *egr1* promoter (Nentwich et al. 2009). *egr1* differentially regulates the expression of known FGF target genes, for example by activating *myod* and repressing the expression of the *tbxt* transcription factor (Nentwich et al. 2009), which *egr1* is a target of (Saka et al. 2000).

egr1 and *fos* are immediate-early genes which can be transcribed within minutes of stimulation, for example by interaction of extracellular growth factors with cell-surface receptors, activating intracellular signalling pathways, including MAPK and ERK (Bahrami and Drabløs 2016). *fos* encodes a basic-leucine zipper transcription factor and was significantly up regulated in FGF4 and iFGFR1 RNA-Seq data sets. It forms AP-1 heterodimers with Jun proteins, for example AP-1, comprised of Fos and JunD, has been shown to mediate FGF and BMP signalling during *Xenopus* development (Lee et al. 2011). FGF2 activates AP-1 activity in *Xenopus* animal caps and AP-1 is implicated in FGF regulated mesoderm induction, due to its ability to induce *tbxt* expression and the inhibition of AP-1-induced mesoderm induction in the presence of a dominant negative form of *tbxt* (Kim et al. 1998).

fos and *spry2* are expressed in regions of FGF activity with dorsal enrichment during gastrula stages (Cowell 2019; Sivak et al. 2005). *spry2* was identified in dnFGFR1, dnFGFR4 and iFGFR2 microarrays and FGF4 and iFGFR1 RNA-Seq data sets. Spry2 is a member of the Sprouty protein family, which modulate FGF signalling intracellularly by inhibiting the PLCγ pathway (Sivak et al. 2005). Spry2 also negatively regulates MAPK signalling via suppression of ERK activation in the ventral marginal zone (Hanafusa et al. 2009).

The expression of these well-supported and previously characterised FGF target genes will be analysed in *fgfr* knockout embryos, as a result of CRISPR/Cas9 targeting, to confirm successful FGF signalling inhibition. This analysis could also highlight unique roles of individual FGFRs in regulating the expression of these FGF target genes.

3.3.3.2 Putative genes uniquely regulated by FGFR1 or FGFR4

A number of genes were identified as being exclusively regulated by iFGFR1 or iFGFR4 in the RNA-Seq data set. iFGFR1 up regulated well-supported FGF target genes *egr1* and *fos* (Section 3.3.3.1), suggesting unique regulation of their expression by FGFR1 in early *Xenopus* development. *junb* encodes Jun transcription factor, which forms AP-1 heterodimers with Fos proteins (Kim et al. 1998; Lee et al. 2011), and was also up regulated by iFGFR1 signalling, along with *notch3*. The increased expression of *notch3* in this data set, conducted on neuralised animal caps, is logical as notch receptors function in notch-delta signalling in lateral inhibition to regulate the transcription of proneural genes (Hori et al. 2013).

iFGFR1 and iFGFR4 regulated the expression of different members of *dusp* and *hes* gene families, namely *dusp1*, *dusp5*, *dusp6* and *hes1*, compared to *dusp22* and *hes2* respectively. *dusp5* is activated by FGF signalling and functions as a negative regulator inhibiting FGF signalling (Branney et al. 2009). Oscillations in the expression of *hes1*, a notch effector molecule, were induced by FGF2 (Nakayama et al. 2008), which has receptor preference for FGFR1c (Itoh and Ornitz 2004; Zhang et al. 2006). This is in line with its increase in expression by iFGFR1 signalling.

These genes could be analysed in *fgfr* CRISPR/Cas9 knockout embryos to investigate the possible unique roles of individual FGFRs in regulating their expression.

3.3.4 Results caveats

Previous studies suggested dnFGFR4 is a more potent FGF signalling inhibitor than dnFGFR1 per mass of injected mRNA (Branney et al. 2009). This could explain the larger number of genes affected by dnFGFR4 than dnFGFR1 signalling inhibition and consequently larger number of biological processes in the PANTHER analysis. This is reminiscent of the stronger enrichment of biological processes in iFGFR1-regulated gene lists than iFGFR4, which could mean FGFR1 has stronger effects during development than FGFR4. Conversely, it could mean that iFGFR1 construct is more potent at activating FGF signalling, which is supported by a stronger ability of iFGFR1 to induce dpERK than iFGFR4 (Brunsdon and Isaacs 2020).

The iFGFR RNA-Seq and microarray data only contained a single replicate (Brunsdon and Isaacs 2020) and consequently there is variability in the expression of transcripts with low FPKM values (Figure 18A), which would likely have been reduced with three biological replicates. Therefore, strict filtering criteria was required to reduce the higher number of false positives. Nevertheless, without biological replication, the observed magnitude of expression can be analysed, but significance of a difference between a pair of samples cannot be assessed.

PANTHER and STRING analysis results should be interpreted with caution. Due to *Xenopus* species being a less widely used model organism, genes from high stringency filtered gene lists were examined using the better annotated *M. musculus* genome for these analyses. Therefore, there could be species-specific differences in the function and interactions of proteins. Furthermore, the microarray data sets contain more unnamed probes than the RNA-Seqs, which likely explains the increased occurrence of insignificant PANTHER and STRING results, rather than as a result of the targeted receptor or experimental design.

3.3.5 Future directions

The experimental design caveats discussed above could be rectified by performing an RNA-Seq experiment in *X. tropicalis*, with triplicate sibling embryo treatment groups, using dnFGFR4, iFGFRs, CSKA-FGF4 and CRISPR/Cas9, accompanied by respective controls. Ding et al. (2018) provides ideas of potential statistical analysis if all embryo treatments were carried out simultaneously, including principal component analysis (PCA) to quantitatively determine the relatedness of the gene expression profiles of embryos subjected to different FGF manipulation approaches. Furthermore, it would be possible to generate quantitative heatmaps to compare the magnitude of gene activation or inhibition as a result of different FGF signalling manipulation techniques, targeting different receptors.

Chapter 4: Protocol development to target FGFRs using CRISPR/Cas9

4.1 Introduction

In this chapter, a CRISPR/Cas9 protocol will be developed to investigate the roles of FGFR1, FGFR4 and FGFR1 in mediating FGF signalling in early *Xenopus* development. CRISPR/Cas9 has been selected due to its gene targeting specificity, which is lacking in other FGF manipulation approaches. As highlighted in Chapter 3, CSKA-FGF4 does not enable the investigation of a unique FGFR because individual ligands exhibit preferences for multiple receptors, stated in Table 1 (Itoh and Ornitz 2004; Zhang et al. 2006). Conversely, iFGFRs activate FGF signalling through specific receptors (Pownall et al. 2003; Brunsdon and Isaacs 2020), however FGF signalling inhibition approaches are crucial to understand the necessity of individual FGFR signalling in *Xenopus* development. Although dnFGFRs remain a useful method to understand the role of FGF signalling broadly, it is likely that they inhibit signalling through multiple FGFRs, exhibiting issues of promiscuity and specificity (Ueno et al. 1992). There is concern over the induction of innate immune responses in embryos subjected to MO-mediated inhibition (Gentsch et al. 2018; Paraiso et al. 2019). Finally, drug-inhibition of FGF signalling, by SU5402 application, inhibits the activation of receptor tyrosine kinases, including FGFRs, by blocking its tyrosine kinase activity (Mohammadi et al. 1998). However, SU5402 has non-specific effects on other receptor tyrosine kinases, such as FLT3, TRKA, FLT4 and JAK3 (Gudernova et al. 2016).

In order to perform genome editing using CRISPR/Cas9, primers are designed to generate a sgRNA, which is a fusion of the traditional CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) (Doudna and Charpentier 2014). The sgRNA forms a complex with and recruits the endonuclease Cas9 to a target DNA sequence, which is followed by a PAM site for recognition. Cas9 cleaves DNA, creating a double strand break, resulting in NHEJ (Doudna and Charpentier 2014), which can lead to indels and the introduction of a premature stop codon (method reviewed by (Nakayama et al. 2013)). This has been successfully implemented in *X. tropicalis* in a number of previous studies (Blitz et al. 2013; Nakayama et al. 2013; Bhattacharya et al. 2015; Aslan et al. 2017; McQueen and Pownall 2017; Sempou et al. 2018).

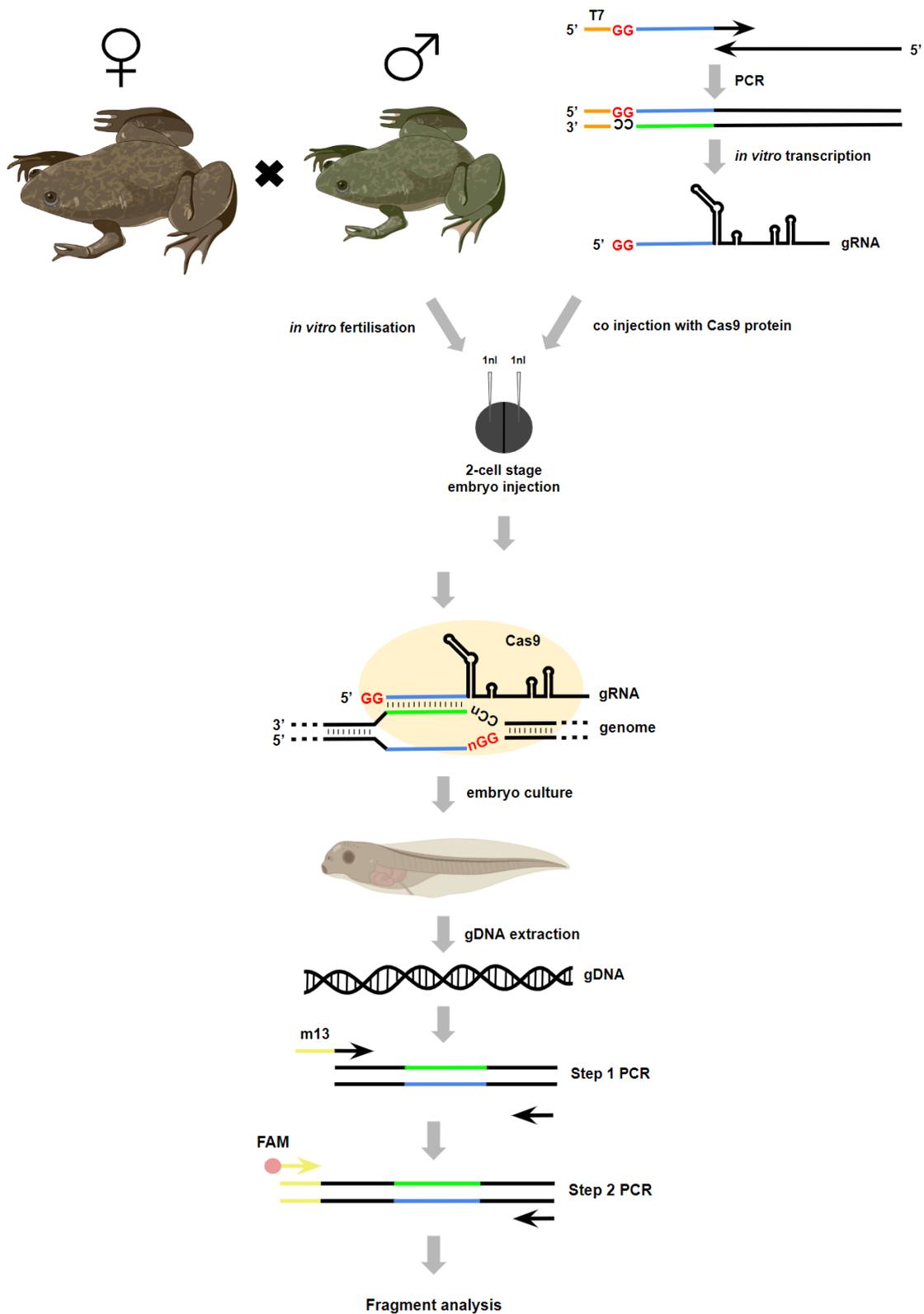


Figure 31: Preliminary CRISPR/Cas9 protocol. sgRNA generation adapted from (Nakayama et al. 2013) and fragment analysis adapted from (Bhattacharya et al. 2015). *Xenopus* adult and embryo illustrations were taken from BioRender.com. 5' primer contains T7 promoter (orange), guanine nucleotide (G) for transcriptional initiation using T7 RNA polymerase and genomic target sequence which is complementary to ~ 20 bp DNA sequence (blue). 3' universal primer partly overlaps with the 5' primer and contains the sgRNA backbone sequence, required for correct RNA folding for Cas9

interactions. sgRNA is generated by PCR followed by in vitro transcription using T7 RNA polymerase. The resulting sgRNA is coinjected with endonuclease Cas9 protein into 1-2 cell stage embryos, whereby the complex is recruited to the target site, via complementary base pairing, upstream of the PAM sequence. Cas9 cleaves the target site. Step 1 PCR involves genomic DNA extracted from injected embryos and a forward (F) primer with a 5'-m13 overhang and a reverse (R) primer which are complementary to the DNA sequence which has been targeted by CRISPR/Cas9. Step 2 PCR involves a common fluorescent m13 primer and the reverse (R) primer used in step 1 PCR, to generate fluorescently tagged PCR fragments, to be analysed using fragment analysis.

In *Xenopus*, it is favourable to coinject sgRNA and Cas9 protein, rather than Cas9 mRNA, which has been shown to be more toxic when injected and less efficacious in gene-editing (Bhattacharya et al. 2015). Furthermore, Cas9 protein elicits genome modifications earlier in development than injecting Cas9 mRNA, which would need to be translated in the embryo, delaying mutation generation (Bhattacharya et al. 2015). CRISPR/Cas9, administered by microinjection of sgRNA and Cas9 protein as a preformed complex into embryos, to knockout all *fgfrs* has been performed in *Xenopus* by a previous MSc student (Zilinskaite 2019).

However, CRISPR/Cas9 frequently produces mosaicism in the indels it generates (Blitz et al. 2013; Nakayama et al. 2013; Bhattacharya et al. 2015; McQueen and Pownall 2017; Lamas-Toranzo et al. 2019), due to the timing of targeting and the error-prone process of NHEJ. Mosaicism is the presence of more than two alleles in an individual and therefore reduces the generation of a knockout in one step without the need of F0 founder breeding (Lamas-Toranzo et al. 2019), which is impossible in the timescale of this project, and reduces the penetrance of the expected phenotype. For these reasons, it is crucial to accurately determine the efficiency of CRISPR/Cas9 sgRNAs in inducing mutations in their target genes. To do this, fragment analysis will be utilised, whereby primers, which amplify the regions targeted by CRISPR/Cas9 are used during a 2-step PCR using genomic DNA from injected embryos to generate fluorescently labelled DNA fragments (Bhattacharya et al. 2015). These labelled fragments are separated according to their size on a capillary electrophoresis gel, at single base resolution, enabling indel detection in embryos injected with CRISPR/Cas9 constructs (Schuelke 2000; Yang et al. 2015). Since fragment analysis is a novel approach to detect mutations, its results have been validated by Sanger sequencing (Yang et al. 2015; Zilinskaite 2019), a more widely used method to determine sequence changes.

In this chapter, *tyrosinase* will be targeted to determine optimal sgRNA concentration and CRISPR/Cas9 system efficiency. sgRNA design for *fgfrs* will be outlined, followed by fragment analysis to enable the selection of the sgRNA with the highest targeting efficiency for future experiments. The development of this genome-modification protocol will enable gene expression analysis of FGF target genes identified by a meta-analysis of high-throughput

transcriptomic data sets. This will reveal the unique roles of individual FGFRs in regulating the expression of FGF target genes.

The aims of this chapter are:

- Optimise sgRNA injection concentration using *tyrosinase* targeting experiments
- Design multiple sgRNAs for each *fgfr*, targeting different constitutive exons
- Perform preliminary experiments to analyse phenotypic defects in *fgfr* knockout embryos
- Predict the sgRNA with the highest targeting efficiency, based on indel frequency in fragment analysis.

4.2 Results

4.2.1 CRISPR/Cas9 system efficiency analysis by *tyrosinase* targeting

To examine CRISPR/Cas9 system efficiency, *tyrosinase* was selected to be targeted, whose protein product is the rate-limiting enzyme in melanin synthesis, which provides the pigmentation to eyes and skin. Therefore, disruption of this gene results in the observable phenotype of oculocutaneous albinism (Ando et al. 2007), when biallelic knockout occurs (Blitz et al. 2013). Furthermore, this gene has previously been disrupted using CRISPR/Cas9 in *X. tropicalis* (Blitz et al. 2013; Nakayama et al. 2013; Bhattacharya et al. 2015).

1-2 cell stage *X. tropicalis* embryos were injected with 1.5ng Cas9 protein and either 400pg or 600pg sgRNA, trialling different concentrations of sgRNA. By stage 42, 57.1% of embryos injected with 400pg sgRNA (n=7) resembled the uninjected (n=36) and Cas9 injected controls (n=19), which displayed wild type pigmentation in their retinal pigment epithelium (Figure 32). The remaining 42.9% exhibited speckled pigmentation (n=7), termed mosaic albinism, which implies mosaic *tyrosinase* gene targeting, resulting from Cas9 targeting one, neither or both of the alleles and the error-prone process of NHEJ (McQueen and Pownall 2017). Injections of 600pg sgRNA yielded higher albinism rates, whereby 36.7% and 23.3% displayed mosaic and complete albinism respectively, and the remaining 40% exhibited wild type pigmentation (n=30). This demonstrates successful tyrosinase targeting by CRISPR/Cas9 and the 600pg sgRNA concentration is more effective in yielding the expected mutant phenotype of complete albinism in the eye, however several neural crest derived melanocytes remain present (Figure 32A).

To quantitatively determine the efficiency of *tyrosinase* targeting, fragment analysis was performed. This approach identifies insertion or deletion mutations (indels) in the *tyrosinase* gene as a result of CRISPR/Cas9 targeting the first exon, which contains the EGF-like domain and Tyrosinase CuA-binding region (de Castro et al. 2006). Primers, amplifying the region targeted by CRISPR/Cas9, were used in a 2 step PCR reaction protocol using genomic DNA from embryos to generate fluorescently labelled DNA fragments (Figure 31). These labelled fragments are separated according to their size on a capillary electrophoresis gel, at single base resolution (Schuelke 2000; Yang et al. 2015). Fragments containing indels, resulting from CRISPR/Cas9 targeting, are visualised as additional peaks of fluorescence, along with those present in control embryos (Bhattacharya et al. 2015).

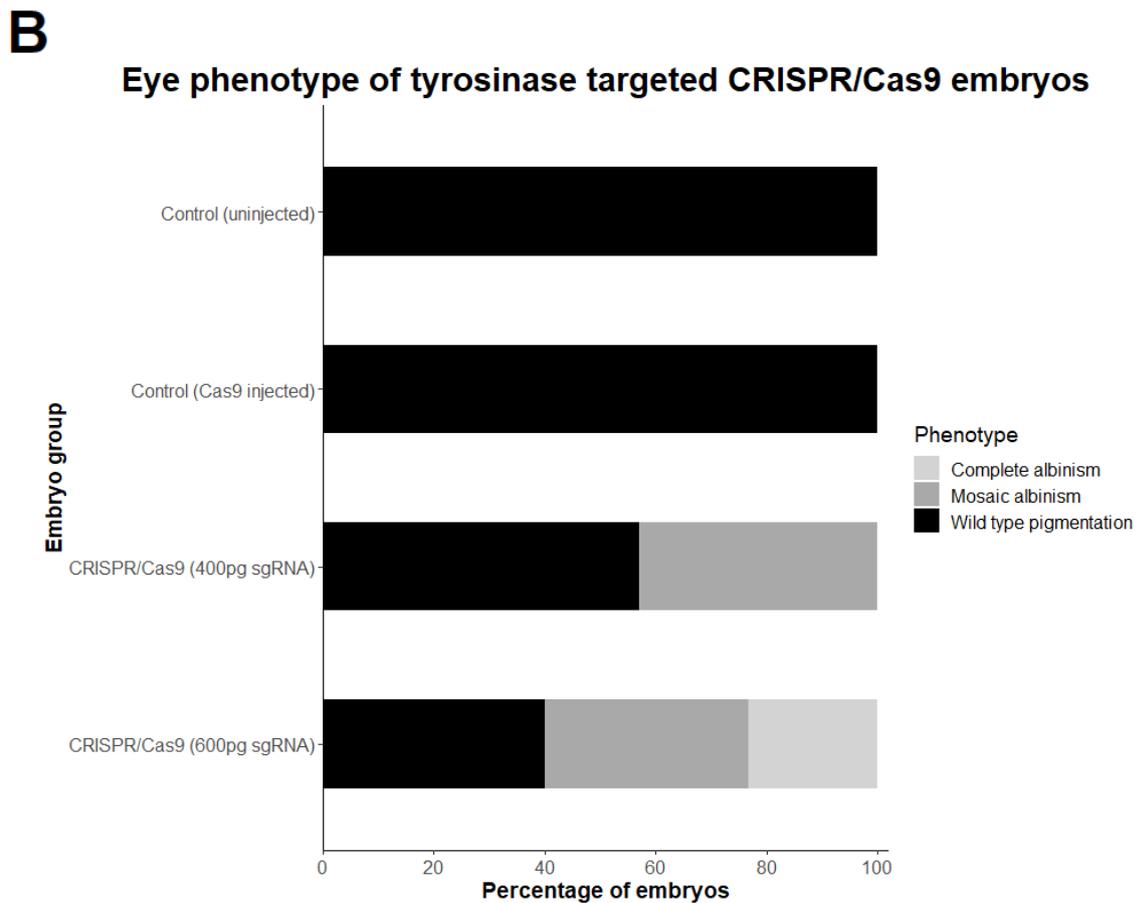
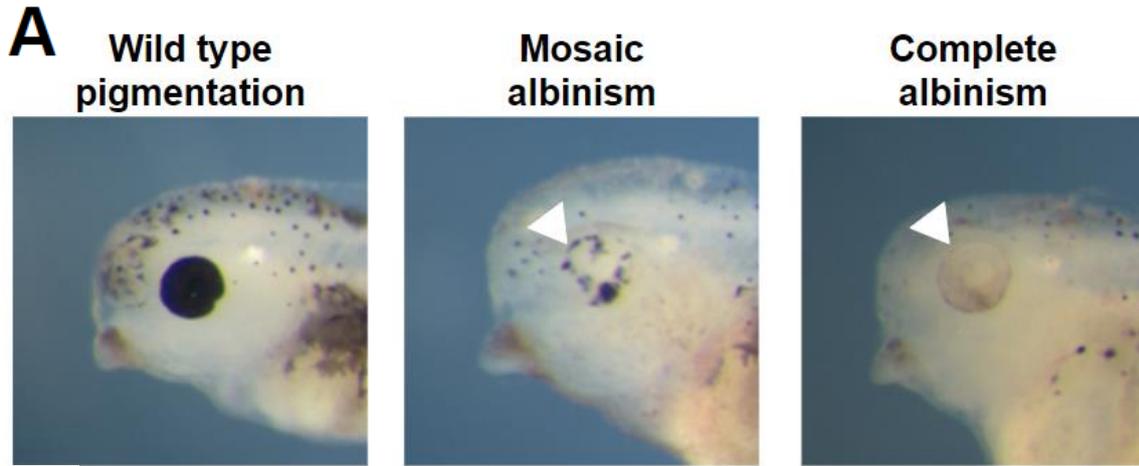


Figure 32: Phenotypic classification of embryos subject to *tyrosinase* gene targeting by CRISPR/Cas9. Stage 42 embryos are either uninjected (n=36) or injected at 1-2 cell stage with 1.5ng Cas9 (n=19) or 400pg (n=7) or 600pg (n=30) *tyrosinase* sgRNA and 1.5ng Cas9 in 2nl. **(A)** White arrow highlights reduced or absent pigmentation in the retinal pigment epithelium. Embryos are staged according to Nieuwkoop and Faber (1994) stages of *Xenopus* development. **(B)** Quantification of eye phenotypes, categorised by those displaying wild type pigmentation (black), mosaic albinism (dark grey) and complete albinism (light grey), corresponding to the classification (A).

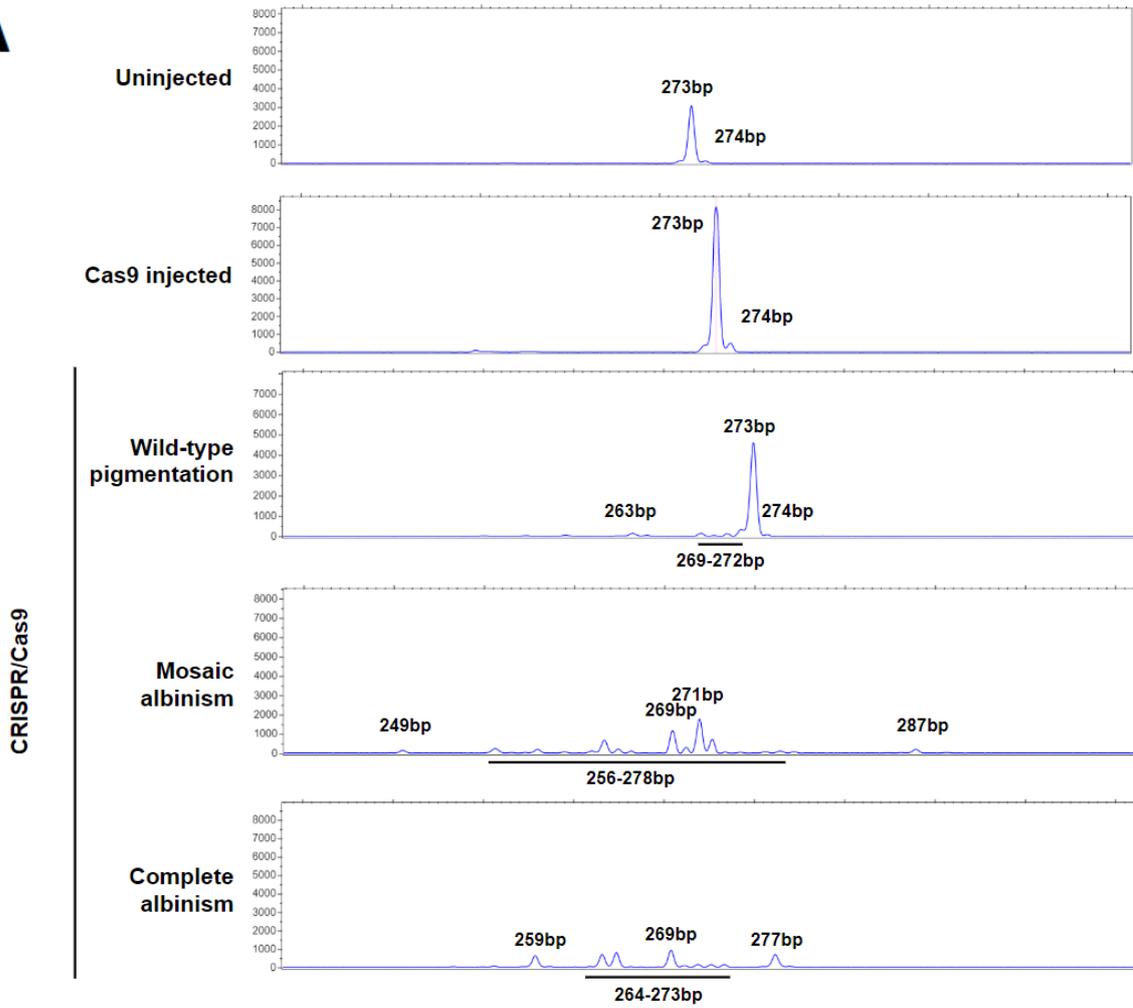
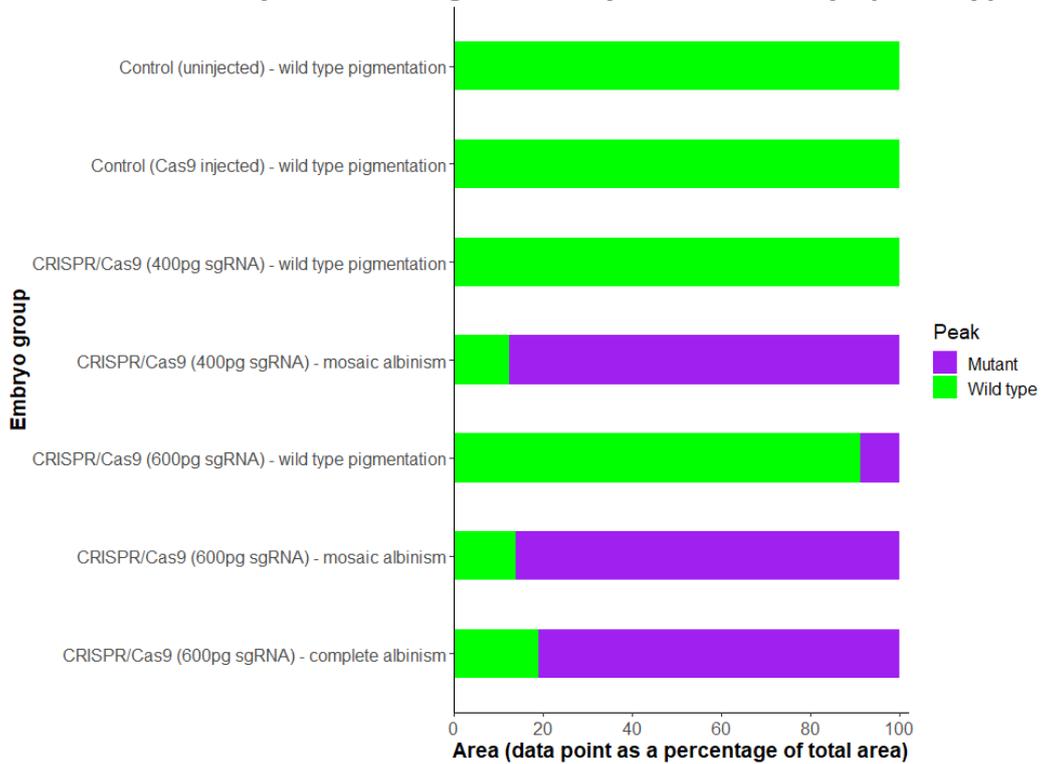
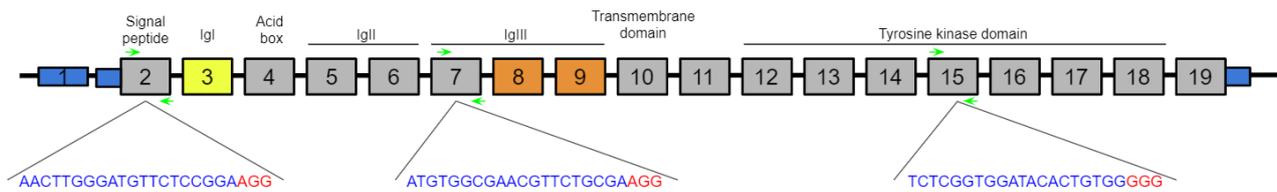
A**B Quantification of tyrosinase fragment analysis based on eye phenotype**

Figure 33: Fragment analysis of embryos subject to tyrosinase gene targeting by CRISPR/Cas9. Stage 42 embryos are either uninjected or injected at 1-2 cell stage with 1.5ng Cas9 or 400pg or 600pg *tyrosinase* sgRNA and 1.5ng Cas9 in 2nl. Embryos are staged according to Nieuwkoop and Faber (1994) stages of *Xenopus* development. **(A)** Fragment analysis of uninjected, Cas9 injected and 600pg sgRNA injected embryos displaying wild type pigmentation, mosaic albinism or complete albinism. Wild type peaks are seen at 273bp and 274bp and are present in all embryos. CRISPR/Cas9 embryos show peaks of varying size. Indels occurred in the range of 247bp-287bp. **(B)** Quantification of fragment analysis peaks, using peak area, based on data points, as a percentage of total area. Embryos were categorised by those displaying wild type pigmentation, mosaic albinism and complete albinism in the retinal pigment epithelium. Number of embryos in each group analysed are as follows: Control (uninjected) - wild type pigmentation (n=1), Control (Cas9 injected) - wild type pigmentation (n=2), CRISPR/Cas9 (400pg sgRNA) - wild type pigmentation (n=1), CRISPR/Cas9 (400pg sgRNA) - mosaic albinism (n=3), CRISPR/Cas9 (600pg sgRNA) - wild type pigmentation (n=5), CRISPR/Cas9 (600pg sgRNA) - mosaic albinism (n=3), and CRISPR/Cas9 (600pg sgRNA) - complete albinism (n=1).

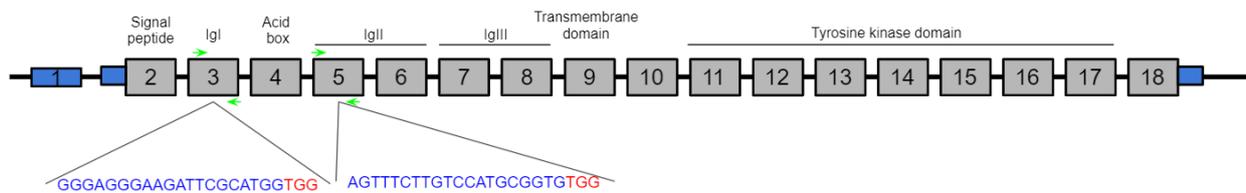
Uninjected and Cas9 injected control embryos showed two peaks at 273bp and 274bp (Figure 33A). These peaks were present in CRISPR/Cas9 targeted embryos displaying wild type pigmentation in the retinal pigment epithelium, along with the presence of fragments containing indels, between 263bp and 272bp in length. However, the wild type fragments were the most abundant. Contrastingly, in embryos displayed either mosaic or complete albinism, the abundance of these wild type fragments was diminished. Fragment analysis of embryos with mosaic pigmentation revealed fragments containing deletions, between 249bp and 272bp, and additionally fragments containing insertions, between 276bp and 287bp. Similarly, embryos displaying complete albinism in the retinal pigment epithelium contained fragments of 256bp to 272bp and a fragment at 277bp representing an insertion.

Fragment analysis results were quantified by measuring the area under each peak, to determine the relative abundance of wild type and mutant fragments, as a percentage of the total area under all the identified peaks (Figure 33B). Embryos were separated based on their treatment (control or CRISPR/Cas9) and phenotype of retinal pigment epithelium (wild type pigmentation, mosaic albinism and complete albinism). 100% of the peaks in uninjected (n=1) and Cas9 injected embryos were wild type (n=2). Similarly, 100% of peaks present in embryos injected with 400pg sgRNA displaying wild type pigmentation were wild type (n=1). However, in embryos with mosaic pigmentation, wild type peaks constituted 12.5% of the total peak area, with the remaining 87.5% being mutant peaks containing indels (n=3). This pattern is consistent with injections of 600pg sgRNA, whereby wild type peaks constituted 91.1% and 13.8% peak area in embryos displaying wild type pigmentation (n=5) and mosaic albinism (n=3) respectively. However, the percentage of wild type peaks in embryos displaying complete albinism is increased to 19% (n=1), relative to that in embryos with mosaic albinism.

FGFR1



FGFR4



FGFRL1

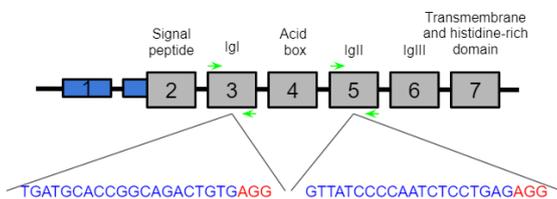


Figure 34: CRISPR/Cas9 targets in *fgfr1*, *fgfr4* and *fgfr11*. Blue boxes represent 5' and 3' UTR. Yellow boxes represent infrequent exon skipping. Orange boxes represent mutually exclusive exons. Blue nucleotides represent CRISPR/Cas9 target and red represents PAM site. Green arrows represent PCR primers.

Taken together, this demonstrates the mosaic nature of CRISPR/Cas9 as a result of the error-prone process of NHEJ and that deletions were more common than insertions. Furthermore, this demonstrates the success of CRISPR/Cas9 and fragment analysis protocols in generating and quantitatively analysing indel mutations in *tyrosinase* respectively and 600pg sgRNA concentration was selected for FGFR experiments, due to increased targeting efficiency from higher sgRNA concentrations (Guo et al. 2014).

4.2.2 CRISPR/Cas9 target selection in *fgfrs*

In order to generate *fgfr* knockouts using CRISPR/Cas9, the targeted sequence must be followed by a 5'-NGG-3' PAM site (where N is any nucleotide) in the genome for Cas9 recognition (Figure 34). sgRNA must be targeted to a sequence in a constitutive exon, which is unique to that receptor but present in all its isoforms. Attractive domains for this are the signal peptide and Ig domains, which constitute the extracellular regions of the FGFR and provide the FGF ligand binding specificity. Exons encoding these regions are in the 5' region of the gene, therefore maximising the chance of a truncated non-functional protein product being translated. Furthermore, the introduction of indels could cause nonsense-mediated

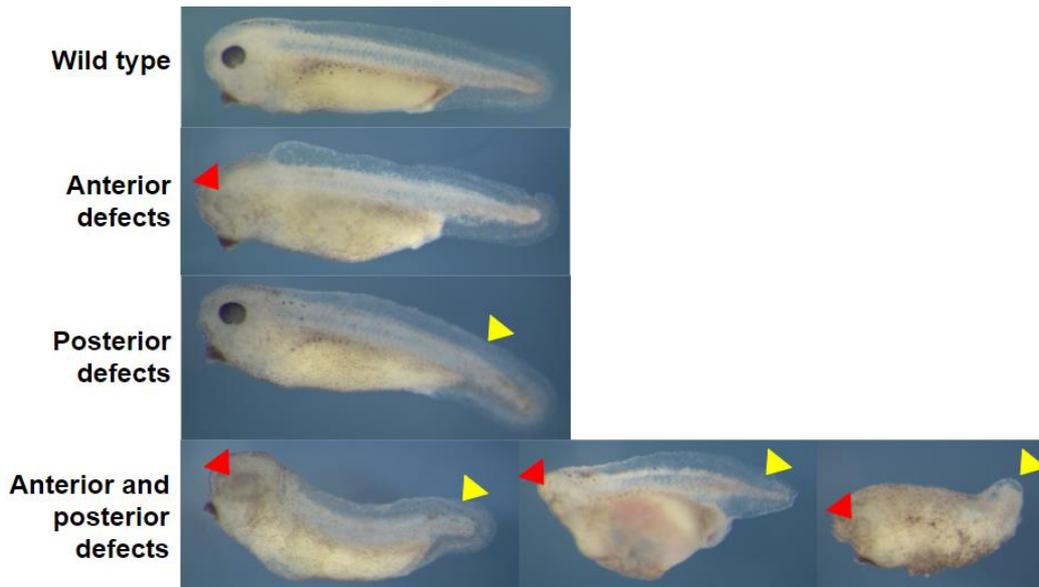


Figure 35: Phenotype of *X. tropicalis* subject to CRISPR/Cas9 targeting of *fgfrs*. Embryo classification. Anterior defects (red arrow) include the presence of oedemas and enlargement, reduction or absence of the head, ventral endodermal yolk mass or eyes, while posterior defects (yellow arrow) include posterior truncations and tail curvature. Embryos were injected with 600pg sgRNA and 1.5ng Cas9 (or 2nl water) at 1-2 cell stage and cultured to stages 37-40. Embryos are staged according to Nieuwkoop and Faber (1994) stages of *Xenopus* development.

decay (Popp and Maquat 2016). Suitable CRISPR/Cas9 target sites in these extracellular domains were found in *fgfr1* exon 2 and exon 7, and exon 3 and exon 5 in *fgfr4* and *fgfr11* (Figure 34). Furthermore, exon 15 of *fgfr1*, which encodes a proportion of the intracellular tyrosine kinase domain, contained a suitable CRISPR/Cas9 target site, however there were no such targets present in *fgfr4*. Multiple sgRNAs, targeting different regions, were designed for each gene to control for potential off-target effects.

4.2.3 Targeting *fgfrs* using CRISPR/Cas9

Preliminary experiments show the percentage of embryos displaying a wild type phenotype or with anterior and/or posterior defects, when injected at the 1-2 cell stage with sgRNA and Cas9 protein and collected for subsequent fragment analysis at stages 37-40 (Table S44). All CRISPR/Cas9 *fgfr* embryos displayed a similar range of phenotypes, allowing the use of this general classification approach. Anterior defects include the presence of oedemas and enlargement, reduction or absence of the head, ventral endodermal yolk mass or eyes, while posterior defects include posterior truncations and tail curvature. Figure 35 shows examples of embryo phenotypes classed as having anterior and/or posterior defects. Embryo phenotypes from individual experiments and the corresponding fragment analysis results are shown in Figure S10-S15.

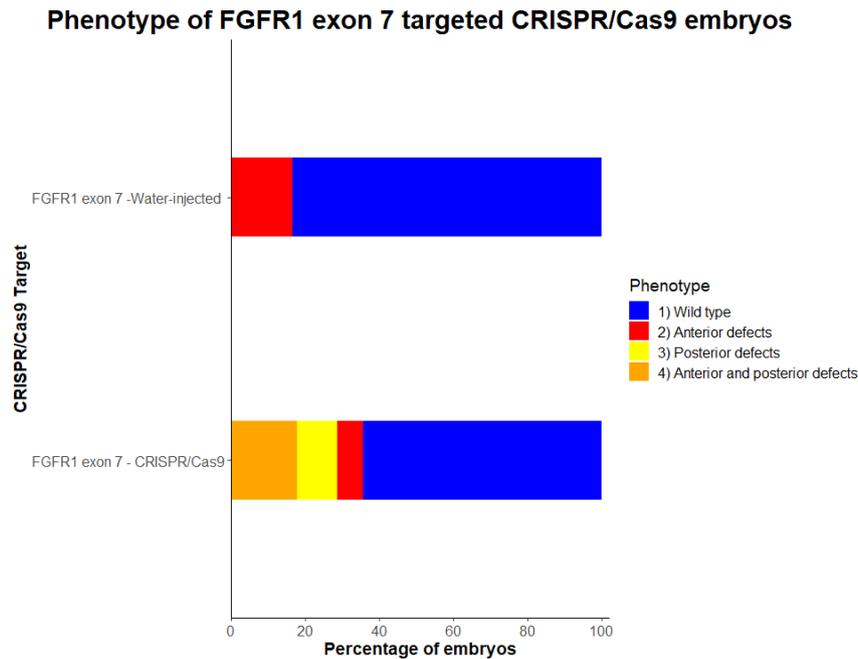
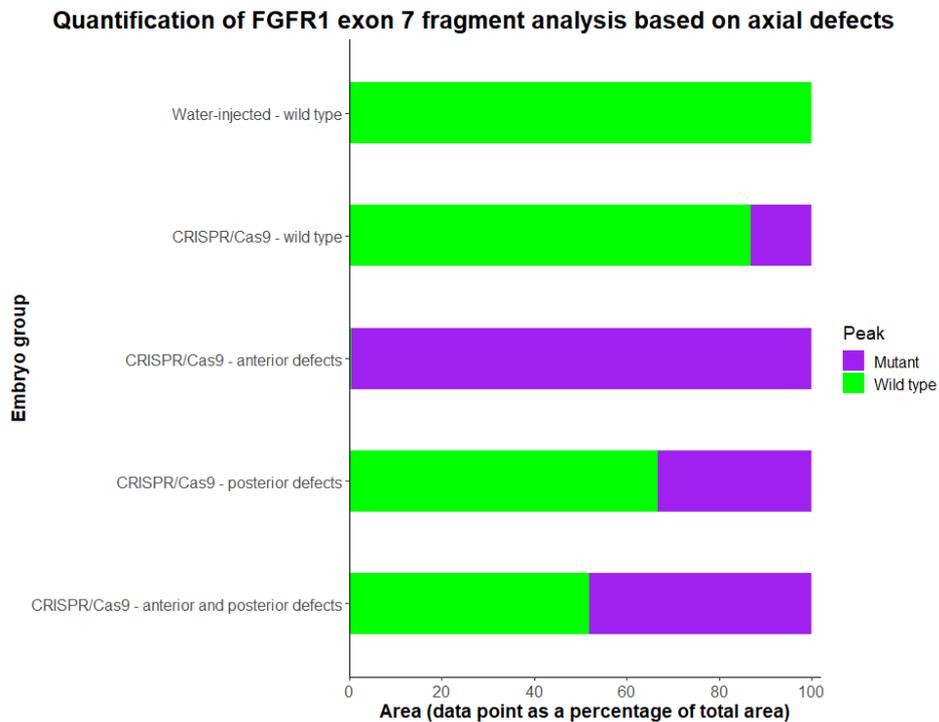
A**B**

Figure 36: Targeting of *fgfr1* exon 7 by CRISPR/Cas9. (A) Phenotypic classification of embryos subject to *fgfr1* gene targeting by CRISPR/Cas9, grouped by axial defects. Number of embryos in each group are as follows: Water-injected (n=18) and FGFR1 exon 7 (n=28). **(B)** Quantification of fragment analysis peaks, using peak area, based on data points, as a percentage of total area. Embryos were categorised by those displaying a wild type phenotype and anterior and/or posterior defects. Number of embryos in each group analysed are as follows: Control (Water-injected) - wild type (n=1), CRISPR/Cas9 - wild type (n=1), CRISPR/Cas9 - anterior defects (n=1), CRISPR/Cas9 - posterior defects (n=3), and CRISPR/Cas9 - anterior and posterior defects (n=2).

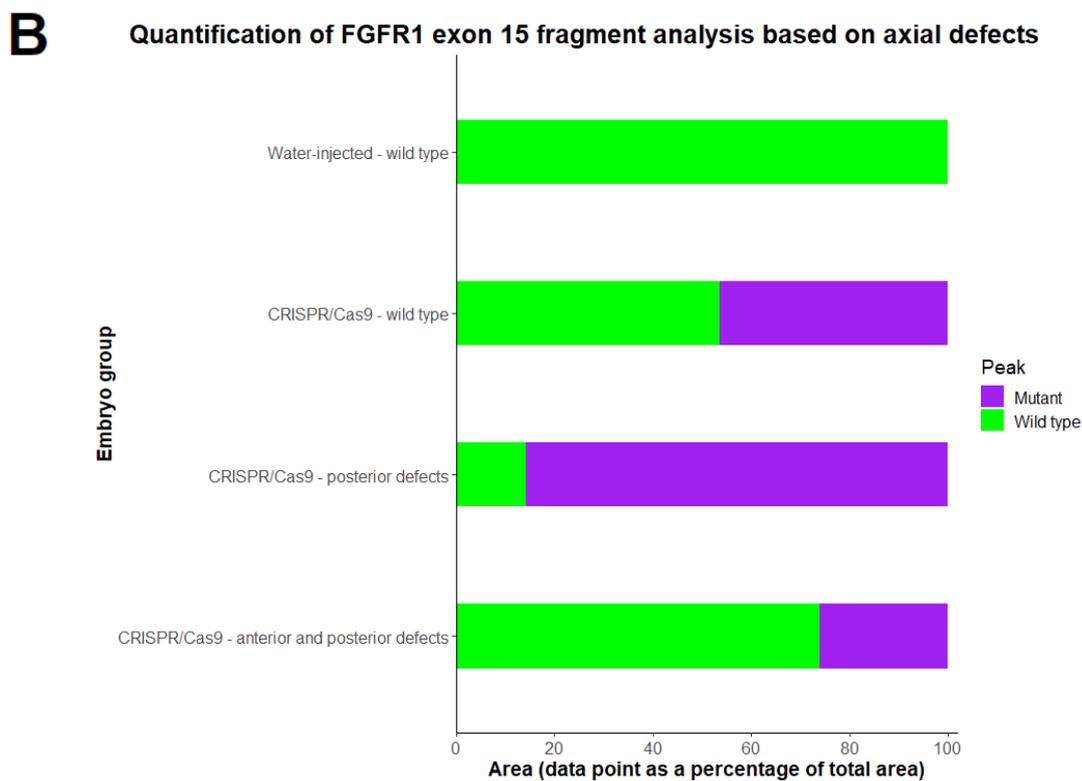
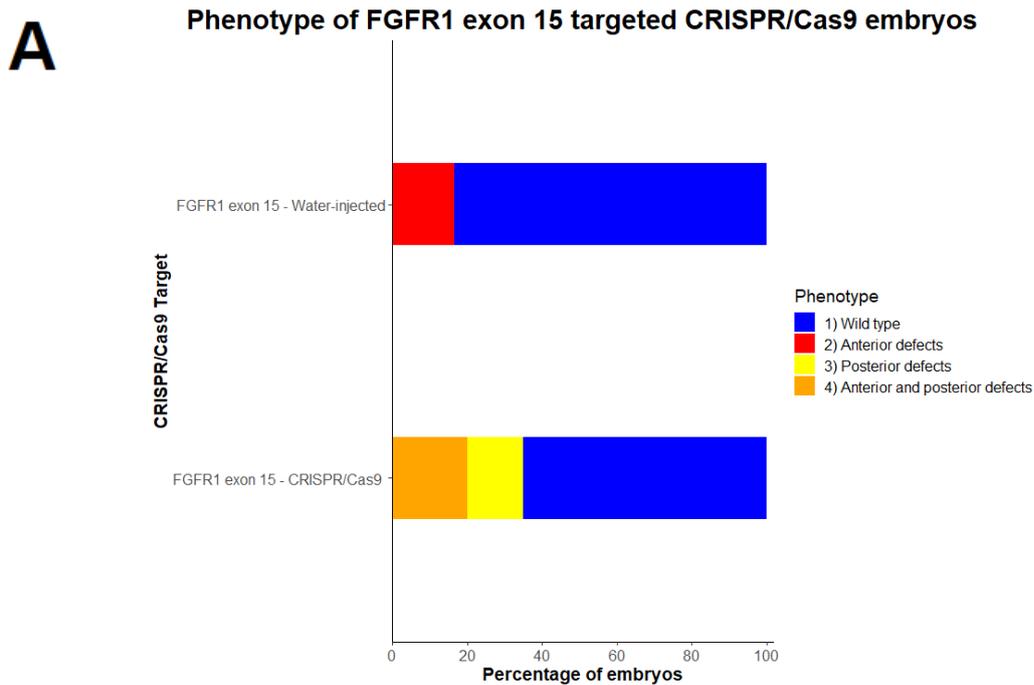


Figure 37: Targeting of *fgfr1* exon 15 by CRISPR/Cas9. (A) Phenotypic classification of embryos subject to *fgfr1* gene targeting by CRISPR/Cas9, grouped by axial defects. Number of embryos in each group are as follows: Water-injected (n=18) and FGFR1 exon 15 (n=20). **(B)** Quantification of fragment analysis peaks, using peak area, based on data points, as a percentage of total area. Embryos were categorised by those displaying a wild type phenotype and anterior and/or posterior defects. Number of embryos in each group analysed are as follows Control (Water-injected) - wild type (n=1), CRISPR/Cas9 - wild type (n=1), CRISPR/Cas9 - posterior defects (n=2), and CRISPR/Cas9 - anterior and posterior defects (n=2).

4.2.3.1 *fgfr1*-targeted embryos

fgfr1 CRISPR/Cas9 targeting exon 7 elicited a higher rate of posterior and a combination of anterior and posterior phenotypic defects in embryos, when compared to respective water-injected control embryos (Figure 36A, S10). Fragment analysis of *fgfr1* exon 7 CRISPR/Cas9 embryos (Figure 36B, S10), separated by phenotypic classification, revealed that 86.8% of fragment area in wild type CRISPR/Cas9 embryos was attributable to those present in the control embryos (n=1). 0.4% fragments were wild type in embryos with anterior defects (n=1), whereas 66.7% and 51.9% were wild type in embryos with posterior defects alone (n=3) and those with a combination of anterior and posterior defects (n=2) respectively.

fgfr1 CRISPR/Cas9 targeting exon 15 elicited a higher rate of posterior and a combination of anterior and posterior phenotypic defects in embryos, when compared to respective water-injected control embryos (Figure 37A, S11). Quantification of fragment analysis results revealed that 53.5% of total peak area in wild type CRISPR/Cas9 embryos represented the wild type peaks (n=1), present in water-injected controls (Figure 37B, S11). This was reduced to 14.1% in those displaying posterior defects (n=2) and 74.0% in those displaying anterior and posterior defects (n=2).

4.2.3.2 *fgfr4*-targeted embryos

The highest percentage of phenotypic defects out of all sgRNAs was due to *fgfr4* exon 3 targeting, whereby almost 60% of embryos displayed axial defects (n=22) (Figure 38A, S12). Quantification of fragment analysis for *fgfr4* exon 3 targeting (Figure 38B, S12) results revealed 100% of fragments in CRISPR/Cas9 wild type embryos were mutant (n=1). 2.8% and 12.8% of peak area represented fragments of wild type length in embryos displaying anterior (n=3) and a combination of anterior and posterior defects (n=4).

Contrasting to *fgfr4* exon 3 targeting, CRISPR/Cas9 embryos displayed comparable rates phenotypic defects to water-injected controls when *fgfr4* exon 5 was targeted (Figure 39A, S13). Furthermore, the proportion of wild type peaks in *fgfr4* exon 5 targeted embryos was considerably higher, at 91.0% and 92.9% in CRISPR/Cas9 embryos with a wild type phenotype (n=1) or displaying a combination of anterior and posterior defects (n=2) respectively (Figure 39B, S13).

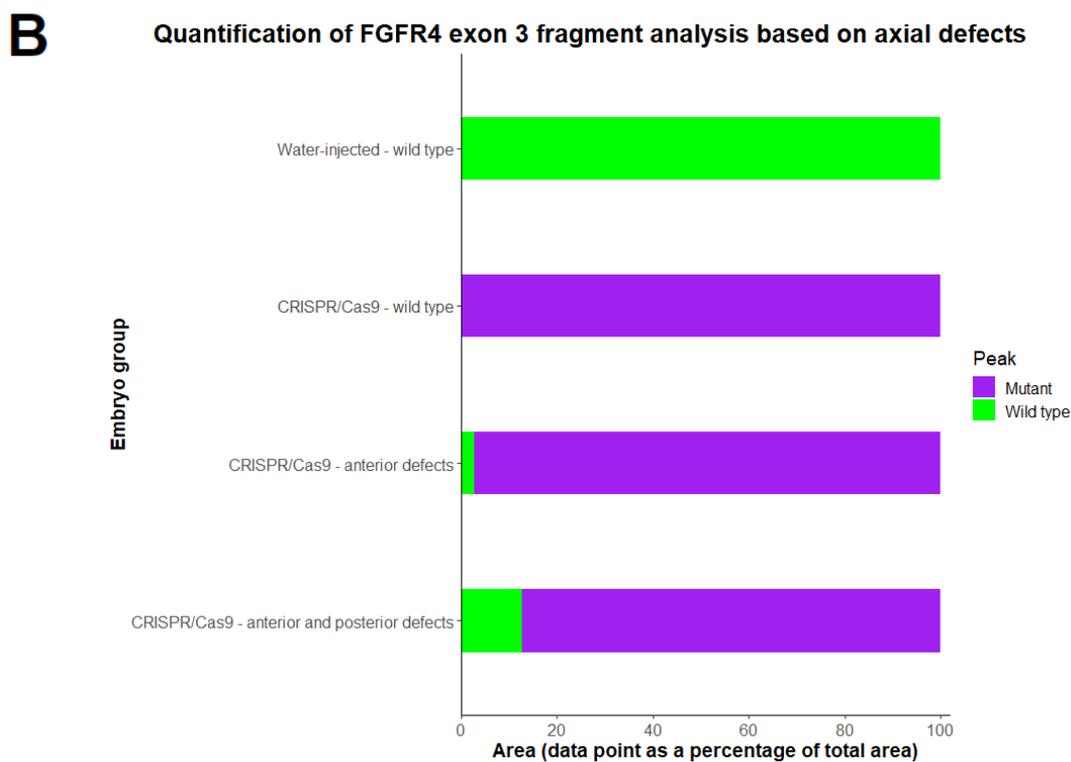
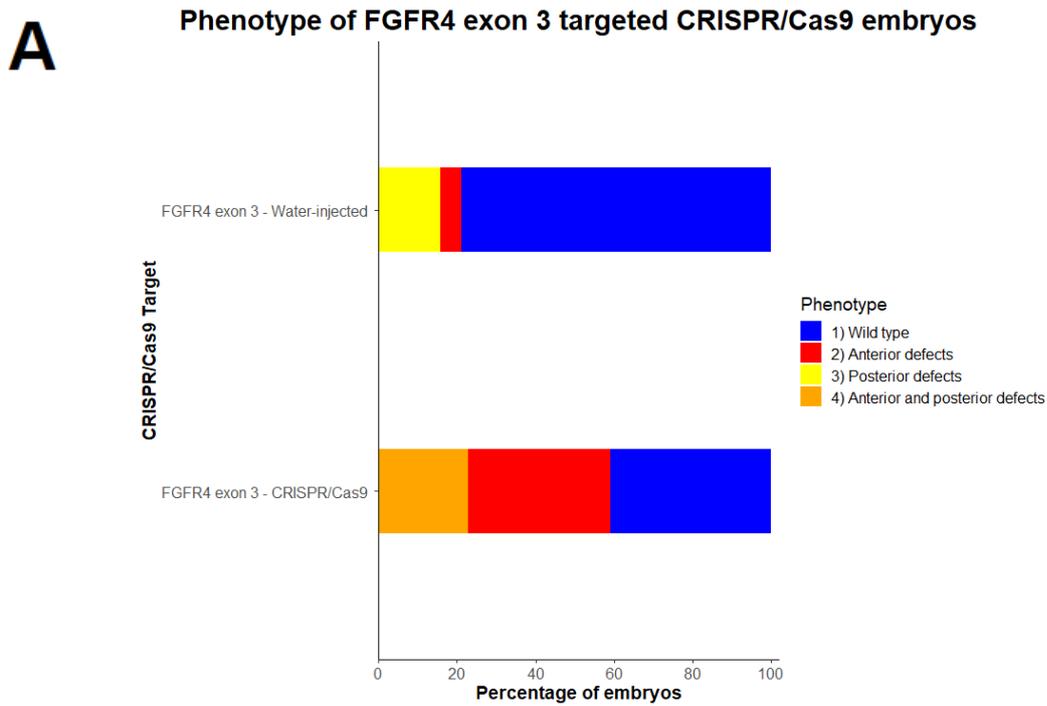


Figure 38: Targeting of *fgfr4* exon 3 by CRISPR/Cas9. (A) Phenotypic classification of embryos subject to *fgfr4* gene targeting by CRISPR/Cas9, grouped by axial defects. Number of embryos in each group are as follows: Water-injected (n=19) and FGFR4 exon 3 (n=22). **(B)** Quantification of fragment analysis peaks, using peak area, based on data points, as a percentage of total area. Embryos were categorised by those displaying a wild type phenotype and anterior and/or posterior defects. Number of embryos in each group analysed are as follows Control (Water-injected) - wild type (n=1), CRISPR/Cas9 - wild type (n=1), CRISPR/Cas9 - anterior defects (n=3), and CRISPR/Cas9 - anterior and posterior defects (n=4).

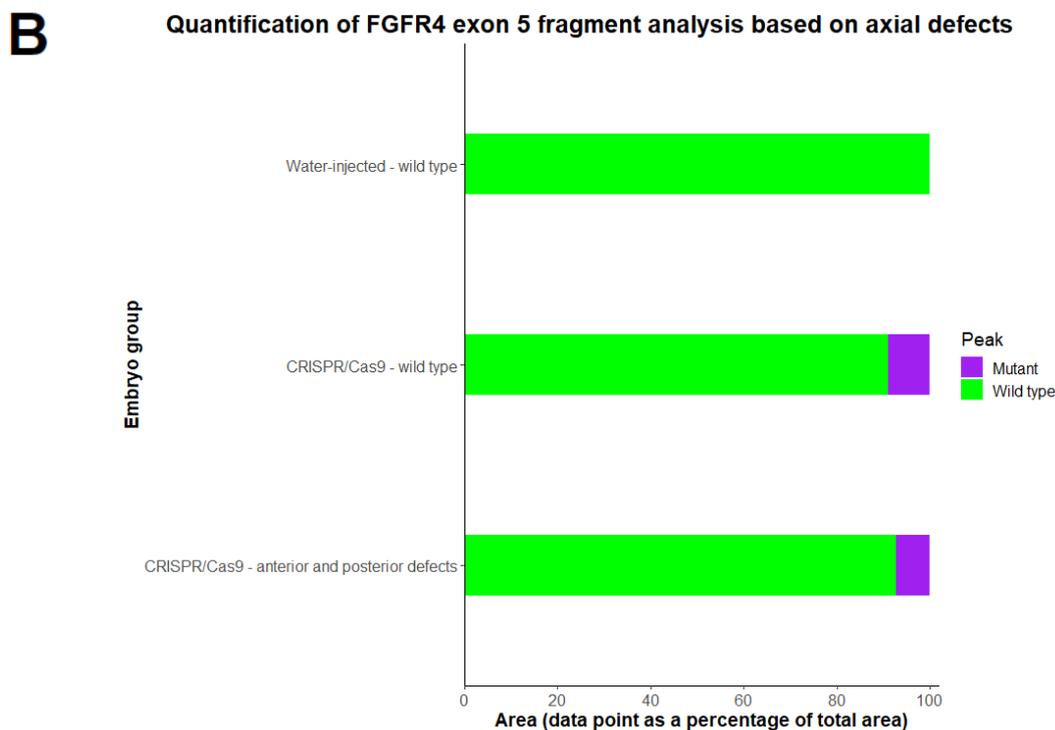
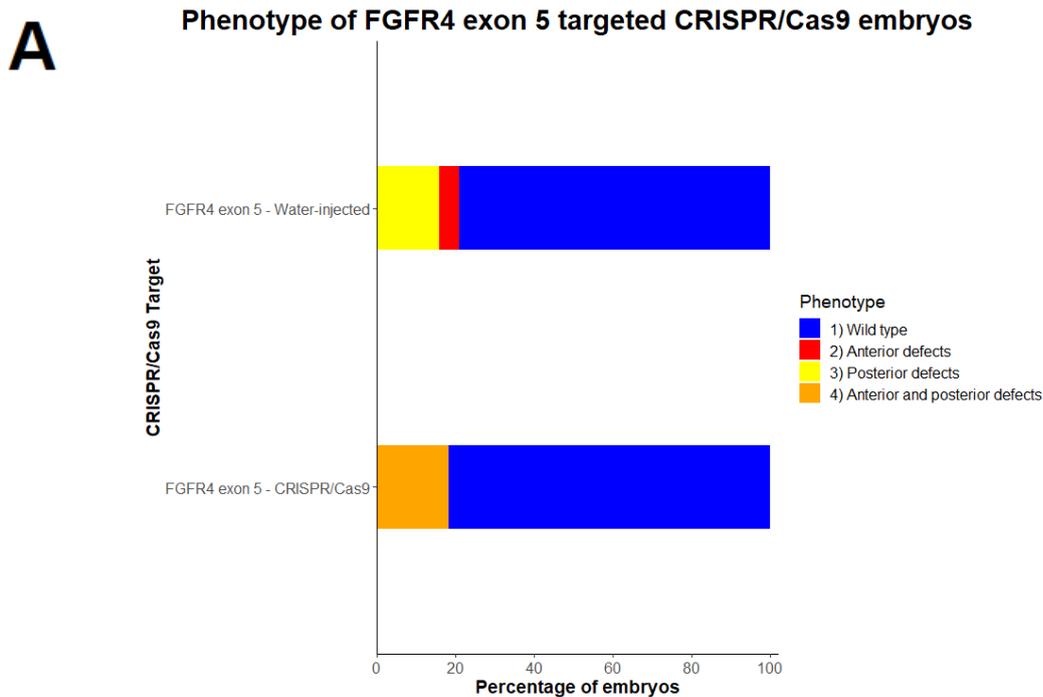


Figure 39: Targeting of *fgfr4* exon 5 by CRISPR/Cas9. (A) Phenotypic classification of embryos subject to *fgfr4* gene targeting by CRISPR/Cas9, grouped by axial defects. Number of embryos in each group are as follows: Water-injected (n=19) and FGFR4 exon 5 (n=22). **(B)** Quantification of fragment analysis peaks, using peak area, based on data points, as a percentage of total area. Embryos were categorised by those displaying a wild type phenotype and anterior and/or posterior defects. Number of embryos in each group analysed are as follows Control (Water-injected) - wild type (n=1), CRISPR/Cas9 - wild type (n=1), and CRISPR/Cas9 - anterior and posterior defects (n=2).

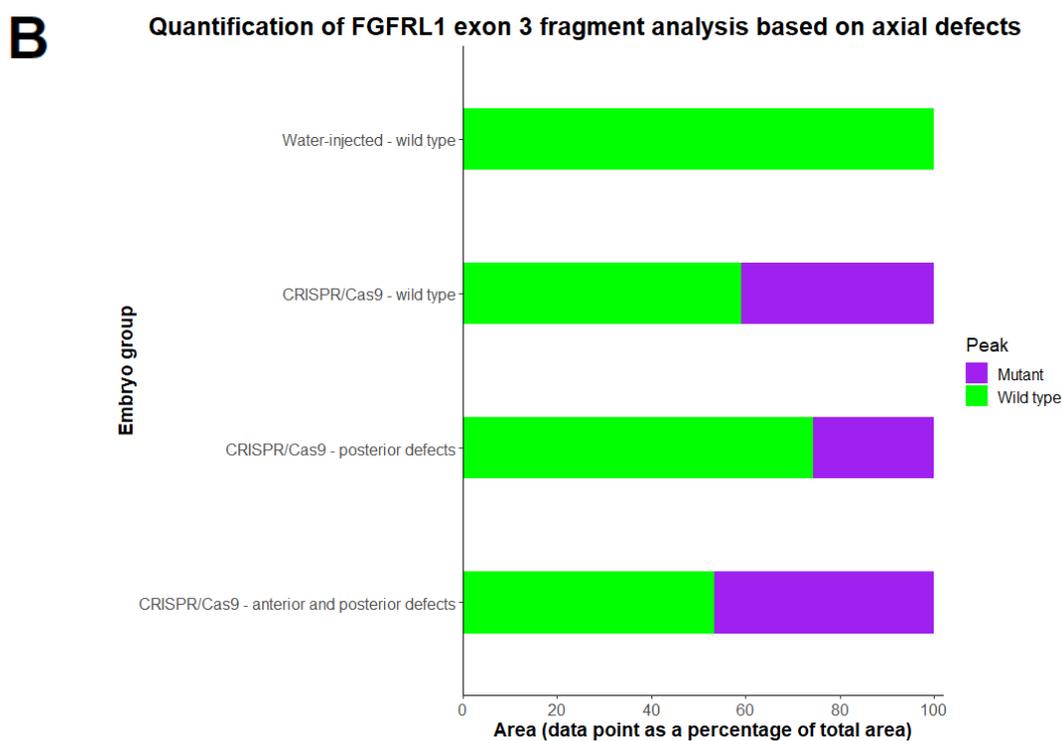
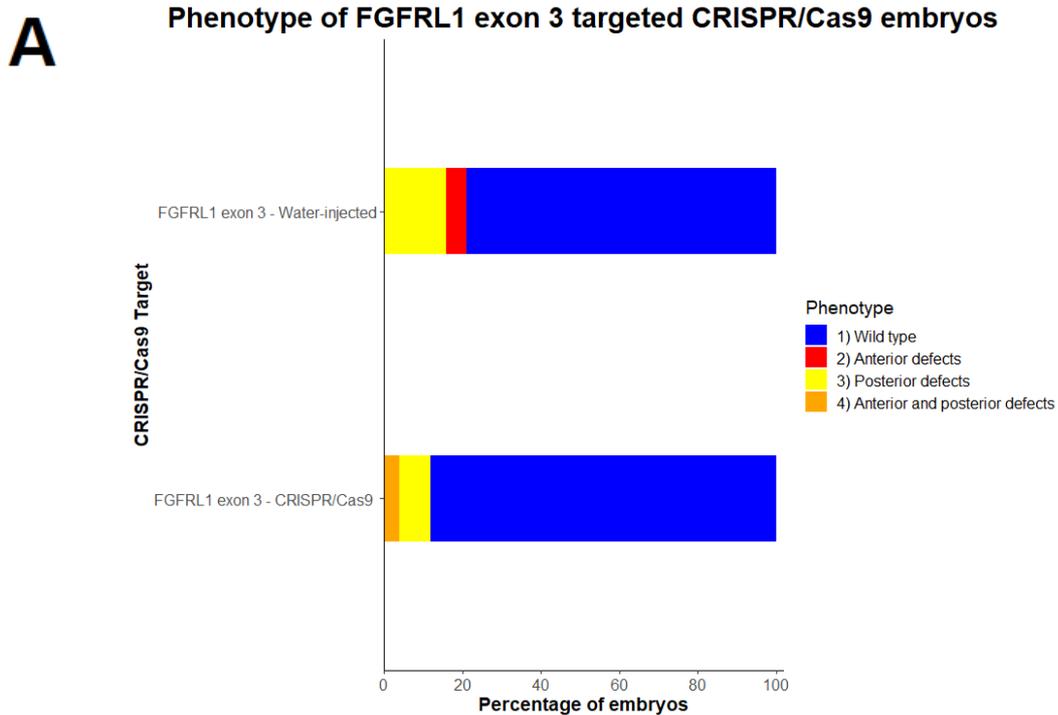


Figure 40: Targeting of *fgfr1* exon 3 by CRISPR/Cas9. (A) Phenotypic classification of embryos subject to *fgfr1* gene targeting by CRISPR/Cas9, grouped by axial defects. Number of embryos in each group are as follows: Water-injected (n=19) and FGFR1 exon 3 (n=25). **(B)** Quantification of fragment analysis peaks, using peak area, based on data points, as a percentage of total area. Embryos were categorised by those displaying a wild type phenotype and anterior and/or posterior defects. Number of embryos in each group analysed are as follows Control (Water-injected) - wild type (n=1), CRISPR/Cas9 - wild type (n=1), CRISPR/Cas9 - posterior defects (n=1), and CRISPR/Cas9 - anterior and posterior defects (n=1).

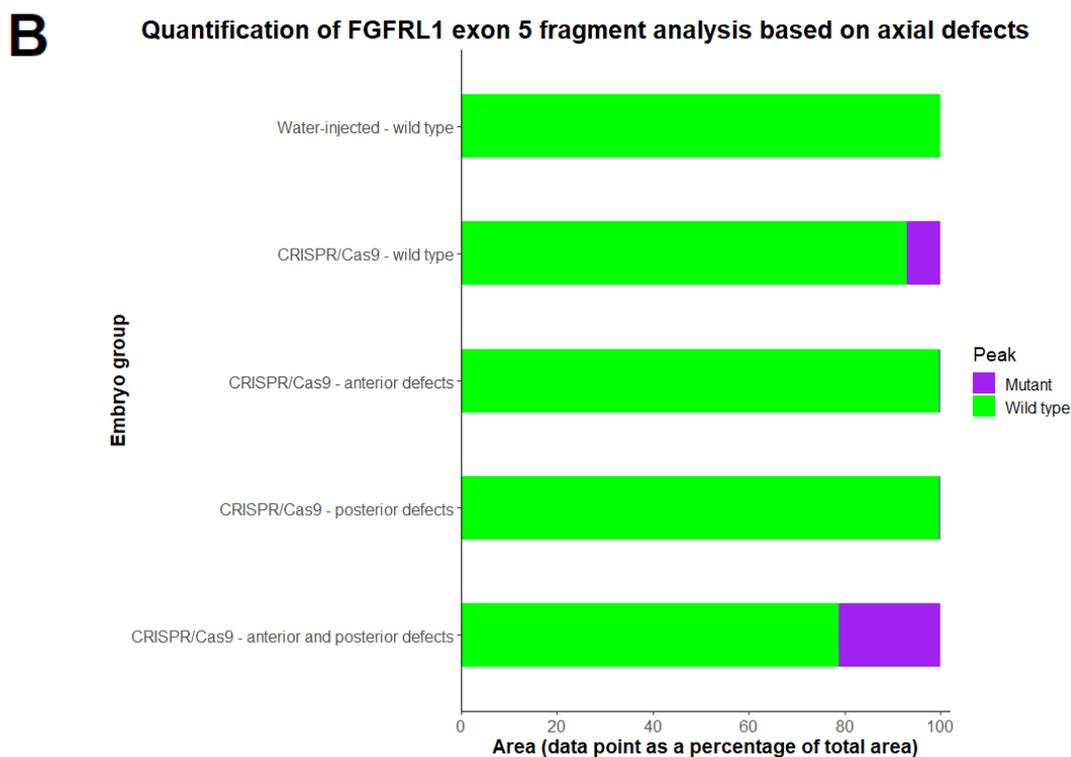
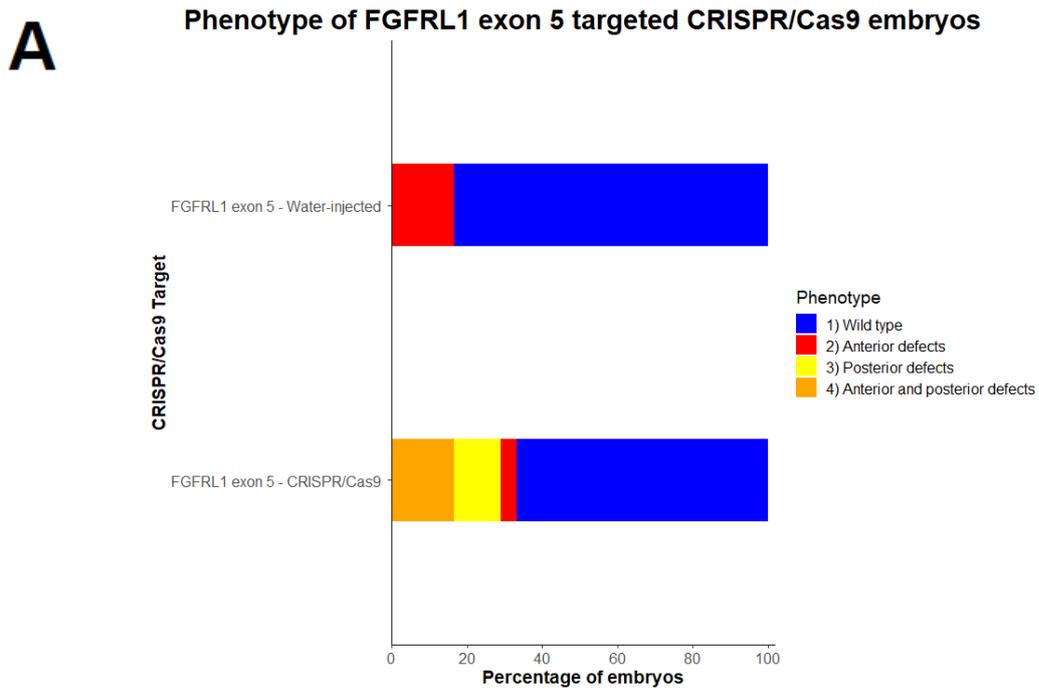


Figure 41: Targeting of *fgfr1* exon 5 by CRISPR/Cas9. (A) Phenotypic classification of embryos subject to *fgfr1* gene targeting by CRISPR/Cas9, grouped by axial defects. Number of embryos in each group are as follows: Water-injected (n=18) and FGFR1 exon 5 (n=24). **(B)** Quantification of fragment analysis peaks, using peak area, based on data points, as a percentage of total area. Embryos were categorised by those displaying a wild type phenotype and anterior and/or posterior defects. Number of embryos in each group analysed are as follows Control (Water-injected) - wild type (n=1), CRISPR/Cas9 - wild type (n=1), CRISPR/Cas9 - anterior defects (n=1), CRISPR/Cas9 - posterior defects (n=2), and CRISPR/Cas9 - anterior and posterior defects (n=2).

4.2.3.3 *fgfr1*-targeted embryos

fgfr1 exon 3 CRISPR/Cas9 targeting a lower higher phenotypic defect rate than in water-injected controls (Figure 40A, S14). Fragment analysis of *fgfr1* exon 3 targeted embryos revealed that 59.1%, 74.4% and 53.4% of the area of peaks represented wild type fragments in *fgfr1* exon 3 CRISPR/Cas9 embryos displaying wild type (n=1), posterior defects (n=1) and a combination of anterior and posterior defects (n=1) respectively (Figure 40B, S14).

fgfr1 targeting exon 5 led to the occurrence of twice as many phenotypic defects than observed in the control embryos (Figure 41A, S15). *fgfr1* exon 5 targeting yielded the lowest efficiencies in indel introduction, leading to the detection of mutant fragments in fragment analysis (Figure 41B, S15). 93.0%, 99.7% and 99.8% of total peak area was constituted by wild type fragments, present in the water-injected control, in embryos displaying a wild type phenotype (n=1), anterior defects (n=1) and posterior defects (n=2) respectively. This was reduced to 78.8% in embryos presenting anterior and posterior defects (n=2).

4.2.4 Summary

Taken together, this demonstrates the successful targeting of *tyrosinase*, *fgfr1*, *fgfr4* and *fgfr1* by CRISPR/Cas9. Furthermore, fragment analysis enabled the identification and quantification of mutant fragments containing indels, as a result of the error-prone process of NHEJ, and this emphasised the mosaic nature of this genome editing approach. Currently, fragment analysis results, together with phenotypic observations, are insufficient to enable the prediction of the most efficient sgRNA to target each receptor for future work.

4.3 Discussion

4.3.1 Key findings

In summary, results in this chapter demonstrate the successful targeting of *tyrosinase*, *fgfr1*, *fgfr4* and *fgfr11* by CRISPR/Cas9 in F0 embryos. Fragment analysis enabled the identification of mutant fragments containing indels, as a result of the error-prone process of NHEJ. These results demonstrate the mosaicism of genome editing by CRISPR/Cas9, visualised in phenotypic observations and quantified by fragment analysis.

4.3.2 Successful preliminary *tyrosinase* targeting

tyrosinase was selected for targeting due to its visual phenotype, resulting from gene mutagenesis (Ando et al. 2007). Phenotypic analysis of *tyrosinase* targeted embryos highlights the mosaicism caused by the CRISPR/Cas9 approach. Although albinism, either mosaic or complete, was observed in 42.9% and 60% of embryos injected with 400pg or 600pg sgRNA respectively, higher rates of albinism have been observed in the literature (Blitz et al. 2013; Nakayama et al. 2013; Guo et al. 2014; Bhattacharya et al. 2015). Other phenotypic defects, including oedemas, were observed in this investigation. However, these have been noted previously and were determined to not be the result of CRISPR/Cas9 injections, as these were also observed in control embryos at a comparable rate (Blitz et al. 2013; Nakayama et al. 2013).

It has been suggested that a Cas9 injected control may be more applicable than a water injected control to demonstrate that the Cas9 protein alone does not elicit phenotypic defects and genome modifications. This has been implemented by a previous lab member (McQueen and Pownall 2017) and was trialled in this series of experiments targeting the *tyrosinase* gene. Uninjected controls have also been utilised in the literature (Blitz et al. 2013), however these embryos do not control for the possibility of the injection itself causing defects, and therefore water injected or Cas9 injected controls are arguably more informative. An additional control of performing a rescue with *tyrosinase* mRNA coinjection has been unsuccessful due to degradation of the injected mRNA prior to expression of the *tyrosinase* gene, commencing during late tailbud stages (Nakayama et al. 2013).

4.3.3 Phenotypic defects of *fgfr* knockout embryos were observed at a low rate but are consistent with known functions of FGF signalling

The phenotypic defects of *fgfr* targeted embryos include anterior defects, such as the presence of oedemas and enlargement, reduction or absence of the head, ventral endodermal yolk

mass or eyes, while posterior defects include posterior truncations and tail curvature. These are consistent with known functions of FGF signalling.

Suppression of the head was observed in some *fgfr* targeted embryos. *In situ* hybridisations detected *fgfr1* expression throughout the anterior head region at early neurula stage and subsequently confined to the forebrain, eyes, midbrain-hindbrain boundary and otic vesicles in tailbud stages (Hayashi et al. 2004). The expression of this putative negative regulator of FGF signalling (Steinberg et al. 2010) could function to inhibit FGF signalling in these regions as FGF posteriorises the embryo. This posteriorising activity was evidenced by the suppression of heads in embryos subject to FGF4 overexpression (Isaacs et al. 1994). Therefore, disruption of the *fgfr1* gene could lead to loss of FGF signalling inhibition in anterior regions, resulting in head suppression.

Eye defects observed in CRISPR/Cas9 embryos included enlargement, reduction or absence of this structure in tailbud stage embryos. All *fgfrs* targeted in this investigation are expressed in the eyes of *Xenopus* embryos (Hayashi et al. 2004; Lea et al. 2009; Sempou et al. 2018), with *fgfr1* and *fgfr4* expression overlapping in the cells surrounding the lens (Lea et al. 2009). Embryos subject to FGF signalling negative regulator *sulf1* overexpression exhibit reduced eyes, which is the equivalent of FGF signalling inhibition by *fgfr* gene disruption. Although this phenotype is consistent with *sulf1* expression in the retina and lens of the eye, this phenotype could be due to Sulf acting on a different signal transduction pathway, such as BMP signalling (Freeman et al. 2008). Increased FGF signalling, either through FGF4 overexpression or iFGFR signalling, resulted in the loss or malformation of eyes (Isaacs et al. 1994; Pownall et al. 1996; Brunson and Isaacs 2020), caused by an increase in posteriorly expressed genes (Pownall et al. 1996).

The posterior truncations observed in CRISPR/Cas9 embryos in this study is characteristic of embryos subject to FGF signalling inhibition by dnFGFR1 (Amaya et al. 1991; Isaacs et al. 1994) and overexpression of transmembrane Sef protein, which is negative regulator of FGF signalling (Tsang et al. 2002). These posterior defects resemble that of *cdx* gene knockdowns using MOs (Faas and Isaacs 2009). It was concluded that FGF signalling regulates the expression of posteriorly expressed *hox* genes, which is mediated by *cdx* genes, for example *cdx4* (Pownall et al. 1996).

Microinjection of 600pg sgRNA and 1.5ng Cas9 in this investigation is higher than when performed by previous lab members, who co-injected with 1ng Cas9 protein and 300pg of sgRNA, when targeting *myod* (McQueen and Pownall 2017), however this could be due to the

different Cas9 utilised. This injection of higher CRISPR/Cas9 reagent concentrations could have elicited higher levels of phenotypic axial defects. However, this work necessitates further improvement as across all experiments 70.8% CRISPR/Cas9 embryos exhibited a wild type phenotype, and the highest percentage of phenotypic defects was due to *fgfr4* exon 3 targeting, whereby almost 60% of embryos displayed axial defects, in comparison with other sgRNAs targeting *fgfr4* eliciting organ laterality defects in 15-30% of embryos (Sempou et al. 2018). Other sgRNAs, namely targeting *fgfr4* exon 5 and *fgfr1* exon 3, elicited phenotypic defects at a comparable or lower rate respectively to that of respective control embryos, which was surprising. This contrasts the rate of expected phenotypes in other CRISPR/Cas9 studies in *X. tropicalis*, for example 50% and 40.8% when *six3* and *pax8* were targeted (Nakayama et al. 2013; Bhattacharya et al. 2015).

Gene-editing by TALENs elicited either partial or complete albinism in over 90% of F0 *tyrosinase*-targeted embryos (Ishibashi et al. 2012) and up to 95.7% efficiency in generating indel mutations in a variety of targeted loci (Lei et al. 2012), which were heritable (Ishibashi et al. 2012; Lei et al. 2012). However, TALEN mRNA translation during rapid cell divisions in early development could delay gene modification. This could be improved using a fusion construct to promote mRNA translation in the oocyte to increase mutagenesis, however this is a laborious process (Nakajima and Yaoita 2015). The CRISPR/Cas9 protocol developed in this project does not require this and the results presented are promising, considering that the phenotypic analysis was only performed once, due to the circumstances. This could explain why similar rates of phenotypic defect classes were not observed when individual *fgfrs* were targeted with different sgRNAs. This could be the result of the mosaic nature of CRISPR/Cas9 targeting or suggestive of off-target mutations occurring. Therefore, firstly, these rates could have been improved with subsequent work, and secondly, statistical analysis cannot be performed to determine the significance of these results.

This low rate of phenotypic defects could be due to CRISPR/Cas9 only mutating transcripts expressed from genes post MBT (Bhattacharya et al. 2015; Leerberg et al. 2019), with maternal *fgfr1* and *fgfr4* transcripts remaining unaffected (Lea et al. 2009). However, *fgfr1* expression is first detected during late gastrula stages (Hayashi et al. 2004), so this possible explanation is not applicable for this receptor. However, this low rate could be attributed to potential FGFR redundancy, hypothesised after developmental defects were observed in double and triple *fgfr* knockouts, but not in single knockout zebrafish embryos (Leerberg et al. 2019). This idea of genetic redundancy among FGFRs was observed in single, double and triple *fgfr* mutant mice, exhibiting telencephalon truncations with increasing severity, in FGFR3 null mice with Cre-Lox recombination to target FGFR1 and FGFR2 (Paek et al. 2009).

Furthermore, FGFRs have overlapping ligand specificities, whereby individual ligands exhibit preferences for multiple receptors, stated in Table 1 (Itoh and Ornitz 2004; Zhang et al. 2006). Therefore, FGF ligands could perhaps signal through an alternative receptor in an FGFR knockout embryo to continue FGF signalling. This could be investigated by co-injecting sgRNAs targeting multiple receptors, creating double or triple knockout embryos. This approach has been successful in simultaneously mutating up to five genes in mouse embryonic stem cells (Wang et al. 2013).

The aim of this chapter was to investigate the effects of FGFRs in mesoderm and anteroposterior patterning. Therefore, embryos were collected prior to the opportunity of observing gut and cardiac looping, as an indication for left-right asymmetry (Sempou et al. 2018). However, due to FGFR1 and FGFR4 previously being implicated in this patterning process (Neugebauer et al. 2009; Sempou et al. 2018), it is logical to predict incorrect looping and internal organ positioning in embryos with these genes disrupted. The involvement in left-right asymmetry could be part of future research.

4.3.4 FGFR studies in other model organisms

This investigation had aimed to compare the function of individual receptors in *X. tropicalis*, which has been explored in other model organisms, including zebrafish and mice.

The breeding of CRISPR/Cas9 *fgfr4*-targeted F0 individuals resulted in phenotypically normal wild type and heterozygotes, along with homozygous mutant *fgfr4* zebrafish. These double knockout fish exhibited posterior truncations and oedemas (Varshney et al. 2015), which are consistent with the observations presented in this study and the role of FGF signalling in patterning the anteroposterior axis (Amaya et al. 1991; Isaacs et al. 1994; Pownall et al. 1996; Faas and Isaacs 2009). However, in another study, all single FGFR mutants were deemed as morphologically wild type, with only double and triple knockout embryos exhibiting phenotypic defects, which could be the result of increased mRNA levels of other receptors to enable continued FGF signalling (Leerberg et al. 2019).

fgfr1 knockouts in mice are embryonic lethal between E7.5 and E9.5 (Yamaguchi et al. 1994), which contrasts with *Xenopus* work, including the results presented in this study using CRISPR/Cas9 and in the wider literature employing dnFGFRs (Amaya et al. 1991; Isaacs et al. 1994; Pera et al. 2003; Delaune et al. 2005; Branney et al. 2009). It is surprising that dnFGFR overexpression is not lethal, since dnFGFRs likely inhibit all FGFR signalling (Ueno et al. 1992). Nevertheless, there are similarities in the absence of somites in both *fgfr1*

knockout mice and dnFGFR1 *Xenopus* embryos (Amaya et al. 1991; Yamaguchi et al. 1994), suggesting a conserved role of FGFR1 signalling in the formation of this structure.

Mice with targeted disruptions in *fgfr4* gene display abnormal muscle regeneration (Zhao et al. 2006), which contrasts with another *fgfr4* null line, which develops normally with no apparent morphological defects (Weinstein et al. 1998). Both groups confirmed a loss of *fgfr4* mRNA and protein in homozygous mutant mice (Zhao et al. 2006; Weinstein et al. 1998). However, *fgfr3 fgfr4* double knockout mice exhibit phenotypes which are not observed in single knockout mice, including dwarfism, lung abnormalities and lack of alveologenesi (Weinstein et al. 1998), further evidence for FGFRs functioning cooperatively (Weinstein et al. 1998; Leerberg et al. 2019). dnFGFR studies in *Xenopus* suggest a greater involvement of FGFR4 in neural induction and patterning than FGFR1 (Hongo et al. 1999; Hardcastle et al. 2000), however dnFGFRs lack the specificity to inhibit a single receptor (Ueno et al. 1992).

Targeted disruptions in *fgfr1* result in death shortly after birth due to decreased diaphragm size (Baertschi et al. 2007; Catela et al. 2009). Catela et al. (2009) also discovered axial, cardiac and skeletal defects, which were not observed in Baertschi et al. (2007). This difference has been hypothesised to be due to differences in mutant gene construction (Catela et al. 2009). Further studies, utilising the method of *fgfr1* knockout mice generation by Baertschi et al. (2007), have pointed to a role of FGFR1 in nephrogenesis, for which functional IgIII is crucial (Gerber et al. 2009; Gerber et al. 2020). These phenotypic defects overlap with those highlighted in *Xenopus* studies, which revealed the resemblance of embryos subject to dnFGFR1 and *fgfr1* overexpression in displaying posterior truncations and incorrect muscle development (Amaya et al. 1991; Steinberg et al. 2010). However, differences in FGFR1 function in mice could be attributed to the replacement of a proportion of the intracellular histidine with unrelated residues (Sleeman et al. 2001; Trueb et al. 2003).

4.3.5 Variable targeting efficiencies of sgRNAs in fragment analysis experiments

Fragment analysis results demonstrated that all the tested sgRNAs were capable of generating indels in targeted genes. Analysis of CRISPR/Cas9 embryos revealed peaks of smaller and larger sizes, other than the expected wild type peak. These peaks correspond to fragments containing the deletion or insertion of bases at the target size, resulting from Cas9 break induced repair (van Overbeek et al. 2016). Cloning and sequencing of DNA from *tyrosinase*-targeted mice revealed that deletions occur more frequently than insertions (Yen et al. 2014). This is consistent with data presented in this thesis and results from *myod*-targeted *X. tropicalis* embryos (McQueen and Pownall 2017). Specifically, indels are typically less than 20bp (van Overbeek et al. 2016).

Unfortunately, sgRNA targeting *fgfr1* exon 2 was not tested in this study before lockdown. Fragment analysis of *fgfr1* exon 7 targeted embryos displaying posterior phenotypic defects revealed almost 100% of peaks were mutant, with a 1bp deletion. This is considerably greater than that of embryos displaying anterior or a combination of anterior and posterior defects. Targeting exon 15, constituting the intracellular tyrosine kinase domain, in *fgfr1* yielded a higher wild type peak area in embryos displaying anterior and posterior defects. Repeats of *fgfr1* targeting with larger numbers of embryos could potentially reveal differences in phenotypes and effects on FGF target gene expression resulting from targeting *fgfr1* at extracellular (exon 2 or exon 7) or intracellular (exon 15) domains. For example, mutations in exon 15, encoding the tyrosine kinase domain, could still produce a functional protein to be inserted into the cell membrane, but instead functionally mimic FGFR1, which lacks an intracellular tyrosine kinase domain (Sleeman et al. 2001).

Similar to *fgfr1* exon 15, fragment analysis for *fgfr4* exon 3 revealed unexpected results. 100% of the peak area in CRISPR/Cas9 wild type embryos was mutant fragments containing indels. Having said that, targeting of *fgfr4* exon 3 and *fgfr1* exon 3 yielded the highest proportion of mutant peak area and therefore, these sgRNA are classified as the most effective sgRNA for the respective receptor. However, similar to that observed in *X. tropicalis* FGFR1 and FGFR2 (Lea et al. 2009), splice variants of mouse FGFR1 have been identified, with the inclusion or exclusion of Igl (Sleeman et al. 2001), which exon 3 encodes and sgRNA was targeted to, by a previous MSc student (Zilinskaite 2019). Furthermore, this series of experiments was only repeated once so statistical analysis cannot be performed to determine significance. Therefore, larger numbers of embryos per treatment with biological replicates are required before more conclusive conclusions can be derived.

CRISPR/Cas9 targeting could be quantitatively measured by T-cloning, as performed by Blitz et al. (2013), which displayed the results of 19 clones from a single embryo, 84% of which contained indels either upstream or proximal to the PAM site. Furthermore, this process of fragment cloning and sequencing would enable the characterisation of indels present, which is crucial since early frameshift mutations are likely to disrupt protein function or cause NSMD (Bhattacharya et al. 2015).

4.3.6 Recommendations for further optimisations of CRISPR/Cas9 protocol

Redesigning sgRNA could increase the incidence of phenotypic defects. This was recommended by Guo et al. (2014), which analysed the efficiency of 10 sgRNAs targeting different *X. tropicalis* genes. A highly variable efficiency rate between 45% to 100% was

elucidated, which could be increased by redesigning or increasing the concentration of sgRNA.

An approach to improve the rate of phenotypic defects and subsequent fragment analysis results would be to inject at only the 1 cell stage, rather than also into each cell at the 2 cell stage. This would likely decrease mosaicism and increase the number of fragments containing indels from the resulting embryos. Bhattacharya et al. (2015) directly compared the efficiency of CRISPR/Cas9 injections at the 1 and 2 cell stage, along with being the first paper to utilise Cas9 protein, rather than mRNA, visualised by *tyrosinase* albinism phenotypes and fragment analysis results. This revealed that Cas9 mRNA is more toxic when injected and less efficacious than Cas9 protein in gene-editing (Bhattacharya et al. 2015). This reduced efficacy of Cas9 mRNA could be due to the requirement of mRNA translation, which is also the case for TALEN approach, delaying genome editing and therefore increasing the presence of wild type sequences. This is particularly crucial in the rapid development of *Xenopus* embryos, enabling a short generation time (Blitz et al. 2013), in relation to mouse embryos, whose cell divisions occur slower. Therefore, translation of Cas9 mRNA does not hinder biallelic mutagenesis in mice (Yen et al. 2014).

The efficiency rate could be further elevated by more microinjections into multiple sites in the embryo to more efficiently distribute the complex, for example Blitz et al. (2013) injected 1nl into four sites above the equator per embryo, with a total injection volume of 4nl. Furthermore, co-injection of multiple sgRNAs targeting the same exon could increase mutagenesis rates (Yen et al. 2014). However, it is important to determine the toxicity of sgRNAs, which could impact on survival (Nakayama et al. 2013).

Contrasting to the method of embryo injection resulting in mosaic NHEJ presented in this investigation, injection of CRISPR/Cas9 reagents and DNA ligase inhibitor into oocytes would enable maternal genome editing and increased homology directed repair activity. This would be followed by the host transfer method and subsequent eggs would be fertilised by sperm from wild type males. This method results in wild type or non-mosaic heterozygous F0 embryos, which can be genotyped and bred to generate homozygous F1 embryos. Although this method is more time-consuming, higher F0 indel frequency has been observed than that resulting from embryo microinjection (Aslan et al. 2017).

4.3.7 Fragment analysis caveats and recommendations

Although all the PCR products were diluted by the same factor for fragment analysis, this was not quantitative and the differences in the fluorescence between embryos could be due to less

PCR product present in the sample, rather than CRISPR/Cas9 targeting. The optical density of samples could be measured and adjusted to one level for all samples, enabling comparisons of peak height within and between runs. However, this will include primers, partial fragments and remaining nucleotides in the measurement, which there could be varying amounts of in different samples. To overcome this, the extension time could be increased further, to minimise the chance of incomplete fragments, and a PCR clean up could be performed to remove impurities.

Unexpectedly, there are additional peaks, as well as the peak corresponding to the wild type fragment, present in uninjected, Cas9 injected and water injected control embryos across multiple experiments with different CRISPR/Cas9 targeting. The appearance and size of these fragments varies within and between experiments, with inconsistent differences from the expected fragment size. However, these fragments were usually smaller in length than the expected fragment, suggesting they are incomplete PCR fragments. This makes concluding on CRISPR/Cas9 targeting efficiency difficult, since CRISPR/Cas9 more frequently elicits deletions than insertions (Yen et al. 2014).

Attempts to amend the reagents and protocol to eliminate these fragments were made by ordering new FAM stock, making fresh primer stocks, reducing the PCR cycle number and increasing the annealing temperature. These alterations were tested using genomic DNA from wild type embryos in step 1 and step 2 PCR reactions. Fragment analysis of the step 2 PCR product revealed the continued presence of multiple peaks in wild type embryos (data not shown).

Other possible explanations include performing the PCR reactions on consecutive days, which could lead to degradation of fragment ends. However, these fragments are up to 330bp in length and therefore highly stable, so this is unlikely. A step 2 PCR clean up would remove impurities in the step 2 PCR product before fragment analysis, however this would likely remove considerably smaller fragments, for example nucleotides and primers but not the fragments closer to the expected size. Nevertheless, this was attempted, however they were unsuccessful in retaining the PCR product and there was no opportunity to repeat this. Finally, the multiple peaks could be due to heterogeneity in the size of the PCR primers, whereby some smaller incomplete fragments may be present in the stock. This could be tested by specifying a more stringent purification protocol, for example high performance liquid chromatography (HPLC), rather than the desalted method.

4.3.8 Future work

At the beginning of this project, a prediction of the most efficient sgRNA for each FGFR was reasonable aspiration. However, there is insufficient data to conclude on this, having only repeated experiments once due to the pandemic. Once this has been determined, *fgfr1*, *fgfr4* and *fgfr1* knockouts using the CRISPR/Cas9 sgRNA with the highest targeting efficiency, validated by fragment analysis, could be carried out, along with appropriate controls, including dnFGFR1 as a positive control for FGF signalling inhibition (Amaya et al. 1991; Isaacs et al. 1994; Branney et al. 2009). Western blots for dpERK levels could be performed on the resulting embryos, which should reveal a decrease in levels in FGFR1 and FGFR4 knockout embryos, representing correct FGF signalling inhibition. FGFR1-knockout embryos could display an increase in dpERK levels as FGFR1 is a putative negative regulator of FGF signalling (Steinberg et al. 2010). Sibling embryos could be used in a RT-PCR experiment to analyse the expression of well-supported FGF gene targets *cdx4*, *egr1*, *myod*, *tbxt* and *spry2*, along with *odc1* loading control (Isaacs et al. 1994; Pownall et al. 1996; Fisher et al. 2002; Isaacs et al. 2007; Branney et al. 2009; Faas and Isaacs 2009; Nentwich et al. 2009). The expression of these FGF target genes should be reduced in *fgfr1* and *fgfr4* knockout embryos but could be increased in *fgfr1* knockout embryos. However, different contributions of individual FGFRs could result in differential effects on the expression of these FGF target genes. I attempted to perform this series of experiments upon the reopening of University laboratories, however unfortunately the embryos did not survive and there was insufficient time to repeat this. Nevertheless, building on this, qPCR analysis of FGF target genes would enable a quantitative comparison of the contributions of FGFR1, FGFR4 and FGFR1 in regulating these genes.

In situ hybridisations would investigate the role of different FGFRs in mesodermal and neuronal cell development, using probes I had generated to study the effect of knocking out specific FGFRs on various markers of these cell types. Generated probes include *sox3* (early neural), *rasl11b* (putative target of FGF signalling), *n-tubulin* (neurons), *en2* (midbrain-hindbrain junction) and *cdx4* (paraxial mesoderm), *myod* (skeletal muscle) and *krox20* (rhombomeres 3 and 5). Genes identified in the transcriptomic analysis, including *egr1*, *fos*, *junb*, *notch3* and *spry2*, could also be investigated.

RNA-Seq analysis of CRISPR/Cas9 knockout *X. tropicalis* embryos could provide an insight into the transcriptomes regulated by individual FGFRs, using targeted genome manipulation (McQueen and Pownall 2017). FGF signalling inhibition would be confirmed using phenotypic analysis and western blots for dpERK levels on sibling embryos. Targeting individual receptors in embryos would reveal overlapping functions of FGFRs in terms of regulating FGF target

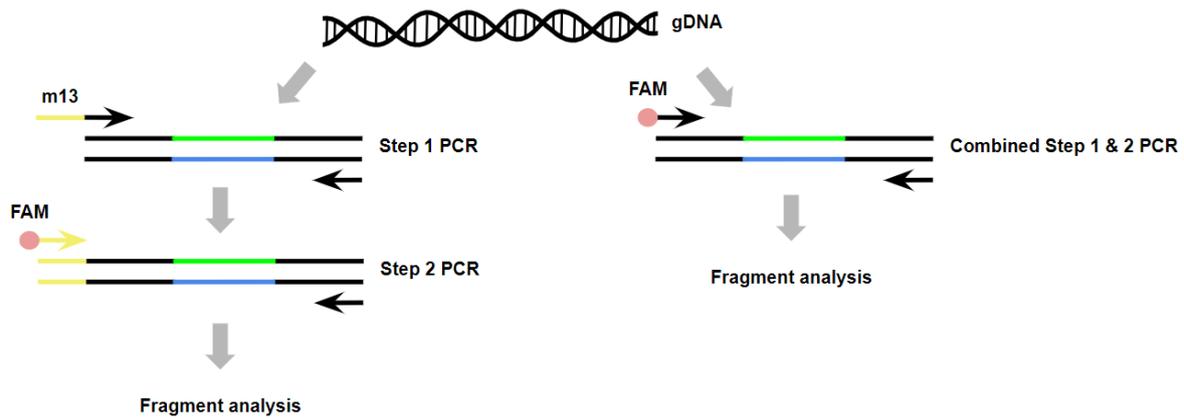


Figure 42: Comparison of 1 step and 2 step PCR protocols for fragment analysis. 2 Step PCR: Step 1 PCR involves genomic DNA extracted from injected embryos and a forward (F) primer with a 5'-m13 overhang and a reverse (R) primer which are complementary to the DNA sequence which has been targeted by CRISPR/Cas9. Step 2 PCR involves a common fluorescent m13 primer and the reverse (R) primer used in step 1 PCR, to generate fluorescently tagged PCR fragments, to be analysed using fragment analysis. **1 Step PCR:** PCR involves genomic DNA extracted from injected embryos and a forward (F) primer, which is FAM labelled, and a reverse (R) primer which are complementary to the DNA sequence which has been targeted by CRISPR/Cas9. This directly generates fluorescently tagged PCR fragments, to be analysed using fragment analysis.

gene expression and also whether signalling through individual receptors regulates the expression of other receptors in feedback loops. Whereas, the targeting of multiple FGFRs could also reveal redundant functions. If these RNA-Seqs are performed at the same time, PCA can be carried out, to determine which transcriptomes regulated by different FGFRs are the most similar. Heatmaps could also be generated to visualise the magnitude of expression change between treatments and controls in colour. These analyses were not possible in this study, since the data sets were generated with different protocols and different transcripts of genes have been identified in each data set.

The CRISPR/Cas9 protocol presented in this thesis enables the generation and genotyping of F0 embryos, which phenocopy previously characterised MO morphant phenotypes (Bhattacharya et al. 2015). F0 generation analysis enables rapid investigation into gene function, without the need for the establishment of mutant lines from breeding and outcrossing, which is time and space intensive (McQueen and Pownall 2017). However, mutations have been shown to be heritable (Blitz et al. 2013; Guo et al. 2014; Yen et al. 2014; Aslan et al. 2017) and potential off target effects of sgRNAs could be bred out in mutant lines (Nakayama et al. 2013). Furthermore, fragment analysis is an effective tool to identify indels, which are stable, as shown by the detection of consistent peaks in adult CRISPR/Cas9 frogs (Bhattacharya et al. 2015). Future fragment analysis should include a 1 step PCR protocol,

whereby the forward primer for the is FAM labelled (Figure 42), which would be more time and cost effective.

4.3.9 Conclusions and implications

This thesis presents evidence towards the unique roles of FGFRs in development through bioinformatic analyses, which also highlights the varying sensitivities to different methods of manipulating FGF signalling through individual FGFRs. CRISPR/Cas9 utilises a more targeted approach and a novel protocol has been devised in this study to generate F0 knockout embryos for analysis, exploiting targeted disruption and less complex design than previously employed approaches, such as zinc finger nucleases and TALENs (Guo et al. 2014). This gene-editing technique could be expanded to investigate the individual roles of FGF ligands in early *Xenopus* development.

References

- Amaya, E. et al. (1993). FGF signalling in the early specification of mesoderm in *Xenopus*. *Development*, 118 (2), pp.477–487.
- Amaya, E., Musci, T. J. and Kirschner, M. W. (1991). Expression of a Dominant Negative Mutant of the FGF Receptor Disrupts Mesoderm Formation in *Xenopus* Embryos. *Cell*, 66, pp.257–270.
- Ando, H. et al. (2007). Approaches to identify inhibitors of melanin biosynthesis via the quality control of tyrosinase. *The Journal of investigative dermatology*, 127 (4), pp.751–761.
- Aslan, Y. et al. (2017). High-efficiency non-mosaic CRISPR-mediated knock-in and indel mutation in F0 *Xenopus*. *Development*, 144 (15), pp.2852–2858.
- Baertschi, S., Zhuang, L. and Trueb, B. (2007). Mice with a targeted disruption of the *Fgfr1* gene die at birth due to alterations in the diaphragm: *Fgfr1* deficient mice die at birth. *The FEBS journal*, 274 (23), pp.6241–6253.
- Bahrami, S. and Drabløs, F. (2016). Gene regulation in the immediate-early response process. *Advances in biological regulation*, 62, pp.37–49.
- Bertrand, S. et al. (2009). FGFR1 is a neglected putative actor of the FGF signalling pathway present in all major metazoan phyla. *BMC evolutionary biology*, 9, p.226.
- Bhattacharya, D. et al. (2015). CRISPR/Cas9: An inexpensive, efficient loss of function tool to screen human disease genes in *Xenopus*. *Developmental biology*, 408 (2), pp.196–204.
- Birsoy, B. et al. (2006). Vg 1 is an essential signaling molecule in *Xenopus* development. *Development*, 133 (1), pp.15–20.
- Blitz, I. L. et al. (2013). Biallelic genome modification in F(0) *Xenopus tropicalis* embryos using the CRISPR/Cas system. *Genesis*, 51 (12), pp.827–834.
- Blum, M. et al. (2009). *Xenopus*, an ideal model system to study vertebrate left-right asymmetry. *Developmental dynamics: an official publication of the American Association of Anatomists*, 238 (6), pp.1215–1225.
- Böttcher, R. T. et al. (2004). The transmembrane protein XFLRT3 forms a complex with FGF receptors and promotes FGF signalling. *Nature cell biology*, 6 (1), pp.38–44.
- Böttcher, R. T. and Niehrs, C. (2005). Fibroblast growth factor signaling during early vertebrate development. *Endocrine reviews*, 26 (1), pp.63–77.
- Branney, P. A. et al. (2009). Characterisation of the fibroblast growth factor dependent transcriptome in early development. *PloS one*, 4 (3), p.e4951.
- Brunsdon, H. and Isaacs, H. V. (2020). A comparative analysis of fibroblast growth factor receptor signalling during *Xenopus* development. *Biology of the cell / under the auspices of the European Cell Biology Organization*. [Online]. Available at: doi:10.1111/boc.201900089.
- de Castro, E. et al. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research*, 34 (Web Server issue), pp.W362–W365.
- Catela, C. et al. (2009). Multiple congenital malformations of Wolf-Hirschhorn syndrome are recapitulated in *Fgfr1* null mice. *Disease models & mechanisms*, 2 (5-6), pp.283–294.
- Chan, W. Y. I. et al. (2007). The paralogous hematopoietic regulators *Lyl1* and *Scl* are coregulated by

Ets and GATA factors, but Lyl1 cannot rescue the early Scl^{-/-} phenotype. *Blood*, 109 (5), pp.1908–1916.

Cowell, Laura May-Seen (2019) Investigation of the role of Capicua in the FGF signalling pathway and the wound response. MSc by research thesis, University of York.

Dale, L., Matthews, G. and Colman, A. (1993). Secretion and mesoderm-inducing activity of the TGF-beta-related domain of *Xenopus* Vg1. *The EMBO journal*, 12 (12), pp.4471–4480.

Delaune, E., Lemaire, P. and Kodjabachian, L. (2005). Neural induction in *Xenopus* requires early FGF signalling in addition to BMP inhibition. *Development*, 132 (2), pp.299–310.

Ding, Y. et al. (2018). Bighead is a Wnt antagonist secreted by the *Xenopus* Spemann organizer that promotes Lrp6 endocytosis. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (39), pp.E9135–E9144.

Dorey, K. and Amaya, E. (2010). FGF signalling: diverse roles during early vertebrate embryogenesis. *Development*, 137 (22), pp.3731–3742.

Doudna, J. A. and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346 (6213), p.1258096.

Emerson, S. E. et al. (2017). Identification of target genes downstream of semaphorin6A/PlexinA2 signaling in zebrafish. *Developmental dynamics: an official publication of the American Association of Anatomists*, 246 (7), pp.539–549.

Faas, L. and Isaacs, H. V. (2009). Overlapping functions of Cdx1, Cdx2, and Cdx4 in the development of the amphibian *Xenopus tropicalis*. *Developmental dynamics: an official publication of the American Association of Anatomists*, 238 (4), pp.835–852.

Fisher, M. E., Isaacs, H. V. and Pownall, M. E. (2002). eFGF is required for activation of XmyoD expression in the myogenic cell lineage of *Xenopus laevis*. *Development*, 129 (6), pp.1307–1315.

Fletcher, R. B., Baker, J. C. and Harland, R. M. (2006). FGF8 spliceforms mediate early mesoderm and posterior neural tissue formation in *Xenopus*. *Development*, 133 (9), pp.1703–1714.

Freeman, S. D. et al. (2008). Extracellular regulation of developmental cell signaling by XtSulf1. *Developmental biology*, 320 (2), pp.436–445.

Fürthauer, M. et al. (2004). Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. *Development*, 131 (12), pp.2853–2864.

Gentsch, G. E. et al. (2018). Innate Immune Response and Off-Target Mis-splicing Are Common Morpholino-Induced Side Effects in *Xenopus*. *Developmental cell*, 44 (5), pp.597–610.e10.

Gerber, S. D. et al. (2009). The murine Fgfr1 receptor is essential for the development of the metanephric kidney. *Developmental biology*, 335 (1), pp.106–119.

Gerber, S. D. et al. (2020). Functional domains of the FgfrL1 receptor. *Developmental biology*, 461 (1), pp.43–54.

Gillespie, L. L., Chen, G. and Paterno, G. D. (1995). Cloning of a fibroblast growth factor receptor 1 splice variant from *Xenopus* embryos that lacks a protein kinase C site important for the regulation of receptor activity. *The Journal of biological chemistry*, 270 (39), pp.22758–22763.

Goetz, R. and Mohammadi, M. (2013). Exploring mechanisms of FGF signalling through the lens of structural biology. *Nature reviews. Molecular cell biology*, 14 (3), pp.166–180.

Goldfarb, M. et al. (2007). Fibroblast growth factor homologous factors control neuronal excitability through modulation of voltage-gated sodium channels. *Neuron*, 55 (3), pp.449–463.

- Golub, R. et al. (2000). Evolutionarily conserved and divergent expression of members of the FGF receptor family among vertebrate embryos, as revealed by FGFR expression patterns in *Xenopus*. *Development genes and evolution*, 210 (7), pp.345–357.
- Gospodarowicz, D. (1974). Localisation of a fibroblast growth factor and its effect alone and with hydrocortisone on 3T3 cell growth. *Nature*, 249 (453), pp.123–127.
- Gudernova, I. et al. (2016). Multikinase activity of fibroblast growth factor receptor (FGFR) inhibitors SU5402, PD173074, AZD1480, AZD4547 and BGJ398 compromises the use of small chemicals targeting FGFR catalytic activity for therapy of short-stature syndromes. *Human molecular genetics*, 25 (1), pp.9–23.
- Guo, X. et al. (2014). Efficient RNA/Cas9-mediated genome editing in *Xenopus tropicalis*. *Development*, 141 (3), pp.707–714.
- Hanafusa, H., Matsumoto, K. and Nishida, E. (2009). Regulation of ERK activity duration by Sprouty contributes to dorsoventral patterning. *Nature cell biology*, 11 (1), pp.106–109.
- Hardcastle, Z., Chalmers, A. D. and Papalopulu, N. (2000). FGF-8 stimulates neuronal differentiation through FGFR-4a and interferes with mesoderm induction in *Xenopus* embryos. *Current biology: CB*, 10 (23), pp.1511–1514.
- Harland, R. M. (1991). In situ hybridization: an improved whole-mount method for *Xenopus* embryos. *Methods in cell biology*, 36, pp.685–695.
- Hayashi, S. et al. (2004). Expression patterns of *Xenopus* FGF receptor-like 1/nou-darake in early *Xenopus* development resemble those of planarian nou-darake and *Xenopus* FGF8. *Developmental dynamics: an official publication of the American Association of Anatomists*, 230 (4), pp.700–707.
- He, H. et al. (2018). Study on the mechanism behind lncRNA MEG3 affecting clear cell renal cell carcinoma by regulating miR-7/RASL11B signaling. *Journal of cellular physiology*, 233 (12), pp.9503–9515.
- Hemmati-Brivanlou, A. et al. (1991). Cephalic expression and molecular characterization of *Xenopus* En-2. *Development*, 111 (3), pp.715–724.
- Holzmann, K. et al. (2012). Alternative Splicing of Fibroblast Growth Factor Receptor IgIII Loops in Cancer. *Journal of nucleic acids*, 2012, p.950508.
- Hongo, I., Kengaku, M. and Okamoto, H. (1999). FGF signaling and the anterior neural induction in *Xenopus*. *Developmental biology*, 216 (2), pp.561–581.
- Hori, K., Sen, A. and Artavanis-Tsakonas, S. (2013). Notch signaling at a glance. *Journal of cell science*, 126 (Pt 10), pp.2135–2140.
- Hou, S. et al. (2007). The secreted serine protease xHtrA1 stimulates long-range FGF signaling in the early *Xenopus* embryo. *Developmental cell*, 13 (2), pp.226–241.
- Isaacs, H. V., Deconinck, A. E. and Pownall, M. E. (2007). FGF4 regulates blood and muscle specification in *Xenopus laevis*. *Biology of the cell / under the auspices of the European Cell Biology Organization*, 99 (3), pp.165–173.
- Isaacs, H. V., Pownall, M. E. and Slack, J. M. (1994). eFGF regulates Xbra expression during *Xenopus* gastrulation. *The EMBO journal*, 13 (19), pp.4469–4481.
- Ishibashi, S., Cliffe, R. and Amaya, E. (2012). Highly efficient bi-allelic mutation rates using TALENs in *Xenopus tropicalis*. *Biology open*, 1 (12), pp.1273–1276.
- Ishitani, T. et al. (2003). The TAK1-NLK mitogen-activated protein kinase cascade functions in the

- Wnt-5a/Ca(2+) pathway to antagonize Wnt/beta-catenin signaling. *Molecular and cellular biology*, 23 (1), pp.131–139.
- Itoh, N. (2010). Hormone-like (endocrine) Fgfs: their evolutionary history and roles in development, metabolism, and disease. *Cell and tissue research*, 342 (1), pp.1–11.
- Itoh, N. and Ornitz, D. M. (2004). Evolution of the Fgf and Fgfr gene families. *Trends in genetics: TIG*, 20 (11), pp.563–569.
- Itoh, N. and Ornitz, D. M. (2008). Functional evolutionary history of the mouse Fgf gene family. *Developmental dynamics: an official publication of the American Association of Anatomists*, 237 (1), pp.18–27.
- Johnson, D. E. et al. (1991). The human fibroblast growth factor receptor genes: a common structural arrangement underlies the mechanisms for generating receptor forms that differ in their third immunoglobulin domain. *Molecular and cellular biology*, 11 (9), pp.4627–4634.
- Kalinina, J. et al. (2009). Homodimerization controls the fibroblast growth factor 9 subfamily's receptor binding and heparan sulfate-dependent diffusion in the extracellular matrix. *Molecular and cellular biology*, 29 (17), pp.4663–4678.
- Karimi, K. et al. (2018). Xenbase: a genomic, epigenomic and transcriptomic model organism database. *Nucleic acids research*, 46 (D1), pp.D861–D868.
- Khokha, M. K. et al. (2005). Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures. *Developmental cell*, 8 (3), pp.401–411.
- Kim, J. et al. (1998). Mesoderm induction by heterodimeric AP-1 (c-Jun and c-Fos) and its involvement in mesoderm formation through the embryonic fibroblast growth factor/Xbra autocatalytic loop during the early development of *Xenopus* embryos. *The Journal of biological chemistry*, 273 (3), pp.1542–1550.
- King, M. G. (2019) Investigating the role of Capicua in mediating FGF transcriptional regulation in *X. tropicalis*. PhD thesis, University of York.
- Kouhara, H. et al. (1997). A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *Cell*, 89 (5), pp.693–702.
- Kovalenko, D. et al. (2006). A role for extracellular and transmembrane domains of Sef in Sef-mediated inhibition of FGF signaling. *Cellular signalling*, 18 (11), pp.1958–1966.
- Kuroda, H. et al. (2005). Default neural induction: neuralization of dissociated *Xenopus* cells is mediated by Ras/MAPK activation. *Genes & development*, 19 (9), pp.1022–1027.
- Lai, J. K. H. et al. (2019). Induction of interferon-stimulated genes and cellular stress pathways by morpholinos in zebrafish. *Developmental biology*, 454 (1), pp.21–28.
- Lamas-Toranzo, I. et al. (2019). Strategies to reduce genetic mosaicism following CRISPR-mediated genome editing in bovine embryos. *Scientific reports*, 9 (1), p.14900.
- Lamb, T. M. and Harland, R. M. (1995). Fibroblast growth factor is a direct neural inducer, which combined with noggin generates anterior-posterior neural pattern. *Development*, 121 (11), pp.3627–3636.
- Lea, R. et al. (2009). Temporal and spatial expression of FGF ligands and receptors during *Xenopus* development. *Developmental dynamics: an official publication of the American Association of Anatomists*, 238 (6), pp.1467–1479.
- Lee, S.-Y. et al. (2011). Inhibition of FGF signaling converts dorsal mesoderm to ventral mesoderm in early *Xenopus* embryos. *Differentiation; research in biological diversity*, 82 (2), pp.99–107.

- Leerberg, D. M., Hopton, R. E. and Draper, B. W. (2019). Fibroblast Growth Factor Receptors Function Redundantly During Zebrafish Embryonic Development. *Genetics*, 212 (4), pp.1301–1319.
- Lei, Y. et al. (2012). Efficient targeted gene disruption in *Xenopus* embryos using engineered transcription activator-like effector nucleases (TALENs). *Proceedings of the National Academy of Sciences of the United States of America*, 109 (43), pp.17484–17489.
- Lewis, T. et al. (1995). XCL100, an inducible nuclear MAP kinase phosphatase from *Xenopus laevis*: its role in MAP kinase inactivation in differentiated cells and its expression during early development. *Journal of cell science*, 108 (Pt 8), pp.2885–2896.
- Louro, R. et al. (2004). RASL11A, member of a novel small monomeric GTPase gene family, is down-regulated in prostate tumors. *Biochemical and biophysical research communications*, 316 (3), pp.618–627.
- Makarenkova, H. P. et al. (2009). Differential interactions of FGFs with heparan sulfate control gradient formation and branching morphogenesis. *Science signaling*, 2 (88), p.ra55.
- Massagué, J. and Chen, Y. G. (2000). Controlling TGF-beta signaling. *Genes & development*, 14 (6), pp.627–644.
- Maude, Olivia (2019). Investigating Rasl11b as a novel target of FGF signalling in *Xenopus*. BSc dissertation, University of York.
- McQueen, C. and Pownall, M. E. (2017). An analysis of MyoD-dependent transcription using CRISPR/Cas9 gene targeting in *Xenopus tropicalis* embryos. *Mechanisms of development*, 146, pp.1–9.
- Mi, H. et al. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic acids research*, 47 (D1), pp.D419–D426.
- Mohammadi, M. et al. (1991). A tyrosine-phosphorylated carboxy-terminal peptide of the fibroblast growth factor receptor (Fg) is a binding site for the SH2 domain of phospholipase C-gamma 1. *Molecular and cellular biology*, 11 (10), pp.5068–5078.
- Mohammadi, M. et al. (1998). Crystal structure of an angiogenesis inhibitor bound to the FGF receptor tyrosine kinase domain. *The EMBO journal*, 17 (20), pp.5896–5904.
- Murphy, L. O. et al. (2002). Molecular interpretation of ERK signal duration by immediate early gene products. *Nature cell biology*, 4 (8), pp.556–564.
- Nakajima, K. and Yaoita, Y. (2015). Highly efficient gene knockout by injection of TALEN mRNAs into oocytes and host transfer in *Xenopus laevis*. *Biology open*, 4 (2), pp.180–185.
- Nakayama, K. et al. (2008). FGF induces oscillations of Hes1 expression and Ras/ERK activation. *Current biology: CB*, 18 (8), pp.R332–R334.
- Nakayama, T. et al. (2013). Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in *Xenopus tropicalis*. *Genesis*, 51 (12), pp.835–843.
- Nentwich, O. et al. (2009). Downstream of FGF during mesoderm formation in *Xenopus*: the roles of Elk-1 and Egr-1. *Developmental biology*, 336 (2), pp.313–326.
- Nicholson, K. M. and Anderson, N. G. (2002). The protein kinase B/Akt signalling pathway in human malignancy. *Cellular signalling*, 14 (5), pp.381–395.
- Nieuwkoop, P. D. and Faber, J. (Eds). (1994). Normal Table of *Xenopus Laevis* (Daudin) (Daudin : A Systematical and Chronological Survey of the Development from the Fertilized Egg Till the End of Metamorp). 1 edition. Routledge.

- Olsen, S. K. et al. (2006). Structural basis by which alternative splicing modulates the organizer activity of FGF8 in the brain. *Genes & development*, 20 (2), pp.185–198.
- Ong, S. H. et al. (2000). FRS2 proteins recruit intracellular signaling pathways by binding to diverse targets on fibroblast growth factor and nerve growth factor receptors. *Molecular and cellular biology*, 20 (3), pp.979–989.
- Ong, S. H. et al. (2001). Stimulation of phosphatidylinositol 3-kinase by fibroblast growth factor receptors is mediated by coordinated recruitment of multiple docking proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (11), pp.6074–6079.
- Orr-Urtreger, A. et al. (1993). Developmental localization of the splicing alternatives of fibroblast growth factor receptor-2 (FGFR2). *Developmental biology*, 158 (2), pp.475–486.
- Ossipova, O. et al. (2020). Pinhead signaling regulates mesoderm heterogeneity via the FGF receptor-dependent pathway. *Development*, 147 (17). [Online]. Available at: doi:10.1242/dev.188094.
- van Overbeek, M. et al. (2016). DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Molecular cell*, 63 (4), pp.633–646.
- Owens, N. D. L. et al. (2016). Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell reports*, 14 (3), pp.632–647.
- Paek, H., Gutin, G. and Hébert, J. M. (2009). FGF signaling is strictly required to maintain early telencephalic precursor cell survival. *Development*, 136 (14), pp.2457–2465.
- Paraiso, K. D. et al. (2019). Morpholinos Do Not Elicit an Innate Immune Response during Early Xenopus Embryogenesis. *Developmental cell*, 49 (4), pp.643–650.e3.
- Pera, E. M. et al. (2003). Integration of IGF, FGF, and anti-BMP signals via Smad1 phosphorylation in neural induction. *Genes & development*, 17 (24), pp.3023–3028.
- Popp, M. W. and Maquat, L. E. (2016). Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. *Cell*, 165 (6), pp.1319–1322.
- Pownall, M. E. et al. (1996). eFGF, Xcad3 and Hox genes form a molecular pathway that establishes the anteroposterior axis in Xenopus. *Development*, 122 (12), pp.3881–3892.
- Pownall, M. E. et al. (2003). An inducible system for the study of FGF signalling in early amphibian development. *Developmental biology*, 256 (1), pp.89–99.
- Raible, F. and Brand, M. (2001). Tight transcriptional control of the ETS domain factors Erm and Pea3 by Fgf signaling during early zebrafish development. *Mechanisms of development*, 107 (1-2), pp.105–117.
- Rebagliati, M. R. et al. (1985). Identification and cloning of localized maternal RNAs from Xenopus eggs. *Cell*, 42 (3), pp.769–777.
- Reece-Hoyes, J. S., Keenan, I. D. and Isaacs, H. V. (2002). Cloning and expression of the Cdx family from the frog *Xenopus tropicalis*. *Developmental dynamics: an official publication of the American Association of Anatomists*, 223 (1), pp.134–140.
- Reis, A. H. and Sokol, S. Y. (2020). Rspo2 antagonizes FGF signaling during vertebrate mesoderm formation and patterning. *Development*, 147 (10). [Online]. Available at: doi:10.1242/dev.189324.
- Saka, Y., Tada, M. and Smith, J. C. (2000). A screen for targets of the Xenopus T-box gene Xbra. *Mechanisms of development*, 93 (1-2), pp.27–39.
- Sapkota, G. et al. (2007). Balancing BMP signaling through integrated inputs into the Smad1 linker.

Molecular cell, 25 (3), pp.441–454.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology*, 18 (2), pp.233–234.

Schulte-Merker, S. and Smith, J. C. (1995). Mesoderm formation in response to Brachyury requires FGF signalling. *Current biology: CB*, 5 (1), pp.62–67.

Sempou, E. et al. (2018). Candidate Heterotaxy Gene FGFR4 Is Essential for Patterning of the Left-Right Organizer in *Xenopus*. *Frontiers in physiology*, 9, p.1705.

Session, A. M. et al. (2016). Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*, 538 (7625), pp.336–343.

Silva, A.-C. et al. (2006). Developmental expression of Shisa-2 in *Xenopus laevis*. *The International journal of developmental biology*, 50 (6), pp.575–579.

Simon, R. et al. (2007). Analysis of Gene Expression Data Using BRB-Array Tools. *Cancer informatics*, 3, p.117693510700300022.

Sivak, J. M., Petersen, L. F. and Amaya, E. (2005). FGF signal interpretation is directed by Sprouty and Spred proteins during mesoderm formation. *Developmental cell*, 8 (5), pp.689–701.

Slack, J. M. et al. (1987). Mesoderm induction in early *Xenopus* embryos by heparin-binding growth factors. *Nature*, 326 (6109), pp.197–200.

Sleeman, M. et al. (2001). Identification of a new fibroblast growth factor receptor, FGFR5. *Gene*, 271 (2), pp.171–182.

Steinberg, F. et al. (2010). The FGFR1 receptor is shed from cell membranes, binds fibroblast growth factors (FGFs), and antagonizes FGF signaling in *Xenopus* embryos. *The Journal of biological chemistry*, 285 (3), pp.2193–2202.

Szklarczyk, D. et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47 (D1), pp.D607–D613.

Tindall, A. J. et al. (2007). Expression of enzymes involved in thyroid hormone metabolism during the early development of *Xenopus tropicalis*. *Biology of the cell / under the auspices of the European Cell Biology Organization*, 99 (3), pp.151–163.

Trueb, B. et al. (2003). Characterization of FGFR1, a novel fibroblast growth factor (FGF) receptor preferentially expressed in skeletal tissues. *The Journal of biological chemistry*, 278 (36), pp.33857–33865.

Tsang, M. et al. (2002). Identification of Sef, a novel modulator of FGF signalling. *Nature cell biology*, 4 (2), pp.165–169.

Tsang, M. and Dawid, I. B. (2004). Promotion and attenuation of FGF signaling through the Ras-MAPK pathway. *Science's STKE: signal transduction knowledge environment*, 2004 (228), p.e17.

Ueno, H. et al. (1992). A truncated form of fibroblast growth factor receptor 1 inhibits signal transduction by multiple types of fibroblast growth factor receptor. *The Journal of biological chemistry*, 267 (3), pp.1470–1476.

Umbhauer, M. et al. (1995). Mesoderm induction in *Xenopus* caused by activation of MAP kinase. *Nature*, 376 (6535), pp.58–62.

Umbhauer, M. et al. (2000). Signaling specificities of fibroblast growth factor receptors in early *Xenopus* embryo. *Journal of cell science*, 113 (Pt 16), pp.2865–2875.

- Varshney, G. K. et al. (2015). High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome research*, 25 (7), pp.1030–1042.
- Wang, H. et al. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 153 (4), pp.910–918.
- Weeks, D. L. and Melton, D. A. (1987). A maternal mRNA localized to the vegetal hemisphere in *Xenopus* eggs codes for a growth factor related to TGF-beta. *Cell*, 51 (5), pp.861–867.
- Weinstein, M. et al. (1998). FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development*, 125 (18), pp.3615–3623.
- Xu, R. H. et al. (1999). Opposite effects of FGF and BMP-4 on embryonic blood formation: roles of PV.1 and GATA-2. *Developmental biology*, 208 (2), pp.352–361.
- Yamagishi, M. and Okamoto, H. (2010). Competition for ligands between FGFR1 and FGFR4 regulates *Xenopus* neural development. *The International journal of developmental biology*, 54 (1), pp.93–104.
- Yamaguchi, T. P. et al. (1994). *fgfr-1* is required for embryonic growth and mesodermal patterning during mouse gastrulation. *Genes & development*, 8 (24), pp.3032–3044.
- Yang, Z. et al. (2015). Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic acids research*, 43 (9), p.e59.
- Yayon, A. et al. (1991). Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell*, 64 (4), pp.841–848.
- Yen, S.-T. et al. (2014). Somatic mosaicism and allele complexity induced by CRISPR/Cas9 RNA injections in mouse zygotes. *Developmental biology*, 393 (1), pp.3–9.
- Yu, S. R. et al. (2009). *Fgf8* morphogen gradient forms by a source-sink mechanism with freely diffusing molecules. *Nature*, 461 (7263), pp.533–536.
- Zakrzewska, M., Marcinkowska, E. and Wiedlocha, A. (2008). FGF-1: from biology through engineering to potential medical applications. *Critical reviews in clinical laboratory sciences*, 45 (1), pp.91–135.
- Zhang, X. et al. (2006). Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. *The Journal of biological chemistry*, 281 (23), pp.15694–15700.
- Zhao, P. et al. (2006). *Fgfr4* is required for effective muscle regeneration in vivo. Delineation of a MyoD-Tead2-*Fgfr4* transcriptional pathway. *The Journal of biological chemistry*, 281 (1), pp.429–438.
- Zilinskaite, Vida (2019). An analysis of FGF receptor function in amphibian development using CRISPR/Cas9 mediated gene editing. Master's thesis, University of York.
- Zimmerman, L. B., De Jesús-Escobar, J. M. and Harland, R. M. (1996). The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell*, 86 (4), pp.599–606.
- Zygar, C. A., Cook, T. L. and Grainger, R. M., Jr. (1998). Gene activation during early stages of lens induction in *Xenopus*. *Development*, 125 (17), pp.3509–3519.

Abbreviations

Abhd15	Abhydrolase domain containing 15
Agr	Anterior gradient
AP-1	Activator protein 1
Atp1b2	ATPase, Na ⁺ /K ⁺ transporting, beta 2 polypeptide
bFGF	Basic fibroblast growth factor (FGF2)
bHLH	Basic helix-loop-helix
BMP	Bone morphogenetic protein
BMPR	Bone morphogenetic protein receptor
Cdx	Caudal type homeobox
Chrd	Chordin
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9
crRNA	CRISPR ribonucleic acid
DAG	Diacylglycerol
dnFGFR	Dominant negative fibroblast growth factor receptor
dpERK	Diphosphorylated extracellular signal-regulated kinase
DUSP	Dual specificity phosphatase
Dynll1	Dynein light chain LC8-type 1
eFGF	Embryonic fibroblast growth factor (FGF4)
Egr1	Early growth response 1
ERK	Extracellular signal-regulated kinase
Erm	ETS related molecule
FCS	Foetal calf serum
FDR	False discovery rate
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FGFRL1	Fibroblast growth factor receptor like 1
FLRT3	Fibronectin leucine rich transmembrane protein 3

FPKM	Fragments per kilobase of transcript per million mapped reads
FRS2 α	FGFR substrate 2 alpha
Frzb	Frizzled
Fzd4	Frizzled class receptor 4
GDP	Guanosine diphosphate
GO	Gene ontology
Grb2	Growth factor receptor bound protein
GRP	Gastrocoel roof plate
Gsc	Gooseoid
GTP	Guanosine triphosphate
hCG	Human chorionic gonadotropin
hFGF	Hormone-like fibroblast growth factor
HSPG	Heparan sulphate proteoglycan
iFGF	Intracellular fibroblast growth factor
iFGFR	Inducible fibroblast growth factor receptor
Ig	Immunoglobulin
IGF	Insulin-like growth factor
Indel	Insertions or deletions of bases
IP ₃	Inositol-1,4,5-trisphosphate
Irg1	Aconitate decarboxylase 1 like gene b (acod1lb)
Krt-b	Keratin 70 (krt70)
KV	Kupffer's vesicle
LRO	Left-right organiser
MAB	Maleic acid buffer
MBT	Mifblastula transition
MAPK	Mitogen-activated protein kinase
MKP	MAP kinase phosphatase
MO	Antisense morpholino oligonucleotides
MRS	Modified Ringers Solution

Mylc2a	Myosin light chain 7
MyoD	Myogenic differentiation
Nek6	NIMA-related kinase 6
NHEJ	Nonhomologous end-joining
Nik	Nemo-like kinase
PAM	Protospacer adjacent motif
PANTHER	Protein ANalysis Through Evolutionary Relationships
PCA	Principle Component Analysis
Pdgfa	Platelet derived growth factor subunit A
Pfkfb3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
PI3K	Phosphoinositide-3 kinase
PIP ₂	Phosphatidylinositol-4,5-diphosphate
Pea3	Polyoma enhancer activator 3
PKB	Protein kinase B
PKC	Protein kinase C
PLC γ	Phospholipase C gamma
Pnhd	Pinhead
PPI	Protein-protein interaction
pSMAD	Phosphorylated SMAD
PTB	Phosphotyrosine-binding
Rasl11b	Ras-like protein, family 11, member B
Rspo2	R-spondin 2
SCL	Stem cell leukemia
sgRNA	Single guide ribonucleic acid
SH3	Src homology 3
SOS	Son of sevenless
Spry	Sprouty
SRF	Serum response factor
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins

TALEN	Transcription activator-like effector nucleases
Tbxt	Brachyury
TGF β	Transforming growth factor beta
TPE	Transcripts per embryo
Tubb2b	N-tubulin