Investigating the role of *Capicua* in mediating FGF transcriptional regulation in *X. tropicalis*

Michael Gavin King

PhD

University of York

Biology

December 2018

Abstract

The Fibroblast Growth Factor (FGF) family of secreted peptides signal through tyrosine kinase receptors (RTKs), leading to the downstream activation of the MAP kinase (MAPK), causing transcriptional change. FGF signalling plays multiple critical roles in the maintenance of mesoderm induction, regulation of differentiation and patterning throughout early embryonic development. Despite extensive research on the FGF dependent transcriptome, there are still gaps in are understanding of gene regulation downstream of FGF signalling.

The aim of this project was to address the gaps in knowledge by investigating a novel regulator of transcription, Capicua (CIC). *D. melanogaster* research has revealed that CIC functions as a transcriptional repressor downstream of RTKs, EGFR and Torso, regulated by Ras-MAPK signal transduction. This project's focus was on establishing if CIC is involved as a transcriptional repressor downstream of FGF signalling, functioning in a similar fashion to EGFR and Torso gene transcription during early amphibian development.

The work in this thesis has established the gene structure of *CIC* in *X. tropicalis,* allowing the analysis of temporal and spatial expression profiles of the prominent isoforms of *CIC*. Analysis has shown that CIC-L is expressed in the maternal phase of embryonic development, whilst CIC-S is zygotically expressed. Knockout of *CIC* in *X. tropicalis* embryos lead to the loss and truncation of anterior structures along the anterior-posterior axis, similar to FGF4 or FGF8 overexpression phenotypes. Treating embryos with FGF4/FGF8 leads to the degradation of the CIC protein. Transcriptome analysis of *CIC* knockdown and *FGF4* overexpression in early development has revealed a subset of FGF regulated genes, which are regulated through inhibition of CIC. Given the deep conservation of developmental mechanisms within vertebrate species and the increasing evidence linking both FGF signalling and the function of CIC in human health and disease, the output of this investigation will have a wide significance.

Table of Contents

List of Figures

List of Tables

List of Accompanying Material

CD containing:

Video of movement phenotypes of CIC-L and CIC-S targeted by morpholinos.

1. CIC-L targeting morpholino injections.
2. CIC-S targeting morpholino injections.
3. CIC-L & CIC-S targeting morpholino injections.
4. Uninjected control.

Annotated *CIC* gene *X. tropicalis* sequence with Snapgene viewer software.

Annotated cDNA sequence of the *X. tropicalis* prominent isoforms *CIC*.

Annotated cDNA sequence of CIC-L

Annotated cDNA sequence of CIC-S

Alignment of *H. sapiens* and *M. musculus* amino acid sequence.

Alignment of isoform of CIC without exon 21

Supplementary Tables:

Spreadsheet of RNA-seq results.

Acknowledgements

I would like to start by thanking my supervisor's Harv and Paul for all their many hours of supervision, advice and expertise. I appreciated all of their feedback throughout my PhD on my various reports and presentations. I would also like to thank them for their support, enthusiasm and commiserations throughout this project. I would like to thank my TAP panel members, Bob and Pegine for their help and guidance during my PhD.

I give thanks to all present and past members of the *Xenopus*/Fish and Genever labs for sharing their friendship, enthusiasm and knowledge along my PhD journey. Amanda, Caitlin, Gideon, Alley and Lewis, I could not ask for a nicer bunch of people and our lab chats will be missed. I would like to thank Betsy and Di for sharing their vast amount of knowledge for every conceivable protocol in the lab. I would also like to thank the support of the staff at the University of York technology facility for all their help with the RNA-seq experiment and all the other bits and bobs of guidance over the years.

I would also like to thank Sue Jones and the team at York St John's for giving me the opportunity to spend time with them experiencing first-hand what it is feels like to be an academic, lecturing and developing material for modules. It was a wonderful experience to be around such a passionate team of people who love their job.

I would like to give a massive thank you to Samantha who has been my rock. She has supported me and has been willing to sacrifice to support me throughout my PhD, through the up and downs and has never stopped believing in me when I did not believe in myself and without her completing the PhD would be impossible. Last of all, I would like to thank my children, Emily and Cavan. They will never know how much them being themselves, their unbound humour and joy to be around has lifted me when times were hard. Most of all I would like to thank my Nan for her uncompromising attitude to life. The security and warmth she provided in some of my darkest days, helping me more than she could have ever known. I had hoped to have completed my PhD so she could attend my graduation but sadly she lost her fight to cancer. I'd like to dedicate this thesis to her memory.

Declaration

I declare that this is an original piece of work conducted under the supervision of H.V. Isaacs and P.G. Genever at the University of York. All the data presented is my own work. All sources are acknowledged as references. Processing of RNA-seq was undertaken by staff at University of York Biology Technology Facility. Processing of the raw data into TPM values was undertaken by John Davey at the University of York Biology Technology Facility.

None of the work presented in this thesis has been previously published or submitted for a qualification either at the University of York or at any other institution. However, some parts of this work have been presented in poster at the 16th (2016) International *Xenopus* conferences (Crete, Chania).

Chapter 1: Introduction

1.1 Fibroblast growth factors

Armelin observed in 1973, whilst looking at the extracts from bovine brains, that brain extracts stimulated growth in *M. musculus* fibroblasts (Armelin, 1973). It was later observed in 1975, that mitogenic growth agent was isolated in the pituitary gland (Gospodarowicz, 1975). This in turn lead to the discovery of the first of the FGF proteins which was acidic fibroblast growth factor (FGF), now known as FGF1 and soon after FGF2 (Basic FGF) which was identified (Esch et al., 1985; Gospodarowicz et al., 1984). Early studies analysing FGF proteins found that they had a mitogenic effect in human foreskin fibroblasts and *M. musculus* fibroblasts maintained in serum. FGF is named for the effects it has on human foreskin fibroblasts and *M. musculus* fibroblasts in creating a two fold increase in cell numbers in primary cultures in comparison to non-treated controls (Gospodarowicz and Moran, 1975).

FGFs are secreted signalling ligands, which have many important roles in early development. FGFs are required to regulate anterior–posterior patterning, control the organisation of mesoderm during gastrulation, influencing cell morphogenesis/movement, neural induction/patterning in the brain and controlling limb development (Amaya et al., 1993; Christen and Slack, 1997a, 1999; Cornell et al., 1995; Delaune et al., 2005; Kim et al., 1998; Slack et al., 1987; Slack et al., 1992). FGFs are also important for the regulation of homeostasis, angiogenesis and wound healing in the adult (Bottcher and Niehrs, 2005). Due to FGF multiple actions on multiple cell types they are often referred to as "pluripotent" or "promiscuous" growth factors (Green et al., 1996).

1.1.1 FGF ligands and receptors.

There are 22 members in the FGF ligand family found in vertebrates and these can be grouped into seven subfamilies which are dependent on their function (Itoh and Ornitz, 2004). FGFs function in a intracrine, paracrine, and endocrine manner (fig 1A) (Itoh et al., 2015). The intracrine-acting FGF11 subfamily of ligands; FGF11, FGF12, FGF13 and FGF14 do not activate FGFRs (Beenken and Mohammadi, 2009). The FGF proteins are highly evolutionary conserved throughout the animal kingdom (Itoh and Ornitz, 2004). FGFs have a high affinity to Heparan sulphate, due to their conserved core of 120-130 amino acids (fig 1B) which leads to them binding with their receptor (Beenken and Mohammadi, 2009; Ornitz and Itoh, 2001).
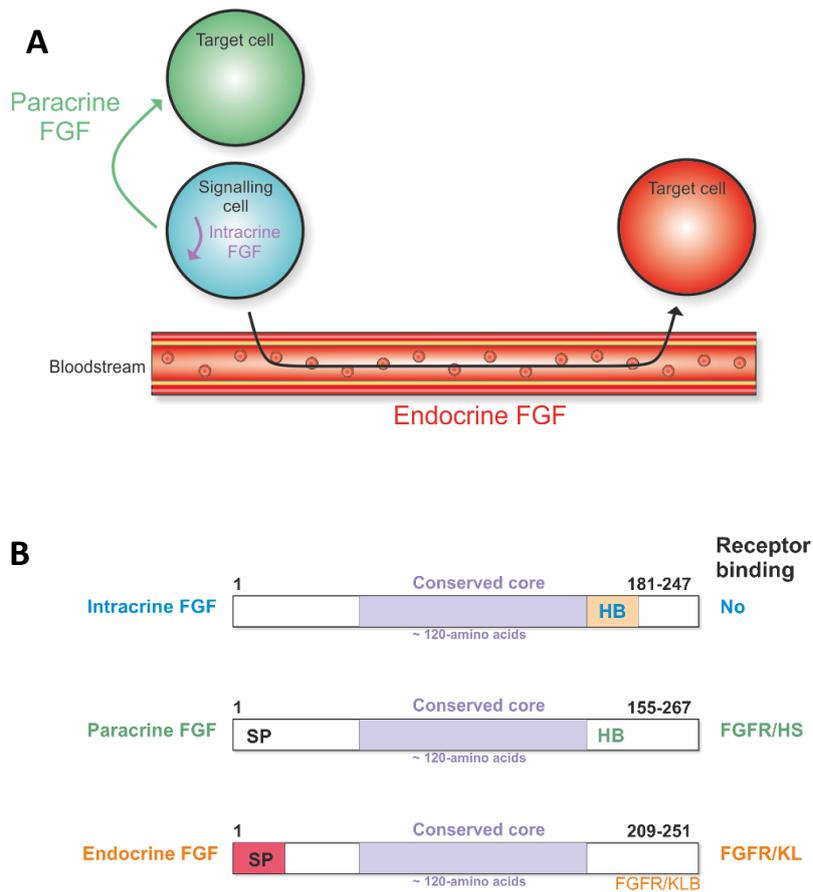
Figure 1, (A) FGFs act in Endocrine, Paracrine and Intracrine manner. Paracrine FGFs are locally secreted that target nearby cells by diffusion. Endocrine FGFs are secreted and target cells through the bloodstream. (B) Schematic diagram of the endocrine, paracrine and intracrine FGF proteins with secreted signals sequences (SP), heparan sulfate-binding site (HB) and and Klotho-binding site.

Heparan sulphate increases the affinity of the FGF ligands to their FGF receptors (FGFR) (table 1). FGFRs are cell surface proteins which belong to a subfamily of receptor tyrosine kinases (RTK) which contain a heparin-binding sequence. The four genes are each spliced to produce multiple isoforms of each receptor. FGF ligands bind to the FGFRs 3 immunoglobulin-like domains Ig1, Ig2 and Ig3 (fig 2) (table 1) (Sakaguchi et al., 1999). FGFRs contain a hydrophobic transmembrane domain and which links the split intracellular tyrosine kinase domain to the extracellular ligand binding domain (fig 2). FGF ligands, along with Heparan sulphate glycosaminoglycan (HS-GAG) chains, bind to the FGFR, which cause the receptor to dimerise (fig 2, 3).

FGF subfamily	Ligands	Receptor preference
FGF1	FGF1, FGF2	FGF1 activates all FGFRs; FGF2 prefers FGFR1c and FGFR2c
FGF4	FGF4, FGF5, FGF6	FGFR1c, FGFR2c
FGF7	FGF3, FGF7, FGF10, FGF22	FGFR2b, FGFR1b
FGF8	FGF8, FGF17, FGF18	FGFR3c, FGFR4, FGFR1c
FGF9	FGF9, FGF16, FGF20	FGFR3c, FGFR2c
FGF19	FGF19, FGF21, FGF23	Hormone class, very weak activation of FGFR1c, FGFR2c
FGF11	FGF11, FGF12, FGF13, FGF14	No activation of FGFRs

Table 1 lists ligands and their receptor preferences. FGF ligands have different affinities for different FGFRs (Itoh and Ornitz, 2004; Zhang et al., 2006).
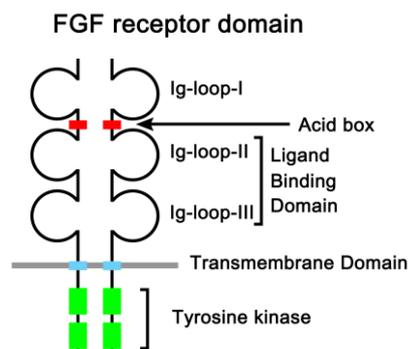
Figure 2, the FGFR protein contains 3 extracellular Ig-loop domains, an acid box, a transmembrane domain and an intracellular tyrosine kinase domain. Ig-II and Ig-III are responsible for the ligand binding. The acid box aids the binding of FGF ligands (Bottcher and Niehrs, 2005).

1.1.2 FGF transduction pathways

FGF ligand binding leads to dimerisation of the FGFR that leads to cross-phosphorylation of conserved tyrosine residues in the intracellular tyrosine kinase domain of the receptor and activates any of the three intracellular signal transduction pathways; Ras/MAPK, P13/AKT and PLC-γ (fig 3) (Thisse and Thisse, 2005).
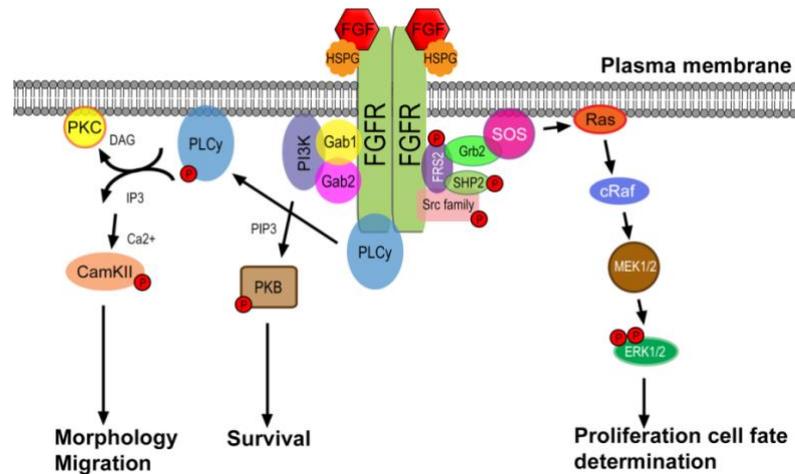
Figure 3, the transmembrane receptor of FGFR with the attached FGF ligand/HS-GAG and three transduction pathways; Ras/MAPK, P13/AKT and the PLC-Y.

The intracellular RTK domain phosphorylates the FGFR substrate 2 (FRS2), which recruits the Growth factor receptor-bound protein 2 (Grb2) adaptor and Son of Sevenless (SOS) proteins (fig 3) (Kouhara et al., 1997; Ong et al., 2000). The Grb2/SOS proteins trigger the activation of Ras, a GTP-binding protein leading to the activation of the Ras/MAPK signalling cascade. Ras activation leads to the phosphorylation cascade of Raf-Mek (MAPK kinase kinase) and finally phosphorylation of MAPK/ERK. MAPK signalling leads to changes in gene transcription (Umbhauer et al., 1995).

The Grb2 adaptor protein is also responsible for the activation of the phosphoinositide-3 kinase (PI3K) pathway through the GRB2-associated-binding protein 1 (GAB1) scaffold protein, which associates with Grb2 (fig 3). Activation of PI3K leads to the activation of the serine threonine kinase Akt/protein kinase B (PKB) (Nicholson and Anderson, 2002). Activation of the phospholipase C-Y (PLC-Y) pathway is caused by the dimerisation of FGFR and phosphorylation of a conserved phosphotyrosine residue (Mohammadi et al., 1991; Ueda et al., 1996). This leads to the phosphorylation of PLC-Y, which hydrolyses phosphotidykinositol-4, 5-diphosphate to inostiaol-1, 4, 5-trisphosphate (IP3) and diacylgycerol (DAG). The formation of DAG activates PKC, whilst IP3 leads to the release of calcium from the intracellular stores of the cell. Although the FGFR is responsible for 3 transduction pathways, this thesis will focus on the Ras-MAPK transduction signalling.

1.1.3 The role of FGF during mesoderm formation

Mesoderm formation is one of the earliest events in germ layer specification in the vertebrate body. Early experiments on *Xenopus laevis* demonstrated that FGF signalling was important for the induction of mesoderm (Slack et al., 1987). FGF is crucial for mesoderm induction in early axis formation of the vertebrate embryo. The first mesodermal inducer to be discovered was basic FGF (FGF2) (Slack et al., 1987; Westall et al., 1978). At least 13 of the 22 FGF genes are expressed in early development in the *Xenopus* embryo (Lea et al., 2009). The importance of FGFs in mesoderm induction was highlighted in several studies using a dominant negative FGFR (dn-FGFR). The dn-FGFR was created by mutating the FGFR gene, creating a receptor that lacked a functional intracellular kinase domain (fig 2). The dn-FGFR functions by forming dimers with the endogenous wild-type FGFR, blocking signalling. Overexpression of the dn-FGFR in *Xenopus* disrupted mesoderm formation, leading to an irregular body axis formation (Amaya et al., 1991). Similar studies using this approach also demonstrated that a subset of genes responsible for mesoderm induction were not expressed when FGF signalling was inhibited (Amaya et al., 1993; Isaacs et al., 1994; Schulte-Merker and Smith, 1995).

In early vertebrate development, two FGF ligand family members are present in the mesoderm forming stages; FGF2 (Slack and Isaacs, 1989) and FGF4 (Isaacs et al., 1992). Although FGF2 is present during gastrulation it is not as effectively secreted as FGF4 due to the lack of a signal sequence within the ligand (fig 1b). Unlike FGF2, FGF4 contains a secretion sequence, which is expressed maternally and has a significant increase of expression during gastrulation which suggests it has a role in induction and maintenance of mesoderm (Isaacs et al., 1994).

Brachyury (*Xbra*) is one such gene that requires FGF signalling for its expression during gastrulation (Smith et al., 1991). *Xbra* is a T-box transcription factor, which is first expressed during mid-blastula transition (MBT) and is required for mesoderm formation. *Xbra* forms an autocatalytic regulatory loop with FGF4 in the early mesoderm. When FGF induces expression of *Xbra*, this then leads to the maintenance of FGF4 expression in a feedback loop (Isaacs et al., 1994; Isaacs et al., 1995b; Schulte-Merker and Smith, 1995). When *Xenopus* embryos were treated with the SU5402 drug, a FGFR-specific tyrosine kinase activity inhibitor or with the dn-FGFR, expression of *Xbra* was lost (Fletcher and Harland, 2008; Isaacs et al., 1994).

In addition to *Xbra*, the caudal-related (Cdx) homeodomain transcription factors, which are part of a paraHox gene cluster, have their expression regulated by FGF signalling (Keenan et al., 2006b; Reece-Hoyes et al., 2002). They play an important role in the development of the vertebrate body axis and gut epithelium (Guo et al., 2004). *Cdx1*, *Cdx2* and *Cdx4* are first expressed in the developing mesoderm due to MAPK and PI3K signalling transduction through FGF (Branney et al., 2009). During gastrulation, only the FGF-MAPK transduction has been shown to regulate *Cdx* expression. Although these genes are known to be activated by FGF signalling the exact mechanisms leading to their transcription is still unknown.

Recent research has shown that if *Xenopus* embryos are treated with SU5402 before mesoderm induction, myogenin D (MyoD) and myogenic regularity factor 5 (myf5), myogenic regulatory factors (MRF) proteins, which regulate myogenesis, have decreased expression during the formation of the early paraxial mesoderm (Weintraub, 1993). Mesoderm patterning not only needs to be induced early in gastrulation but needs to be maintained by FGF.

While the transforming growth factor beta (TGF-β) family of proteins (*Xenopus* nodal related (Xnr1, 2, 4 & Vg1), are inducers of mesoderm formation, FGF acts as a competence factor, giving cells the ability to respond to specific inductive signal (Cornell et al., 1995). FGF is required for activin-mediated mesoderm induction (Cornell et al., 1995). Activin has been found to activate pathways down stream of *Xnr1*, *2*, *4* & *Vg1* and when the activin is blocked, the mesodermal markers are not expressed.

Early growth response 1 (Egr1) is a gene that encodes for a zinc-finger protein, which acts as a transcriptional regulator, binding to its DNA targets at promoters and is expressed throughout the marginal zone of the *Xenopus* embryo during gastrulation, much like *Xbra* (Panitz et al., 1998). Ets-box transcription factor Elk-1 can be phosphorylated by FGF MAPK transduction, which leads to formation of a protein complex with a dimer of serum response factor (SRF). This complex formation binds to a promoter upstream of the *Egr1* gene, allowing the expression of *Egr1.* Egr1 acts to repress *Xbra,* whilst at the same time, allows the expression of *MyoD*. This highlights diverse FGF functions in controlling differing cell fate decision in early development of the embryo.

In *Xenopus*, *Bone morphogenetic protein-4 (BMP-4)* expression is regulated by an auto-regulatory loop with *c-Jun* (Knöchel et al., 2000). *c-Jun*'s expression is regulated

by FGF (Kim et al., 1998; Knöchel et al., 2000). *c-fos*, like *Egr1,* is a downstream target of Elk-1 (Babu et al., 2000). Elk-1 stimulation by MAPK transduction lead to its phosphorylation and binding to the SRF complex which leads to binding of the complex within the *c-fos* promoter. The binding of the Elk-1-SRF complex leads to *c-fos* expression (Besnard et al., 2011). c-Jun and c-Fos proteins form the heterodimer, Activator Protein-1 (AP-1) an early response transcription factor (Kim et al., 1998). The AP-1 complex is a DNA binding transcription factor functioning by binding to the palindromic DNA sequence, TGAC/GTCA. AP-1 is an activator of BMP-4 and is essential for ventral mesoderm formation (Knöchel et al., 2000). The family members of Jun and Fos can vary significantly and lead to changes in interactivity with other proteins, alternating function.

1.1.4 The role of FGF in neural induction

Neural induction, is the process in which the nervous system is created during early embryonic development. Embryonic cells which originate from ectoderm are induced to differentiate towards a neural fate rather than that of the epidermis and mesoderm. *Noggin* was the first neural inducer to be identified (Smith and Harland, 1992) and not long after *Chordin* was discovered (Sasai et al., 1995). Chordin and Noggin act to repress the BMP signalling by binding to BMP ligands preventing activation of the BMP receptor (Piccolo et al., 1996; Zimmerman et al., 1996b). This inhibition of BMP is necessary for neural induction and is known as the "default" model of neural induction (Muñoz-Sanjuán and Brivanlou, 2002). In several experiments involving *D. rerio*, *M. musculus*, *Xenopus* and *Gallus domesticus* embryo animal models, BMP has been shown to be a negative regulator of neural induction (Rogers et al., 2009). When the BMP receptor is inhibited, it allows naïve embryonic ectoderm to become neural tissue rather than epidermis. Inversely, if the expression of either Chordin or Noggin is blocked, allowing a maintained BMP signal, neural tissue such as the neural plate is prohibited from being formed (Wessely et al., 2004). When FGF signalling was blocked with the SU5402 drug, which inhibits the tyrosine kinase activity of the receptor, in *G. domesticus*, neural patterning was absent (Delaune et al., 2005; Furthauer et al., 2004; Mohammadi et al., 1997). Later studies found in *Xenopus*, FGF regulates the transcription of Chordin and Noggin organiser factors, which are BMP antagonists (Marchal et al., 2009). FGF signalling is required for BMP antagonists *Chordin* and *Noggin* expression (Branney et al., 2009; Delaune et al., 2005). This inhibition of BMP expression enables the formation of the dorsal-ventral axis in the embryo.

SMAD family member 1 (Smad1) is a key intracellular mediator of BMP signalling (Pera et al., 2003). SMAD1 functions downstream of the BMP receptor. When SMAD1 is phosphorylated at the serine residues of the carboxy-terminus, due to the BMP receptor kinase phosphorylation, it leads to the movement of the protein into the nucleus leading to gene transcription (Pera et al., 2003). Phosphorylation of SMAD1 by FGF-MAPK signalling at the linker region of the protein leads to the opposite effect, preventing SMAD1 from being internalised into the nucleus (Kretzschmar et al., 1997).

1.1.5 Other roles FGF in embryonic development

FGF also has a role in morphogenesis, the process in which cells change their cytoskeletal structure and behaviour in response to internal or external stimulus. This process allows cell movement during embryonic development. Convergent extension a process in which a sheet of cells in the embryo changes shape by narrowing in one direction and extending in another direction (Wallingford et al., 2002). The movement of cells during gastrulation is example of convergent extension. FGF plays a role in the morphogenetic movement through signalling of the PLC-γ pathway. Convergent extension is promoted by *Xbra,* an FGF target, which inhibits adhesion of fibronection in migrating cells (Kwan and Kirschner, 2003). *Xbra* promotes convergent extension by triggering the expression of *Xwnt11* (Tada and Smith, 2000) and *prickle* (Branney et al., 2009), both factors of the noncanonical Wnt pathway, known to regulate convergent extension. *XSprouty2* is a gene which encodes for a protein that acts as an inducible inhibitor, behaving as an antagonist of FGF-dependant calcium signalling. Sprouty proteins impedes the PKC pathway of FGF (Chow et al., 2009; Sivak et al., 2005), whilst allowing the Ras/MAPK pathway intact leading to mesoderm induction. Unlike sprouty, the Spred protein has the opposite effect, preventing the ras/MAPK pathway, whilst leaving the PKC pathway intact, allowing cells to undergo morphogenesis and cellular movement. FGF signalling in gastrulation has two functions, early MAPK signalling that signals specifies and maintains axial paraxial mesoderm and a later morphogenic role, which is not ERK-dependent leading to the moment of cells during gastrulation and neurulation.

Somitongenesis is the process by which somites form. Somites are mesodermal structures that form from balls of epithelial cells from the paraxial presegmental (or presomitic) mesoderm (PSM) (Dequeant and Pourquie, 2008). Somites give rise to the skeletal muscle, cartilage, vertebrae and dermis of the skin. The somites provide a framework structure for the migration of neural cells to form the neural crest and

allow the patterning of the nervous system. Somitongenesis occurs in an anterior to posterior direction with the formation of pairs of somites at regular intervals. FGF8 was discovered to be highly expressed in the PSM (Dequeant et al., 2006), with a suggested role at keeping the cells in the PSM in an immature undifferentiated state (Delfini et al., 2005). FGF MAPK transduction maintains cells in a highly mobile state in the posterior PSM. Due to a FGF gradient cells in the anterior, PSM are less motile becoming somites. This process of restriction of differentiation by FGF has also been found to occur in neural precursor cells which require depleted FGF signalling in order to undergo differentiation (Diez del Corral et al., 2002).

FGF signalling is involved in the patterning and neurogenesis of the eye in early development (Martinez-Morales et al., 2005; McFarlane et al., 1998). Gain of function studies which activate FGF-MAPK in a number of species show that retinal pigment epithelium (RPE) can be differentiated in to neural retina (Fuhrmann, 2010; Spence et al., 2007; Yoshii et al., 2007) and when FGF-MAPK signalling is blocked distal optic vesicle neural retina is differentiated into RPE (Cai et al., 2010). FGF does this by regulating the expression of *Etv1,* a Pea3/Etv4-subfamily of ETS-domain transcription factors by MAPK transduction. In Xenopus only Etv1 is expressed in the retina (Chen et al., 1999). *Etv1* expression takes place prior to retinal neurogenesis and is downregulated after differentiation retinal neuroepithelium.

1.1.6 FGF regulation of multipotency in mammalian stem cells

In other contexts, FGF has a regulatory role in the maintenance and differentiation of embryonic and somatic stem cells (Coutu and Galipeau, 2011; Kunath et al., 2007). FGF has been shown to play an important role in tissue homeostasis, repair, self-renewal and inhibition of cellular senescence (Coutu and Galipeau, 2011). In *H. sapiens* and *M. musculus*, FGF acts as a negative regulator of mesenchymal stem cell (MSC) senescence. FGF regulates cellular senescence, by activation of the PI3K/Akt transduction pathway, which leads to the phosphorylation and release of murine double minute 2 (Mdm2) from its inhibitor, resulting in the import of Mdm2 into the nucleus. The internalisation of Mdm2 into the nucleus allows ubiquitin-ligase activity, leading to proteasomal degradation of p53, an important regulator of cellular senescence (Qian and Chen, 2013; Wasylyk et al., 1999).

FGF has a mitogenic effect on a range of cell types (Oliver et al., 1990), owing to the Ras/MAPK signal transduction. FGF is also known to regulate MSCs by maintaining them in an undifferentiated state during proliferation in vitro, whilst permitting them to

be multipotent (Tsutsumi et al., 2001). For example, the FGFR2 inhibits osteoblast differentiation by inducing the expression of Sox2, a HMG-box DNA binding transcription factor, which antagonises Wnt signalling (Basu-Roy et al., 2010). Sox2 binding inhibits β-catenin, an intracellular signal protein involved in the Wnt signalling pathway.

Knockouts of both FGFR1 and FGFR2 have highlighted their importance in bone specific lineage pathways in MSCs in vitro (Eswarakumar et al., 2002), displaying decreases in osteoblast cell proliferation, shape and size (Verheyden et al., 2005). In other studies, FGFR2 knockout, resulted in severe dwarfism, due to reduced amounts of skeletal lineage cells (osteoblasts & chondrocytes) (Verheyden et al., 2005). This was a result of reduced osteoblast cell proliferation, demonstrating that FGFR2 has a key role in the proliferation of osteoprogenitor cells.

1.2 FGF in disease

1.2.1 Receptor misregulation

Misregulation of FGF signalling has been implicated in several human diseases, caused by either gain- or loss-of-function mutations in the FGFRs or FGF ligands. Mutations in FGFRs lead to several human diseases such as Achondroplasia, Pfeiffer syndrome, Crouzon syndrome, all of which are skeletal abnormalities, which highlight FGF's importance in bone development, a mesoderm derivative (Beenken and Mohammadi, 2009). Like many genes which are involved in embryonic development, misregulation of FGFR's signal transduction pathways by mutations of the FGFR are associated with a number of cancers, this highlights the importance of FGF in regulation of cell proliferation (Turner and Grose, 2010).

1.2.2 Ligand misregulation

The autosomal dominant condition Hypophosphataemic rickets is caused by a gain of function mutation within the FGF23 gene (Consortium, 2000). Deafness can be caused by loss of function mutations in FGF3 (Tekin et al., 2007). Both Lacrimo-auriculo-dento-digital syndrome (LADD) caused by a loss of function mutation in FGF10 (Milunsky et al., 2006) and Kallmann syndrome which is caused by a loss of function in FGF8 (Falardeau et al., 2008) can also be seen in other components of the FGF signalling pathway in loss-of-function mutations in the FGFRs (Dode et al., 2003; Rohmann et al., 2006).

1.3 Capicua, a transcriptional repressor downstream of MAPK signalling pathway

We have considerable knowledge regarding the FGF dependent transcriptome in *Xenopus* development, however, it is still unclear how the mechanism by which FGF signal transduction leads to changes in gene transcription downstream of the signalling pathway. In addition, assays using cyclohexamide, an inhibitor of protein synthesis show some downstream targets of FGF are expressed when treated with cyclohexamide, which suggest that they are repressed by a labile repressor (Fisher et al., 2002b). The novel transcriptional repressor, Capicua (CIC) potentially provides such a link.

1.3.1 Regulation of CIC by receptor tyrosine kinases

The *CIC* gene was first described and identified in the *Drosophila melanogaster* embryos (Jimenez et al., 2000), acting downstream of the RTKs (fig 4) (Jimenez et al., 2012), the maternally active torso (tor) receptor and epidermal growth factor receptor (EGFR) (Ajuria et al., 2011). The tor pathway is known to be involved with the patterning of the head and tail structures in the early developing *Drosophila* larva. CIC acts downstream of the tor receptor to repress *tailless* (*tll*) and *huckebein* (*hkb*) expression which are expressed at the poles of the *Drosophila* embryo (Duffy and Perrimon, 1994). *tll* and *hkb* are genes responsible for patterning of head and tail structures. Its involvement with patterning of the anterior-posterior poles of the larva that CIC gets its name, meaning 'head-and-tail' in the Catalan dialect (Jimenez et al., 2000). The CIC protein can be found from cnidarians to vertebrates, highlighting its high conservation in the animal kingdom. The human homolog of *CIC* first was discovered in the immature granule cells of the cerebellum, hippocampus and olfactory bulb of the central nervous in 2002 (fig 4) (Lam et al., 2006; Lee et al., 2002).
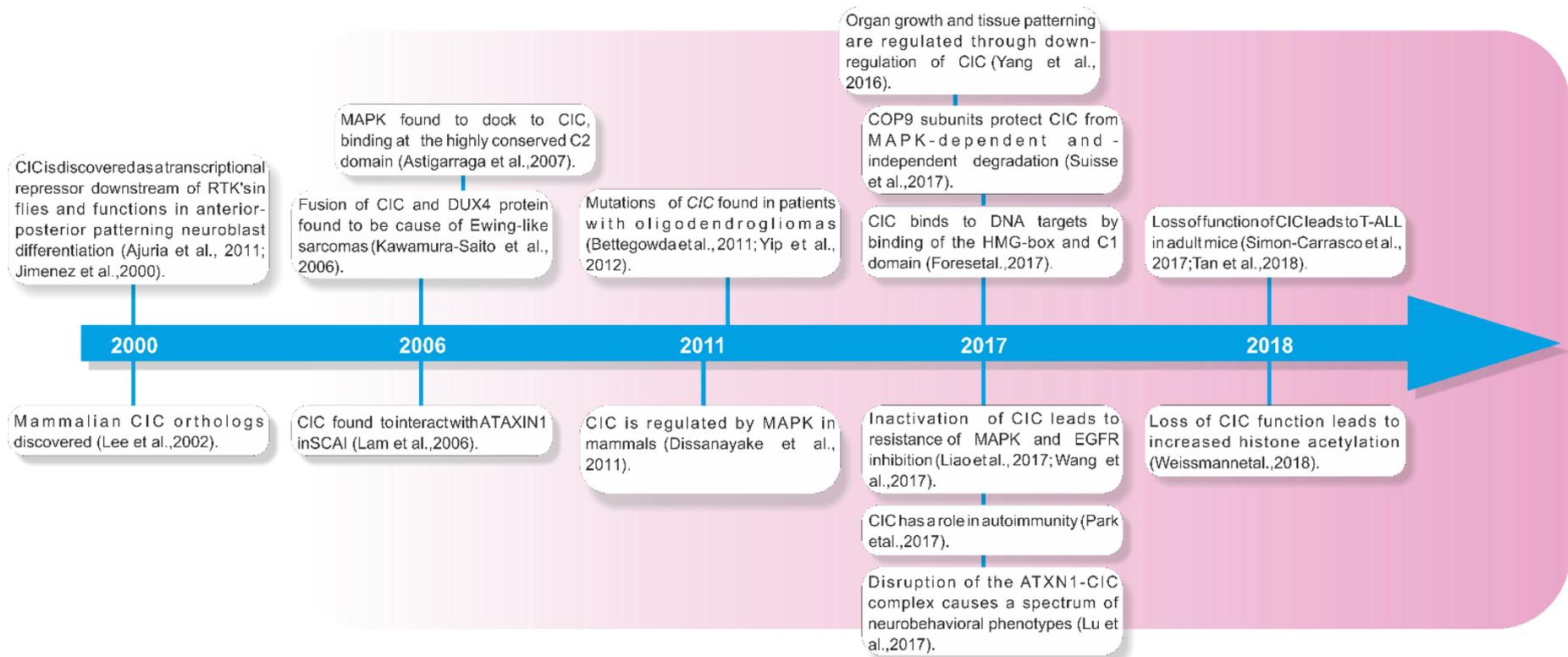
Figure 4, a time line of CIC research discoveries, from its discovery to latest findings for its role in autoimmunity and function in histone acetylation (All key findings have references in text boxes).

Like the FGF pathway, the tor pathway utilises the cell surface tor receptor tyrosine kinase (RTK) which signals through the Ras-Raf-MAPK cascade, regulating downstream gene activation. In the early developing *Drosophila* embryo, the tor receptor is ubiquitously expressed in the embryo, but its activation is restricted to the poles where its ligand, trunk, is expressed (Furriols and Casanova, 2003; Gabay et al., 1997; Greenwood and Struhl, 1997). This activation of the tor receptors creates a gradient of MAPK signalling along the anterior-posterior axis. MAPK signalling leads to decreases in concentration of the CIC protein due to its phosphorylation, leading to reduction in repressional activity of CIC, allowing the expression of *tll* and *hkb*. Phosphorylation leads to the relief of CIC repressional activity, allowing the expression of RTK downstream gene targets.

Although CIC targets both *tll* and *hkb*, it is more effective at repressing *hkb* than *tll*, resulting in *tll* having a broader spatial expression profile than *hkb* (de las Heras and Casanova, 2006; Jimenez et al., 2000). In the posterior pole of the embryo the *hkb* and *tll* spatial expression profiles created by relief of CIC repression allow the formation of the posterior hindgut and posterior spiracles (Furriols and Casanova, 2003; Greenwood and Struhl, 1997). In CIC loss of function studies, reduced repression of *hkb* and *tll* leads to ectopic expression causing irregular formation of the thoracic imaginal discs within the body of the embryo (Jimenez et al., 2000).

EGFR, is another RTK, which utilises the Ras-Raf-MAPK cascade to regulate downstream gene activation by downregulation of CIC in *Drosophila* (Astigarraga et al., 2007b; Goff et al., 2001; Jimenez et al., 2000; Roch et al., 2002). Tor is involved in patterning of the most anterior and posterior poles of the early developing fly embryo, whilst EGFR restricts CIC within the body of the embryo. Both EGFR and tor control cell growth, differentiation, proliferation and survival, switching genes on and off in different context. They both use MAPK transduction as a mechanism to regulate CIC, although the impact on CICs abundance and localisation is very different (fig 5-B). Tor signalling leads to degradation of CIC within the nucleus in the early *D. melanogaster* embryo (Grimm et al., 2012), whilst EGFR signalling causes the relocalisation of CIC from the nucleus in to the cytoplasm, in ovarian follicle cells (Astigarraga et al., 2007b; Roch et al., 2002). Both mechanisms of CIC inhibition are likely due to differences in the phosphorylation of CIC amino acid residues.

Recent *D. melanogaster* data examining wing growth suggests that the mechanism by which EGFR regulates CIC is by proteasomal degradation by Cullin 1 (Cul1)-mediated ubiquitination (Suisse et al., 2017a). EGFR activation by its ligands Spi,

leads to the activation of the MAPK cascade. MAPK phosphorylation targets the CIC protein for ubiquitylation and later proteasomal degradation (Suisse et al., 2017b). There are even suggestions that ubiquitin modification of CIC alone could lead to loss of repressional function. Cul1 acts as a scaffolding subunit of Cullin-RING E3 ubiquitin ligase complexes (Cul1-SkpA-Ago) (CRLs) once neddylated (small ubiquitin-like protein). The CRLs complex consists of Neddylation allows conformational change of the RING domain allowing ubiquitin transfer to target proteins marking them for proteasomal degradation (Merlet et al., 2009).

Conversely, the COP9 signalosome (CSN) complex deneddylates CRLs preventing auto-ubiquitylation (Lyapina et al., 2001). The CSN complex consists of nine-subunits. One such subunit, the CSN1b subunit acts in a protective manner towards CIC, preventing its degradation in the absence of a signalling (Suisse et al., 2017a). This data suggests that that the CSN complex restricts EGFR gene expression by preserving CIC from ubiquitylation and proteasomal degradation. Mutations in one of the subunits the CSN complex, the CSN1b subunit lead to loss of function of the CSN complex increasing the activity of Cul1 increasing degradation of CIC which results in a phenocopy like that of EGFR overexpression. Direct binding of CSN1b to CIC has been shown to effect stability of the CIC protein (Suisse et al., 2017a).
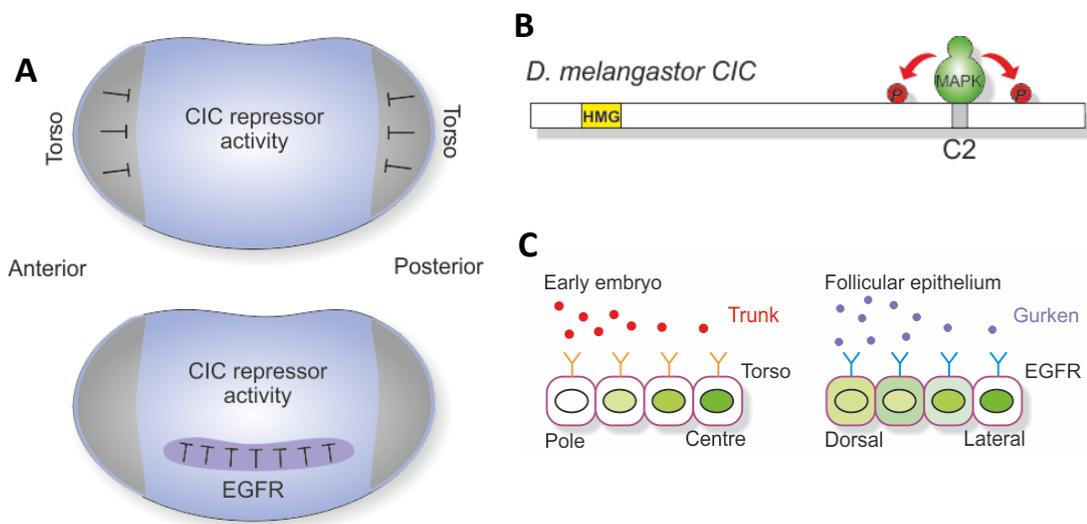
Figure 5, (A) *CIC* first discovered in *Drosophila*, is phosphorylated by MAPK via the tor RTK signalling pathway at the most anterior/posterior regions of the early fly embryo. This prevents CIC from acting as a transcriptional represser allowing the expression of huckebein and tailess, gap genes responsible for segmentation in the fly embryos (Andreu et al., 2012a; Jimenez et al., 2000). Similarly, the EGFR RTK signalling pathway relieves CIC repression within the body of the embryo allowing the induction of neurectoderm via expression of ind. (B) MAPK activation allows binding at the C2 domain of CIC leading to its phosphorylation. (C) Although both EGFR/tor activate the Ras-MAPK signalling pathway to relive CIC repression, the mechanism in which they do this is very different. Tor signalling degrades CIC within the nucleus, whereas EGFR signalling leads to the relocalisation of the CIC protein from the nucleus into the cytoplasm (Grimm et al., 2012).

EGFR is responsible for a number of roles such as direct cell fate choices, cell division, cell survival, cell migration and patterning in *Drosophila* (Shilo, 2003). Several of these pathways are known to be regulated by CIC repressional activity. EGFR is responsible for patterning of the neuroectoderm by inducing the expression of neuroblasts defective (*ind*) gene by relief of CIC repression, which leads to neuroblast differentiation (fig 5A) (Ajuria et al., 2011; Weiss et al., 1998).

In the context to the developing wing vein cells of *Drosophila*, CIC represses *argos* when no MAPK signal is present. Upon activation of the MAPK transduction signalling by EGFR activation of the *argos* gene is allowed to be expressed. The *argos* gene encodes for a ligand of EGFR which acts in an inhibitory loop restricting the level and duration of the EGFR signal during the formation of normal wing vein patterning (Ajuria et al., 2011; Brunner et al., 1994). Loss of repressional function of CIC leads to irregular wing formation (Ajuria et al., 2011).

In the maternal dorsoventral follicle cells during oogenesis, EGFR activation is required to induce expression of the *mirror* (*mirr*) gene. The *mirr* gene is located at the *iroquois* locus and encodes for an homeobox transcription factor (Gomez-Skarmeta et al., 1996). The expression of *mirr* enables the formation of the dorsal appendages in eggshell appendages. In the ventral cells CIC represses expression of *mirr* (Atkey et al., 2006; Goff et al., 2001). In addition, ventral follicle cells require the expression of the *pipe* gene, which encodes for a Heparan sulfate 2-O-sulfotransferase (HSST) (Goff et al., 2001). The HSST acts to create division of the dorsal-ventral axis of the developing embryo (Moussian and Roth, 2005). Expression of *pipe* is repressed in the dorsal follicle cells by the expression of *mirr* which is dependent on EGFR signalling. In the absence of the EGFR transduction signal, CIC represses *mirr* allowing the expression of *pipe* in the ventral cells (Andreu et al., 2012b; Fuchs et al., 2012).

Although CIC has been shown to play important roles downstream of tor and EGFR pathways in patterning of the developing *Drosophila* embryo, it also has a role in cellular proliferation. CIC has been shown to be involved with the proliferation of the larval imaginal disc cells downstream of the EGFR-Ras-MAPK transduction pathway. Larval imaginal discs are internal structures within the *Drosophila* embryo, which during the pupal stage of development, rapidly give rise to external structures such as wings, legs, antennae or other organs in the adult (Beira and Paro, 2016). Interruption of this pathway results in reduction in proliferation capability of the imaginal disc cells, reducing growth rates of these structures (Karim and Rubin, 1998;

Tseng et al., 2007). Recent studies suggest hindering CICs repressional function in this context downstream of the EGFR pathway can increase cell proliferation. In addition, evidence suggests that loss-of-function mutations of CIC circumvent any need for EGFR altogether when promoting proliferation in the larval imaginal discs cells. This highlights CIC role in cell proliferation and important relation between EGFR and CIC.

Additional research in the *Drosophila* model showed that in the intestinal midgut epithelial stem cells, the EGFR-CIC pathway plays an essential role in regulation and maintenance of epithelial stem cell proliferation, in a similar feedback system as the larval imaginal disc cells mentioned above (Jiang et al., 2011). Downregulation of CIC leads to increase in proliferation of the epithelial stem cells. Evidence has shown that CIC loss-of-function leads to ectopic growth of the intestinal stem cells. This phenotype mimics the phenotype observed in EGFR ligand overexpression and in embryos with a constitutively active Ras-MAPK transduction signalling, which is likely owing to the induction and increase of expression of proliferation genes (Jiang et al., 2011). Intestinal midgut epithelial cells with a defective EGR signal lack the ability to grow or divide and lack there regenerative/homeostatic properties. This research suggests CIC has an important role in early development as a negative regulator of growth.

1.3.2 Alternative mechanisms of CIC regulation

Recent research has shown that tor and EGFR are not the only regulators of CIC repression, Minibrain (Mnb) a kinase and Wings apart (Wap) an adaptor protein can act independently of Ras-MAPK transduction signalling to regulate CIC. In the context of wing development Mnb and Wap act in an additive manner along with Ras-MAPK signal transduction by binding to CIC, which leads to CICs phosphorylation, inhibiting its ability to act as a transcriptional repressor (Yang et al., 2016).

1.3.3 Hippo signalling regulates *CIC* at the RNA level

The Hippo pathway is known to control tissue proliferation and growth by regulation of transcription factors YAP/Yorkie. Yorkie is regulator of genes responsible for cell proliferation and suppression of apoptosis. Genes in this pathway include; *cyclin E, DIAP, vein, wg, dMyc* and *E2F* (Cho et al., 2006; Goulev et al., 2008; Huang et al., 2005; Neto-Silva et al., 2010; Nolo et al., 2006; Tapon et al., 2002; Thompson et al., 2006; Wu et al., 2003; Zhang et al., 2009a). The Hippo/Yorkie pathway also regulates

bantam miRNA expression (Nolo et al., 2006). Yorkie can activate EGFR MAPK transduction or act independently of EGFR to induce *bantam* expression. The activation of the EGFR-MAPK and increase expression of *bantam* has a greater effect on CIC concentration in the cell than MAPK transduction signalling alone. Crosstalk between Hippo and EGFR pathways is not only at the MAPK level, amphiregulin a ligand of EGFR is induced by Hippo pathway signalling. In *Drosophila, bantam* microRNA acts to limit CIC expression at the mRNA level by a negative feedback loop mechanism (Herranz et al., 2012b). *Bantam* miRNA itself is a target of the EFGR and Hippo pathway signalling which links these two pathways in a mode of regulation of CIC. CIC represses *bantam* miRNA expression but when CIC repression is alleviated by EGFR MAPK signalling, *bantam* is upregulated further inhibiting CIC expression, leading to a MAPK signal amplification. Bantam may reduce the threshold by which MAPK needs to act to relieve repression of CIC (fig 6) (Degoutin et al., 2013; Herranz et al., 2012a).
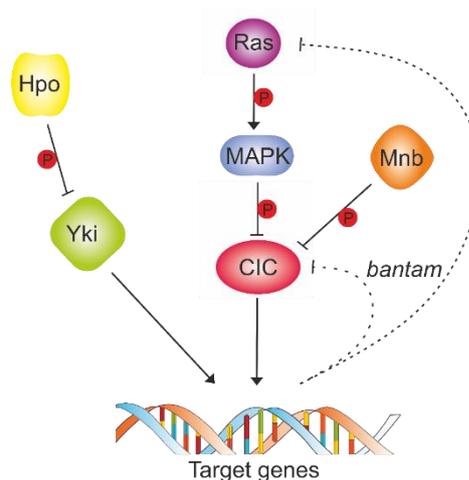
Figure 6, schematic of the CIC regulatory pathways in *D. melanogaster* proliferation and growth control.

1.4 Functional domains of CIC

The *M. musculus* and *H. sapiens* homologs of the CIC protein contain 7 functional domains (fig 7). The shorter isoform CIC-S, does not contain the N1 domain which can only be found within N-terminus of the longer isoform CIC-L, which is encoded within exon 1. All other domains are found to be common between both isoforms (Jimenez et al., 2012).
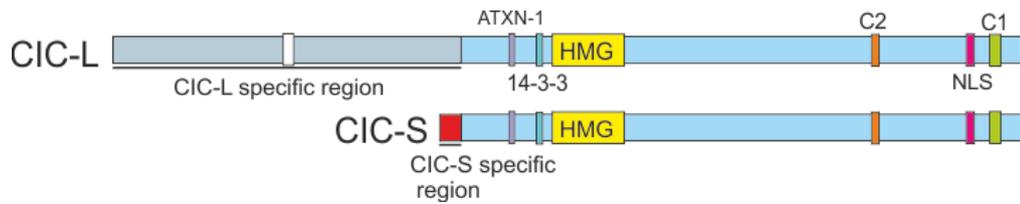
Figure 7, the CIC protein and the 7 functional domains. The shorter isoform CIC-S, does not contain the N1 domain which can only be found within the 5' end of the longer isoform CIC-L, which is encoded within exon 1. All other domains are found to be common between both isoforms (Jimenez et al., 2012).

1.4.1 HMG box and C1 motif

The CIC protein contains seven highly conserved domains (fig 7) and is regulated by RTK MAPK transduction signalling (fig 6) (Ajuria et al., 2011; Astigarraga et al., 2007a; Dissanayake et al., 2011; Fryer et al., 2011; Futran et al., 2015). CIC binds to DNA by means of a High Mobility Group-box (HMG-box) domain, which recognises octameric T(G/C)AATG(A/G)A sites in target gene enhancers and promotors (Jimenez et al., 2012). Other HMG-box transcription factors include Sox/SRY (Sox) and (TCF/LEF1 (TCF) family members. Studies have shown that although these proteins utilise the HMG-box for binding to their DNA targets, the mechanisms by which they do this is very distinct (Fores et al., 2017). Transcription factors that utilise the HMG-box for transcriptional repression often require additional factors or other domains due to its relatively weak affinity and specificity for DNA. The Sox family of proteins utilise two mechanisms to bind to DNA targets, either the recruitment of partner protein which binds to a recognition site adjacent to a Sox binding site or the formation of a Sox dimer to bind to DNA (Peirano and Wegner, 2000). Alternatively, TCF family members utilise a C-clamp domain found within the TCF transcription factor which binds to GC-rich DNA sequences known as Helper sites (<10 bp from) near the TCF binding site (Atcha et al., 2007). In contrast, the CIC transcription factor requires a conserved C1 motif which is only found within the CIC protein (Astigarraga et al., 2007b; Fores et al., 2017). Unlike the Sox and TCF transcription factors families, CIC's C1 motif is found a long distance from the HMG-box, has no known DNA-binding activity and is not involved in dimerization (Fores et al., 2017).

1.4.2 ATAXIN-1 domain

In the vertebrate homologs of the CIC protein, isoforms contain an ATXN1 binding domain which enables binding of ATAXIN-1 (ATXN1) or ATAXIN-1-LIKE proteins to form a protein repressor complex with the CIC protein (Lam et al., 2006; Zoghbi and

Orr, 2009). Discovered when looking at the mechanistic cause of Spinocerebellar ataxia type 1 (SCA1) (Fryer et al., 2011; Lam et al., 2006). Unlike the vertebrate homologs, no evidence has been found to suggest that ATXN1 interacts with the CIC protein in flies. ATXN1 also associate with co-repressors NCOR2/SMRT and histone deacetylases 3 and 4 (Mizutani et al., 2005; Tong et al., 2011; Tsai et al., 2004). Studies have also shown that ATXN1/ATXN1L proteins stabilise CIC, likely through protein-protein interaction (Lam et al., 2006). When expression of ATAXN1/ATXN1L was knocked down CIC protein concentrations were also reduced but *CIC* mRNA was unchanged, although the exact molecular mechanism by which ATAXN1/ATAXN1L functions to stabilise CIC is still unknown. It must be stated that knockout of ATAXN1L has a greater impact on the stabilisation of CIC in comparison to knockout of ATAXN1 (Lee et al., 2011).

Atxn1L$^{-/-}$ and *Atxn1*$^{-/-}$ mice models were created to look at the interactions between ATAXN1/ATXN1L and CIC. The *Atxn1*$^{-/-}$ or *Atxn1L*$^{-/-}$ mice phenotypes displayed lung alveolarization defects and overexpression of the *Matrix metalloproteinase (Mmp)* genes which are responsible for extracellular matrix (ECM). Mmp proteins are responsible for the breakdown of the ECM in several universal processes such as reproduction, embryonic development, bone development, cell migration and wound healing as well as learning and memory. In mice ATXN1/ATXN1L and CIC regulate extracellular matrix remodelling and lung alveolarization (Lee et al., 2011). Knockout of either ATXN1L or CIC leads to increase of expression of *Etv4* a repressional target of CIC and activator of *Mmp* genes. *Etv4* activates transcription of *matrix metallopeptidase 9 (MMP9)*. The *MMP9* is a regulator of neutrophil migration across ECM and is essential for ECM remodelling (Delclaux et al., 2012). PEA3 family of genes consists of three members Etv1 (also known as Er81) (Brown and McKnight, 1992), Etv4 (also known as Pea3 and E1Af) and Etv5 (also known as Erm) (Chotteau-Lelièvre et al., 1997; Monte et al., 1994; Nakae et al., 1995) which all contain a ETS-domain and at least one CIC binding motif in their promoter regions.

1.4.3 14-3-3 motif

In mammals the CIC protein homolog contains a 14-3-3 domain which is a common recognition motif for 14-3-3 phosphoserine family of binding proteins (Dissanayake et al., 2011). The motif contains a serine or threonine residue which once phosphorylated recruits the 14-3-3 proteins. In human HEK-293 cells studies, p90 ribosomal S6 kinase (p90^{RSK}) targets CIC by phosphorylating the 14-3-3 motif serine residues flanking either side of the HMG-box domain (Dissanayake et al., 2011).

p90[RSK] expression is activated by MAPK signal transduction.14-3-3 proteins are a highly conserved family of proteins which are expressed in all eukaryotic cells and involved with a diverse range of signalling processes such as; cell cycle control, apoptosis, and mitogenic signal transduction (Fu et al., 2000). When the serine is phosphorylated at the 14-3-3 motif by p90[RSK] it creates a docking site for 14-3-3 regulatory proteins binding, blocking or reducing optimal binding efficiency of the neighbouring HMG-box domain leading to relief of repression from CIC. Repression of PEA3 Ets transcription factors Etv1, Etv4 and Etv 5 is relieved once 14-3-3 protein dimers are bound to the 14-3-3 motif (Dissanayake et al., 2011). In contrast, in human melanoma cells which express a form of CIC which prevents binding of the 14-3-3 dimer inhibit expression of *Etv1*, *Etv4* and *Etv5* (Dissanayake et al., 2011).

1.4.4 N1 and N2 domains

The N1 domain contained within the large CIC-L specific region of the CIC protein is highly conserved but has no known function, although CIC-L has a role in oogenesis, which might suggest this domain is involved in some way in this process (Fores et al., 2015). A recent study in flies led to the discovery of the N-terminal N2 domain unique to CIC-S isoform in flies (Fores et al., 2015). The N2 domain functions as a docking site for the corepressor Groucho. This mechanism is specific to dipteran insects and suggests an alternative function between the two prominent isoforms of CIC. N2 motif arose approximately 250 million years ago and likely came about due to genetic drift. It must be noted that CIC can also function as a repressor independently of Groucho in *Drosophila.* Although Groucho forms a corepressor complex with CIC in fruit fly models no evidence has been found to suggest Groucho forms a repressional complex with CIC in vertebrates.

1.4.5 C2 domain

The C2 domain is highly conserved in invertebrates and vertebrates. Research in the fly model demonstrated that the C2 domain was critical for CIC repression. The C2 motif was found to function as a novel docking site for di-phosphorylated MAPK in fly and human model systems. Binding of phosphorylated MAPK to CIC mediates further phosphorylation of CIC in response to signalling. In CIC[ΔC2] mutant fly lines which lack the C2 motif showed a similar phenotype to the knockout of *tor* suggesting the C2 domain is essential for relief of CIC repression and CIC protein lacking the C2 motif produces dominant, constitutively active repressor that is insensitive to tor RTK-mediated inactivation. Regarding EGFR, CIC[ΔC2] fly strains show a less server

phenotype in comparison to EGFR knockout phenotypes. Suggesting at least in the case of *mirr* that its regulation by EGFR is not exclusively regulated by CIC repression activity.

1.4.6 NLS

The human homolog of CIC contains a common nuclear localization sequence (NLS) located at the C-terminal. The NLS site allows Importin α4 (KPNA3) to bind to CIC leading to its import into the nucleus. MAPK signal transduction leads to the phosphorylation of NLS site (Ser^{1382} and Ser^{1409}) and or the phosphorylation of KPNA3 preventing binding of KPNA3 to CIC.

1.5 CIC in early vertebrate development

Prior to this project no research had been performed using the Xenopus animal model to investigate the CIC. Little research had been completed on vertebrate animal models with most of the research and are understanding of CIC coming from the D. melanogaster model. Although advances in are understanding of CIC have come from the H. sapiens cancer field (Bettegowda et al., 2011a; Davoli et al., 2013; Yip et al., 2012) and analysis from the M. musculus model (Lam et al., 2006) the amount of research on CIC remains to be limited. In early development M. musculus CIC is expressed in the developing lung and is important for development of alveoli in the lung (Lee et al., 2011). Strong evidence has been found that CIC functions in the M. musculus brain retaining a neurogenesis role, regulating neural patterning and development of the brain, seen previously in invertebrates (Ahmad et al., 2018; Lu et al., 2017b). Knockout of CIC leads to embryonic fatality in M. musculus. Further indication of CIC's important role in neurogenesis has been found in germline heterozygous CIC truncating mutations in humans. These mutations have been found to cause intellectual disability, attention deficit hyperactivity disorder and autism spectrum disorder (Lu et al., 2017b).

No studies other than in D. melanogaster model have found differences in the expression or function of the prominent isoforms of CIC (CIC-L & CIC-S). In the D. melanogaster model, CIC-L has a suggested specific role in oogenesis (Goff et al., 2001; Rittenhouse and Berg, 1995). Only one study has so far been published using the Danio rerio model to study CIC (Chen et al., 2014). The study found that CIC is highly expressed maternally (Chen et al., 2014). Although it is important to consider that unlike other vertebrate models, D. rerio has two copies of the CIC gene (cica and

cicb), likely due to a genome duplication event (Taylor et al., 2003). Later in shield stage of D. rerio development cica and cicb become weakly expressed, maintaining a low level of expression throughout early development and into adulthood (Chen et al., 2014). Given the limited data available outside of invertebrate model CIC research the findings from this study could prove to be invaluable.

1.6 CIC misregulation in disease

1.6.1 Ewing-like sarcoma

Misregulation of CIC leads to several diseases and the first to be discovered was Ewing-like sarcoma in 2006, which is a member of the Ewing's family of tumours (EFTs) (Kawamura-Saito et al., 2006). Ewing-like sarcoma is caused by the chromosomal translocation of the *CIC* gene at t(4;19)(q35;q13) to the *double homeodomain 4* gene (*DUX4),* producing a chimeric fusion protein CIC-DUX4. The CIC-DUX4 contains the N-terminal of the CIC protein and the C-terminal of DUX4, which acts as a dominant oncogene resulting in strong transcriptional activity, rather than a transcription repression of several of its target genes.

When analysing changes in gene expression in the human tumour tissue samples, the polyoma enhancer activator 3 (PEA3) family of genes (Etv1, Etv4 and Etv5) showed upregulated expression. These genes were found to contain CIC octameric binding sites within their promoter regions. This was the first evidence that CIC regulates the expression of PEA3 genes. In addition, upregulation of the PEA3 genes have been observed in some human invasive metastatic malignant neoplastic (breast, ovarian and gastrointestinal) cancers (Davidson et al., 2003; de Launoit et al., 2000; Horiuchi et al., 2003; Roberts et al., 1992; Yamamoto et al., 2004). CIC has been shown to be involved in other cancers which only further highlight its role as an oncogene.

1.6.2 Spinocerebellar ataxia 1

In 2006, a study found that CIC was involved in Spinocerebellar ataxia 1 (SCA1). SCA1 is neurodegenerative disease caused by the polyglutamine (polyQ) tract expansion of ATAXN1. SCA1 is a condition in humans that leads to progressive loss of movement control, coordination, balance. The condition can also lead to difficulty in cognitive brain function due to problems with processing, learning and memory. The SCA1 condition is not caused by a mutant *CIC*, but rather through the binding of

the polyQ tract expanded ATAXN1 to the wild type CIC. Interestingly, SCA type 14 (Brusse et al., 2006) and SCA type 27 (Coebergh et al., 2014) were both found to be caused by defects of the PKCγ signal transduction pathway regulated by FGF14 growth factor.

Further studies in heterozygous *CIC*^{+/−} mice found disruption of the ATXN1-CIC complex in developing forebrain resulted in multiple behavioural abnormalities (Lu et al., 2017a). It must be stated that CIC ^{−/−} mice were not viable for life. Behavioural tests showed that mice exhibited hyperactivity. Continued knockout experiments of either ATXN1, ATXN1L and CIC displayed impaired learning and memory. The reason for the abnormal brain function was likely due to postnatal reduction in the cortical thickness and the number of CUX1⁺ cells (cortex cell marker) (Nieto et al., 2004) due to disruption of the ATXN1-CIC complex. Brain sections from 5-week-old mice were analysed and showed a reduced number of CUX1⁺ cells were likely due to defects in neuronal maturation or maintenance. CUX1 has an essential role in the developing brain in promoting dendritic branching and due to CUX1 levels being reduced in mutant mice neuronal dendritic branching was reduced. Mice had *CIC* knocked out in the hypothalamus and medial amygdala brain tissue using a Cre line which utilised the promoter from orthopedia homeobox (Otp). The results of the *CIC* knockout hypothalamus and medial amygdala showed that mice became more aggressive which suggests CIC has an important role to modulate social interaction in the brain (Lu et al., 2017a).

1.6.3 Oligodendroglioma

Studies have shown that inactivating mutilations in the C1 regions are a hotspot for tumour growth and, oligodendroglioma (ODG), a low-grade brain tumour which is caused by my mutations in the HMG-box and C1 coding region. In adults, ODG is the second most common malignant brain tumour, which account for 20% of all brain tumours. Studies in 2011, found that in 50% ODG cases were caused by a mutation within the *CIC* gene coding for the highly conserved HMG-box region (Bettegowda et al., 2011b). This demonstrates the importance of the C1 domains function in CICs transcriptional repression and regulation of proliferation (Bettegowda et al., 2011a; Yip et al., 2012). Phosphorylation of CIC by MAPK inhibits its ability to act as a transcriptional repressor. When CIC is not repressed by MAPK signalling, CIC is able to bind to octameric DNA target sequences via the HMG-box domain repressing gene expression (Ajuria et al., 2011).

Research examining simultaneous mutations in *CIC* and *isocitrate dehydrogenase* (*IDH1-R132H* or *IDH2-R172K*) genes in cases of ODG, observed that the wildtype CIC-S protein was localised in the cytoplasm after the cells were fractionated and purified (Chittaranjan et al., 2014; Yip et al., 2012). CIC-S was discovered to be found predominantly near the mitochondria in the cytoplasm, whilst the CIC-L protein was predominantly localised within the nucleus (Yip et al., 2012). CIC was found to be associated with ATP-Citrate Lyase (ACLY) in LC-MS/MS experiments. ACLY is protein which is responsible for synthesis cytosolic acetyl-CoA in the Krebs cycle (Fatland et al., 2000). Acetyl-CoA is a molecule involved in many biological reactions in protein, carbohydrate and lipid metabolism. Stable human embryonic kidney (HEK293) and Human Oligodendroglioma (HOG) cell lines were created which ectopically co-expressed wildtype *CIC* and *IDH1* and exhibited increased clonogenicity. Cell lines were also created co-expressing mutant *CIC* and *IDH1,* which unlike the non-mutants, displayed reduced clonogenicity (Chittaranjan et al., 2014). Concentrations of cellular 2-Hydroxyglutarate (2HG) were increased in mutant CIC-S cell lines in comparison to wildtype CIC clones. Levels of ACLY were also found to reduced, much like those seen in ODG patient samples. Interestingly, both FGF2 (bFGF) and FGF21 associated with mitochondria function (Mäkelä et al., 2014; Srisakuldee et al., 2018).

1.7 Pathways linked between Capicua and FGF signal transduction

There is already some evidence that CIC is involved with some of the same pathways regulated by FGF. In the mammalian models, CIC requires the formation of a protein complex with either ATXN1 or its paralog ATXN1-L, to function as a transcriptional repressor (Lee et al., 2011), whereas invertebrates, require the formation of a protein complex with Groucho (Fores et al., 2015). The Ataxin complex has been found to be important for the regulation of the extracellular matrix in lung alveolarization (Lee et al., 2011).

Both paralogs have a role in stabilising CIC and knockout of both ATXN1/ATXN1-L leads to a decrease of CIC protein within the lung cell. Dual knockouts also display increased expression of *ETS translocation variant 1* (*Etv1*)*, Etv4* and *Etv5. Etv1*, *Etv4* and *Etv5* belong to one of the subfamilies of the E-twenty-six (ETS) family of transcription factors, Pea3. All the members contain a highly conserved ETS domain, a winged helix-turn-helix structure, which binds to DNA. There is evidence in *Xenopus*

and in *Danio rerio* that FGF also regulates the Pea3 sub-family genes. The knockdown of all 3 of the Pea3 ETS genes lead to a phenotype similar to FGF inhibition (Znosko et al., 2010). Both Etv4 and Etv5 are regulated by FGF transduction in mammalian lung development. Knockouts of Etv4/Etv5 lead to the malformation of the lung branching. In other contexts, such as limb development, FGF4 activation leads to the increased expression of *Etv4*/*Etv5,* which has been shown to inhibit *sonic hedgehog* (*Shh*) expression in *M. musculus* limb buds (Zhang et al., 2009b).

In human melanoma cell lines, Etv1, Etv4 and Etv5 mRNA levels have been shown to be elevated when CIC is inhibited by small interfering RNA (siRNA) knockdown, which also lead to increased cell migration. The study suggested that loss of CIC desensitises the cell to the effects of MAPK regulation (Dissanayake et al., 2011). In humans, Oligodendrogliomas are a type of tumour caused by loss of the chromosomal 1p and 19q arms, which are usually found in the cerebral hemisphere, frontal and temporal lobes of the brain, effecting the supportive, glial cells. *CIC* is located at 19q arm and in 50% of cases of Oligodendrogliomas CIC has mutations within its gene (Padul et al., 2015). In the Oligodendrogliomas mutated CIC cases, the *Etv* transcription factors were significantly upregulated. *Sprouty* was also found to be upregulated, which is a known target of FGF regulation and a MAPK inhibitor (Branney et al., 2009). Interestingly, *FGFR1* was shown to have increased expression in the mutated *CIC* cases (Padul et al., 2015).

1.8 Aims of this project

Even though extensive work has been completed on the FGF dependent transcriptome in early amphibian development (Branney et al., 2009) there are still gaps in what we understand about the mechanism of gene regulation downstream of FGF signalling. This project aims to address some of the gaps in knowledge by investigating a potential novel regulator of FGF dependent transcription, CIC. There is a large body of work on FGF signalling in the *Xenopus* model (Amaya et al., 1993; Christen and Slack, 1997b; Isaacs et al., 1994; Isaacs et al., 1992; Pownall et al., 1996b), making it an attractive model for investigating the downstream effects of FGF signalling in early vertebrate development. Interestingly, in *Drosophila*, MAPK signalling is activated by several different RTKs in early development, whilst in *Xenopus*, FGF transduction is solely responsible for di-phosphorylation and activation of MAPK in the early embryo (Branney et al., 2009; Christen and Slack, 1999; Shinya et al., 2001). In the fly model, CIC loss phenocopies activation of the RAS - MAPK

pathway and if CIC is present in early development of the Xenopus embryo, it is likely regulated by FGF signal transduction.

The overall aim of this project is to establish if CIC is involved as a transcriptional repressor downstream of FGF in early amphibian development, acting in a similar fashion to EGFR and tor regulated by the Ras-MAPK transduction pathway (fig 6) (Grimm et al., 2012; Roch et al., 2002). If CIC is regulated in the early stages of *Xenopus* development by MAPK transduction, it would be more than likely regulated by FGF.
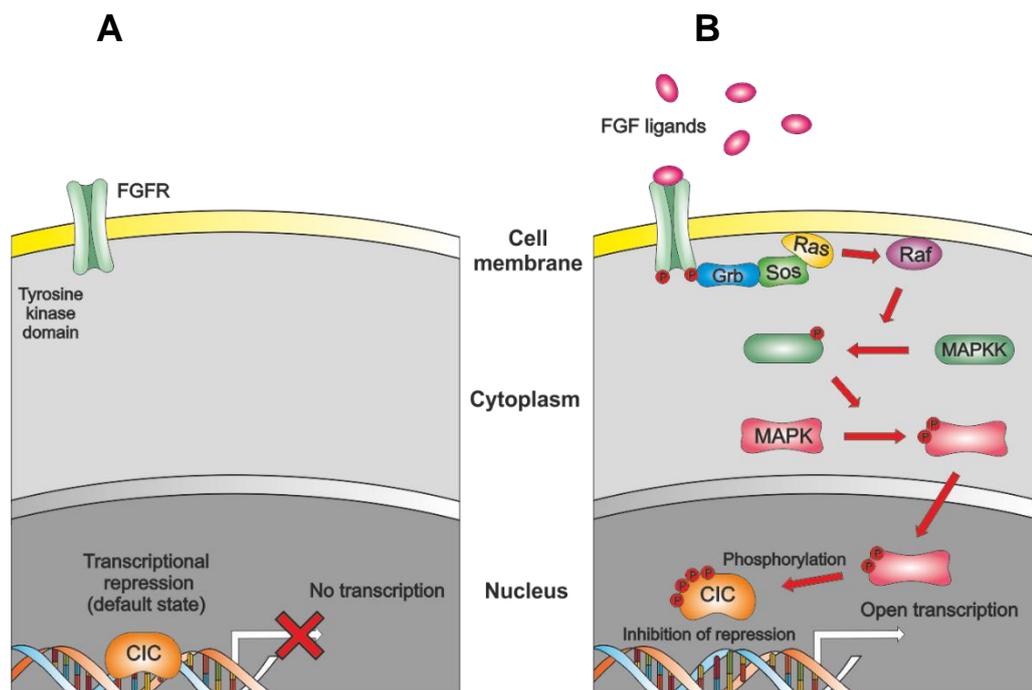
Figure 8, (A) the hypothesis that CIC function downstream of the FGFRs leads to a default state of repression when there is no Ras-MAPK signal transduction, preventing the expression of FGF gene targets. (B) Alternatively, when the Ras-MAPK signalling pathway is activated by the binding of FGF ligands to the FGFRs it leads to a signalling cascade that causes the phosphorylation of CIC at various sites of the protein, leading to relief of repression and allowing the expression of FGF targets.

The hypothesis: Transcription of a subset of FGF target genes is dependent on MAPK mediated inhibition of CIC transcriptional repression.

The overall aim of this project is to determine whether FGF activated MAPK signalling regulates CIC function and to:

- Characterise the genomic locus and sequence of amphibian *CIC*.
- Characterise the temporal and spatial expression of *CIC* isoforms during amphibian development.

- Determine how FGF dependent MAPK phosphorylation of CIC effects its abundance.
- Determine the effects on amphibian development resulting from inhibiting and overexpressing *CIC*.
- Determine if CIC functions downstream of FGF signalling by finding overlapping transcriptional changes of *CIC* knockdown and *FGF* overexpression using RNA-seq.

Chapter 2: Materials and methods

2. 1 Embryological methods

2.1.1 *Xenopus laevis in vitro* fertilization and embryo culture

Female *Xenopus laevis* were induced to lay by subcutaneous injection of 250-350 units of human chorionic gonadotrophin (hCG: Chorulon) 16 hours prior to egg collection. Eggs were fertilised with a fresh suspension of macerated testes in distilled water, obtained from a culled male. Embryos were cultured in NAM/3 (1/3rd Normal Amphibian Medium) (Sive et al., 2000) at 14-24°C, on 1.5% agarose-coated 55 mm dishes. Embryos were de-jellied 30 minutes post fertilisation prior to cleavage in a solution of 2.5% L-cysteine hydrochloride monohydrate (Sigma) in distilled water, pH7.8-8. Embryos were changed to NAM/10 (1/10th NAM) before the onset of gastrulation. Embryos were staged according to Nieuwkoop and Faber (1967).

Normal Amphibian Medium (NAM)

Component	Concentration
NaCl	110 mM
KCl	2 mM
Ca(NO3)2	1 mM
MgSO4	1 mM
EDTA	0.1 mM
NaHCO3	1 mM
Sodium phosphate pH 7.4	2 mM

Make to final volume 1 litre using distilled water

2.1.2 *Xenopus tropicalis in vitro* fertilization and embryo culture

Female *Xenopus tropicalis* were primed to lay by subcutaneous injection with a low dose of 10 units of hCG 24 hours prior to egg collection. Females were induced 2.5-4 hours before egg collection by injection of 100 units of hCG. Eggs were fertilised with a fresh suspension of macerated testes in L-15 medium (Sigma) + 10% foetal calf serum, obtained from a culled male. Embryos were cultured in MRS/9 (1/9th Modified Ringer's Solution) at 21.5-27°C, on 1.5% agarose-coated 60mm dishes. Embryos were de-jellied prior to cleavage in a solution of 2.5% L-cysteine hydrochloride monohydrate in distilled water, pH7.8-8. Embryos were changed to

MRS/20 (1/20th MRS) before the onset of gastrulation. Embryos were staged according to Nieuwkoop and Faber (1967).

Modified Ringers Solution

	1X	Mol. Wt.	10X	for 1L 10X*
NaCl	0.1M	58.44	1 M	58.44 g
KCL	1.8 mM	74.56	18 mM	1.34 g
MgCL2	1.0 mM	203.31	10 mM	2.03 g
Hepes	5.0 mM	238.31	50 mM	11.92 g

Make to final volume 1 litre using distilled water

2.1.3 Micro injections

X. laevis embryos were injected in NAM/3 + 5% Ficoll solution (type 400 DL; Sigma) and transferred to NAM/10 prior to gastrulation. 1-4 cell staged embryos were injected with a total volume of 18.6nl of mRNA per embryo, either using the gas PM 1000 Cell Microinjector or Drummond Microinjector with pulled needles (Narishige). *X. tropicalis* embryos were injected in MRS/9 + 3% Ficoll solution and transferred to MRS/20 prior to gastrulation. 1-4 cell staged embryos were injected with a total volume of 4.6nl of mRNA using the gas PM 1000 Cell Microinjector with pulled needles (Narishige).

2.1.4 Photography

Embryo images were created using a SPOT 14.2 Colour Mosaic camera (Diagnostic Instruments Inc.) and SPOT Advanced software, with a Leica MZ FLIII microscope. Images were processed using Adobe Photoshop CS6.

2.2 Molecular biological methods

2.2.1 First-strand cDNA synthesis

Total RNA was isolated by Trizol extraction mentioned previously. cDNA was synthesised with SuperScript II or IV Reverse Transcriptase from mRNA using either gene specific primers (2 µM), random hexamers (50 µM) (Invitrogen, Cat no: 10609275) or Oligo d(T) primers (50 µM) (Invitrogen, Cat no: 10249034). The following components were combined to anneal the primers to the template RNA:

Volume	Component
1 µl	Oligo d(T) primers, random hexamers* or gene-specific reverse primer
10 µl	dNTP mix (10mM each)
Up to 11µl	Template RNA (10 pg-5 µg total RNA or 10 pg -500 ng mRNA)
to 13 µl	Nuclease-free water

(Left margin: Annealing primer to template RNA)

The mixture was briefly heated up to 65°C in a thermocycler (need name of cycler) and chilled on ice for 2 minutes before adding the following components:

Volume	Component
4 µl	SSIV buffer
1 µl	100mM DTT
1 µl	RNaseOUT Recombinant RNase Inhibitor
1 µl	Superscript IV Reverse transcriptase (200U/ µl)

(Left margin: RT reaction mixture)

If the reaction mixture was setup using random hexamers*, the reaction had additional 10 minute 23°C incubation. This was not required for gene specific primers or Oligo d(T) primers prior to incubation at 55°C for 10 minutes. The reverse transcription reaction was inactivated by incubation at 80°C. Once inactivated cDNA from the reverse transcriptase (RT) reaction, any products which were larger than 1 kb were incubated at 37°C with the addition of 1 µl *E. coli* RNase H (New England Biolabs, Cat no: M0297L) for 20 minutes to remove any RNA. All cDNA was used immediately for PCR amplification or stored at -80°C.

2.2.2 *In vitro* transcription of mRNA from template

Antisense transcription of mRNA from linear plasmid and PCR template was created using SP6 or T7 MEGAshortscript™ (Invitrogen, Cat no: 10628495), MEGAScript™ (Invitrogen, Cat no: 10065754, 10584245), mMessage mMachine™ kits (Ambion, Cat no: 10391175). MEGAshortscript is more efficient at generating transcripts from small templates. mMessage mMachine is more efficient at generating transcripts from larger templates. Kits were used following manufacturer's protocols except for the

following adaptions to increase efficiency for larger transcripts using mMessage mMachine kit. When creating transcripts larger than 5 Kb, GTP becomes a limiting factor which leads to lower yields or premature termination of transcription. To overcome this problem the final cap analog ratio was decreased from 4:1 to 2:1 (cap analog:GTP) with the addition of 1 µl cap analog to the 20 µl reaction volume.

The transcription reaction was incubated at 37°C for 4 hours using 0.1–1 µg of PCR template. After 4 hours incubation the DNA template was removed with the addition of 1 µl DNAse I (Qiagen, Cat no: 79254) to the reaction at 37°C for 15 minutes. mRNA was checked for appropriate transcription by gel electrophoresis. mRNA was purified by adding 115 µl of nuclease-free water and 15 µl of ammonium acetate stop solution to the 20 µl mRNA reaction, taking the volume up to 150 µl. An equal volume of phenol/chloroform was added, samples were vortexed and centrifuged at 13,000 rpm. The aqueous phase was transferred to another tube and an equal volume of chloroform was added. Sample were vortexed and centrifuged at 13,000 rpm and the aqueous phase was transferred to a new tube. 150 µl of isopropanol was added to the samples and tubes were chilled at -20°C for 15 minutes. Samples were centrifuged at 13,000 rpm to pellet the mRNA and dried by desiccation. The mRNA was resuspended in 20 µl nuclease-free water. mRNA was further purified by lithium chloride precipitation by increasing the volume to 50 µl by adding nuclease-free water. 30 µl of Lithium chloride was added and samples were chilled at –20°C for 30 minutes. Tubes were centrifuged at 13,000 rpm for 15 minutes to pellet and dried by desiccation, before being resuspended in 20 µl nuclease-free water. All mRNA was stored at -80°C.

2.2.3 Agarose gel electrophoresis

DNA and RNA samples were run on either ethidium bromide (Polysciences, Cat no: 23590-100) or SYBR safe (Invitrogen, Cat no: 10646353) (1/1000th) stained agarose gels at 0.8-2% using TAE (40 mM Tris-acetate; 1 mM EDTA, pH 8).

2.2.4 Extraction of total RNA

For each extraction, either 5 x *X. laevis* embryos or 10 x *X. tropicalis* embryos were collected. All total RNA extraction used the Trizol reagent (Invitrogen, Cat no: 12034977), following the manufacturers protocol. Embryos were stored at -80°C and thawed on ice before being homogenised by pipetting in 1ml of Trizol. Samples were left at room temperature for 5 minutes and centrifuged at 13,000 rpm for 10 minutes

at 4°C. The supernatant was transferred to a new tube and left at room temperature for 5 minutes. 200 µl of chloroform was added to the sample, vortexed and left to stand at room temperature before being centrifuged at 13,000 rpm for 15 minutes at 4 °C. 400 µl of top aqueous phase was transferred to a new tube. An equal volume of chloroform was added to the sample, before being centrifuged for 5 minutes at 13,000 rpm. The top phase was again transferred to a new tube. 500 µl of propan-2-ol (Sigma-Aldrich, Cat no: SSO5408) was added and samples were placed at -20°C for 30 minutes. Samples were centrifuged at 13,000 rpm for 15 minutes at 4°C to pellet the RNA. The supernatant was removed, and pellet washed with 1ml of ice cold 70% ethanol. The samples were centrifuged for 5 minutes at 13,000 rpm at 4°C. This step was repeated before drying the pellet by desiccation. RNA was resuspeneded in 100 µl of nuclease-free water and 120 µl of 7.5 M LiCl/50 mM EDTA and stored at -80°C overnight to precipitate. Samples were centrifuged at 13,000 rpm the next day, washed with 70% ethanol using the above steps and dried by desiccation. Samples were resuspened in 20 µl of nuclease-free water. The NanoDrop 2000 spectrophotometry and agarose gel electrophoresis were used to analyse quality and quantity of total RNA.

2.2.5 Quantification of DNA and RNA

The NanoDrop 2000/8000 Spectrophotometer (Thermo Scientific) were used to quantify the DNA and RNA concentrations measuring absorbance at the 260nm wave length.

2.2.6 DNA minipreps and midipreps

Single colonies were cultured in Lysogeny broth (LB)-amp medium overnight in either 2-5ml or 50ml depending on the requirements of quantity of plasmid required for Miniprep (Qiagen, Cat no: S27104) or Midiprep (Qiagen, Cat no: 12145) respectively. Plasmids were isolated using the QIAprep Spin Miniprep/Midiprep Kit systems (Qiagen), following the manufacturer's protocols.

Lysogeny Broth (LB) Medium

Component	Weight (g)
Tryptone	10
yeast extract	5
NaCl	10
Molecular Agarose	15

Make to final volume 1 litre using distilled water

2.2.7 Sequencing

DNA sequencing was carried out using the University of York Technology Facility Genomics Lab, using the 3130 Genetic Analyser (Applied Biosystems), or by the GATC Biotech Sanger postal sequencing service, Germany. Sequencing analysis was performed using the SeqMan software from the Lasergene Genomics Suite (DNA Star).

2.2.8 Bacterial transformation

JM109 (Promega, Cat no: L2005) and DH5α (Invitrogen, Cat no: 12017519) competent cells were used for transformations. Cells were thawed on ice from -80°C for 30 minutes. 1–50ng of DNA was added to 100 µl of the competent cells. Bacteria were chilled for a further 10 minutes. Bacteria were heat shocked for 45 seconds in a water bath at 42°C and immediately placed on ice to chill for 2 minutes. 900 µl of LB media was added to each transformation reaction and incubated for 1 hour at 37°C in a shaking incubator. Bacteria were then plated out onto (LB) Ampicillin (Amp) agar plates (15g/1 agar), plating cells out in 100 µl LB medium (1/10, 1/100 dilutions).

2.2.9 Colony PCR

To screen successful transformation of plasmids, colony PCR was used to detect which colonies contained plasmids with inserts of the correct size. The reactions used the following setup:

10 µl	2x PCR Master Mix (Promega)
1.5 µl	Plasmid specific forward primer (10µM)
1.5 µl	Plasmid specific reverse primer (10µM)
2 µl	Colony
5 µl	Nuclease-free water

Colonies were streaked onto a LB-amp plate and numbered for identification. The colony PCR programme used the following setup:

Initial denaturation   2 minutes   95°C

30 cycles:

(Denature   30 seconds   95°C)

(Annealing	1 minute [per Kb]	50°C)	
(Extension	30 seconds	72°C)	
Final extension	10 minutes	72°C	

Colonies identified by agarose gel electrophoresis to have products of the correct size were selected for Mini/Midiprep and sequencing.

Primer name	Sequence (5' to 3' direction)
T7 primer	TAATACGACTCACTATAGGG
SP6 primer (pGEM-T easySP6)	CTATAGTGTCACCTAAATAG

2.2.10 DNA purification

200 µl of DNA/mRNA samples in nuclease-free water were purified/extracted with an equal volume of phenol/chloroform, samples were vortexed and centrifuged at 13,000 rpm. The aqueous phase was transferred to another tube and an equal volume of chloroform (200 µl) was added. Sample were vortexed and centrifuged at 13,000 rpm and the aqueous phase was transferred to a new tube. DNA samples were precipitated with 20 µl of 3M NaOAc and 500 µl 100% ethanol (Sigma, Cat no: SSO5461). DNA samples were chilled at -20°C for overnight. Samples were centrifuged at 13,000 rpm to pellet for 30 minutes. DNA was washed and centrifuged twice with 1ml 70% ethanol at 13,000 rpm for 5 minutes before being dried by desiccation. The DNA was resuspended in 30 µl nuclease-free water. All DNA was stored at -20°C.

2.2.11 Detection of CIC proteins by western blot

5 *X. laevis* embryos and 20 *X. tropicalis* embryos were analysed per blot. Samples were collected at the appropriate developmental stage and stored at -80°C post microinjection. Samples were homogenised in 50 µl of ice cold Phosphosafe buffer (Novagen) supplemented with protease/phoshatase inhibitors, cOmplete mini™ EDTA-free and PhosSTOP™ (Sigma, Cat no: 4906845001). Samples were refrozen at - 80°C for 10 minutes and then centrifuged at 13,000 rpm for 20 minutes at 4°C. Supernatant was added to 120 mM Tris/Cl pH 6.8; 20% glycerol; 4% SDS, 0.04% bromophenol blue; 10% β-mercaptoethanol (2x SDS-sample buffer) and heated at 95°C for 5 minutes. 20 µl of sample was loaded onto 7.5-10% SDS-PAGE

gels dependent on expected size of protein, along with PageRuler Plus prestained protein ladder (ThermoScientific) and allowed to run at 150 volts for 1 hour 30 minutes. Proteins were transferred to Immobilon-P Transfer Membrane (Millipore) by electroblotting wet (tank) transfer at 30 volts overnight at 4°C. Larger proteins were transferred in 5% methanol transfer buffer to prevent precipitation, whilst smaller proteins were transferred in a 10% transfer buffer. Membranes were washed in PBSAT (PBSA + 0.1% Tween) and blocked in PBSAT + 5% milk powder (blocking solution) for 1 hour at room temperature. Primary antibodies were added to fresh blocking solution (see concentrations in below) and left overnight to bind at 4°C.

Phosphate buffered saline	
Start with initial volume of 800ml of distilled water	
Component	Weight (g)
NaCl	8
KCl	0.2
Na_2HPO_4	1.44
KH_2PO_4	0.24
Adjust the pH to 7.4 with HCl.	
Make to final volume 1 litre using distilled water	

Membranes were washed in PBSAT and reblocked with the blocking solution before being blotted with the secondary antibody. BM Chemiluminescence Blotting Substrate kit (Roche) and ECL Hyperfilm (Amersham) were used to detect proteins. Myc Antibody (SIGMA ALDRICH, Cat no: A7470-1ML), GAPDH Antibody (NOVUS BIOLOGICALS, Cat no: NOVUNB300-285), dpERK Polyclonal Antibody (Cambridge Bioscience, Cat no: 3518-100) & GFP Polyclonal Antibody (Takara, Cat no: 632593).

Primary	Dilution	Secondary	Dilution
GFP	1: 4,000	Anti-mouse	1: 4,000
dpERK	1: 4,000	Anti-mouse	1: 4,000
Myc	1: 4,000	Anti-mouse	1: 4,000
GAPDH	1:10,000	Anti-mouse	1: 4,000

2.2.12 Immunoprecipitation of GFP-CIC protein

Immunoprecipitation of the GFP-tagged CIC protein was performed using the GFP-Trap® agarose bead system (ChromoTek, Cat no: 632593). The equivalent of 50 *X. tropicalis* embryos were injected with the mRNA generated from the PCR template containing the ORF of human CIC-S (pcDNA5 FRT/To GFP CIC) (Dissanayake et al., 2011). Samples were collected at the appropriate developmental stage and stored at -80°C post microinjection. Samples were lysed in 200 µl ice cold RIPA buffer (Sigma) supplemented with protease/phoshatase inhibitors, cOmplete mini™ EDTA-free and PhosSTOP™ (Sigma). The GFP-Trap agarose beads were equilibrated in 10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA (dilution buffer). Beads were left to bind to the protein for 1 hour at 4°C under constant mixing. After bead-protein binding, supernatant was removed and beads were washed in ice cold dilution buffer. Proteins were dissociated from the beads and eluted into 2x SDS-sample buffer by boiling at 95°C for 10 minutes. Samples were stored at -80°C until ready for analysis by western blot.

2.2.13 *In vitro* synthesis of digoxigenin-labelled antisense RNA probes

Synthesis of digoxigenin-labelled antisense RNA probes were created using the following reaction: 15µl 5x Transcription Buffer (NEB), 2.5µl 10x DiG dNTP mix (Roche, Cat no: 733-1270), 5µl Dithiothreitol (DTT) (Invitrogen), 2µl RNasin (Promega), 4µl DNA polymerase (Promega), 2µl template linear DNA template and 44µl H2O. The reaction was incubated at 37°C for 4 hours. 10 *X. laevis* embryos or 10 *X. tropicalis* embryos fixed in MEMFA were analysed per sample.

Plasmid	Linearization enzyme	Polymerase enzyme
pGEM-T easy CIC-L (939bp)	NdeI	T7
pGEM-T easy CIC-L (556bp)	NcoI	SP6
pGEM-T easy CIC-S (956bp)	NcoI	SP6
pGEM-T easy CIC-S (454bp)	NdeI	T7
pGEM-T easy CIC-S specific		

2.3 *CIC* knockout and knockdown methods

2.3.1 TALEN targeting of the *CIC* allele in *Xenopus*

TALEN was designed to target the HMG-box of *CIC* in exon 6. The pCS2+ TALEN plasmids (left + right) were designed by Harv Isaacs and made in the University of York technology facility. Both plasmids incorporated sequence for a Flag (left) and HA (right) tag for immunodetection once translated within the embryo. Plasmids were linearised using the NotI restriction enzyme before synthesising mRNA (see section 2.3.2) and micro injection (section 2.1.3). Detection of targeting was performed using the same steps as section above.

Region in trop genome targeted:
```
Tatgatcttcagcaagcggcatagggcccttgtacatcagcgccacccaaacc
```

tal1 target:	`tatgatcttcagcaagcgg`
tal2 target:	`atcagcgccacccaaacc`

2.3.2 Antisense morpholino targeting of the *CIC* specific exons

Antisense morpholinos were created to block translation of the CIC-L and CIC-S homologs by targeting the specific exons 1 (CIC-L) and exon 2 (CIC-S). Morpholinos were designed to be complementary to the mRNA start site preventing translation by sterically blocking protein recruitment (Heasman et al., 2000). All morpholinos were designed and created by GeneTools.

Morpholino	Sequence (5' to 3')
CIC-L morpholino	`5'-GGTAGCTTTCTTTACAGATTTCATT-3'`
CIC-S morpholino	`5'-GCTCAGATGAGAACATGCTGACCAC-3'`

Prior to microinjection morpholinos were diluted to required concentrations and heated to 80°C for 5 minutes. Morpholinos were stored at 4°C.

2.4 CIC overexpression methods

Subcloning of coding regions of genes for mRNA synthesis

2.4.1 Generation of pCS2+ CICf construct

To generate the pCS2+ CICf construct, the *Mus musculus* CIC-S (CICf) (GenBank: AF363690.1) isoform open reading frame (ORF) was subcloned from pMyc-CICf (fig 44) (Kim et al., 2013) into the polylinker at the StuI and XbaI restriction sites of pCS2+. pMyc-CICf contained the ORF for a myc-epitope tag at the 5' of the CICf gene used for detection of the translated protein. The CICf region of pMyc-CICf contains the 4828 bp coding sequencing of the CIC-S transcriptional repressor. pCS2+ vector is used for *in vitro* transcription to generate mRNAs with 5'- termini GpppG cap and 3' termini poly-A signal. The Myc-CICf plasmid was obtained from the Addgene plasmid (Plasmid #48185) repository donated by the Huda Zoghbi lab (Kim et al., 2013).

2.4.2 Generation of pCS2+ XCIC-S construct

The pCS2+ XCIC-S construct was generated by TA cloning the entire CIC-S isoform ORF of *X. tropicalis* into pGEM-T easy (Promega). The ORF sequence was obtained by PCR amplification from the cDNA generated from the total RNA. cDNA was created using SuperScript IV Reverse Transcriptase (ThermoFisher, Cat no: 15387686). GoTaq® Long PCR Master Mix (Promega, Cat no: M4021) was used to generate the 5.5 Kb amplicon. The CIC-S ORF sequence was subcloned into the polylinker at the StuI and XbaI restriction sites of pCS2+. The plasmid map for pCS2+ XCIC-S can be found in the appendices (Appendix 1). The following PCR conditions were used for the PCR amplification:

25 µl	GoTaq® Long PCR Master Mix, 2X (Promega)
1.5 µl	Forward primer (10µM)
1.5 µl	Reverse primer (10µM)
1 µl	Template DNA (0.1–0.5µg)
21 µl	Nuclease-free water

Initial denaturation	2 minutes	95°C

35 cycles:

(Denature	30 seconds	94°C)

(Annealing	30 seconds	2°C below Tm)
(Extension	1 minute [per Kb]	72°C)
Final extension	10 minutes	72°C

Primer	Sequence (5' to 3' direction)	Notes
18|1|18CIC-S For V1	gaga<u>ATCGAT</u>GACCAGATTGTAGAGGAACG	Contains *ClaI* recognition site.
18|1|18CIC-S Rev V1	gagaTCTAGAGAGCCCAGATAACCCTAAAG	Contains *XbaI* recognition site.

2.4.4 Generation of *H. sapiens* GFP-CIC DNA template

Primers were created to amplify a *H. sapiens* CIC-S PCR based template from the pcDNA 5 FRT/TO GFP CIC (fig 2) (Dissanayake et al., 2011). pcDNA 5 FRT/TO GFP CIC contains the ORF for the human homolog of CIC-S which contains an upstream GFP tag for florescence and immunoprecipitation experiments. Primers were designed to amplify an amplicon containing the GFP, *H. sapiens* CIC-S and bGH poly(A) signal (see below). The forward primers contained the SP6 promoter for mRNA synthesis. Three PCR templates designs were created with 2 variations with different distances of the forward primer (containing the SP6) to the initiating codon of the ORF and 1 variation which would not contain the ORF for the bGH poly(A) signal once amplified. The mRNA generated using this template design was polyadenylated *in vitro* to enhance translation initiation efficiency using a Poly(A) Tailing Kit (ThermoFisher).

Primer	Sequence (5' to 3' direction)	Notes
HsGFPCIC-S19/1/17 Forward	GAGA**ATTAGGTGACACTATAGAA***GACGAGCTC GTTTAGTGAACCG *SP6 promoter	126 bp length from AUG
HsGFPCIC-S19/1/17 Reverse	GGGCAAACAACAGATGGCTGGCAACT	

HsGFPCIC-S19/10/16 Forward	GAGA**ATTAGGTGACACTATAGAA***GCCACCATG GTGAGCAAGGGCGAG *SP6 promoter	6 bp length from AUG
HsGFPCIC-S19/10/16 Reverse	ATACCCCCTAGAGCCCCAGCTGGTTCTTTCCG	
HsGFPCIC-S19/1/17 Reverse	GGGCAAACAACAGATGGCTGGCAACT	No bGH poly(A) signal

2.4.5 Differential expression analysis of RNA-seq data

X. tropicalis mRNA samples were collected in triplicate for *CIC* knockdown by TALENs, molecular water injection and *FGF4* overexpression. Total RNA quality control was analysed (bioanalyzer) to confirm RNA quality. Library preparation and Illumina sequencing was performed by the staff at the Bioscience Technology Facility at the University of York. Samples were sequenced using a single lane of the Illumina HiSeq 2000 platform. Random hexamers and reverse transcriptase were used for first strand cDNA synthesis. After the construction of the library each cDNA was sequenced in a high-throughput manner to obtain a read count. The number of reads from a given gene was a measure of its level of expression. To take into account for the varying number of reads sequenced for each sample, and the varying expression of transcripts across the whole transcriptome, transcripts per million (TPM) was used as a measure of expression.

2.4.6 Gene Ontology analysis of RNA-seq data

Initial RNA-seq bioinformatics analysis was performed at the Bioscience Technology Facility at the University of York.

Results were produced by aligning the raw reads for each sample to the reference transcriptome aligned to the *X. tropicalis* reference transcriptome (genome v9.1) (http:// www.xenbase.org/common/displayJBrowse.do?data= data/xt9_1) with a tool called Salmon (http://salmon.readthedocs.io), which produces estimated read counts for each transcript for each sample.

The differential expression results (Q values and effect sizes) were then calculated by fitting a statistical model to the estimated read counts using Sleuth (http://pachterlab.github.io/sleuth/). 43558 transcripts were found for 23635 genes. Transcripts that were found to have overlapping upregulated or downregulated expression in *CIC* knockdown and FGF4 overexpression were selected for gene

ontology (GO) analysis. All GO analysis was performed using the Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification tool (Mi et al., 2013) (http://www.pantherdb.org/) and the Xenbase website GO term tools (http://www.xenbase.org/entry/).

Chapter 3: Characterisation of the CIC *Xenopus tropicalis* gene

3.1 Introduction

CIC was not identified by expressed sequence tag (EST) databases analysis in *Xenopus*. ESTs analysis is used to examine expressed genes and gene discovery (Lindlof, 2003). To generate ESTs, pools of mRNA from the whole organism, or from specific tissues, are isolated by 3' poly-A tail selection (Parkinson and Blaxter, 2009). Reverse transcription of the mRNAs is used to create cDNA libraries for analysis. EST analysis is cheap and relatively quick, although it has limitations. The technique does not give full coverage of the whole transcriptome and is only a snap shot of genes which are expressed at a given time point or specific tissues. Genes which have low expression may be difficult to detect (Lindlof, 2003) and can be of low quality. Genes which have large 3' UTR may be difficult to detect due to ESTs being short in length (300-800bp), this is the likely reason as to why CIC was not detected and will need further analysis to establish the length of the UTRs.

One of the first major aims of this study was to identify the complete exons structure of *CIC* in *X. tropicalis*. *CIC* has not been identified in the *X. tropiclias* genome, although there are predicted *Xenopus* cDNA and protein sequences (Accession XM_004916258) created by automated computational analysis (NCBI's GNOMON). Preliminary analysis of domains has shown high conservation of the *CIC* gene between species. *CIC* is conserved in vertebrates; orthologs of *CIC* exist in *Homo sapiens* (Lee et al., 2002) and *M. musculus* (Lam et al., 2006). The gene sequence/structure information of *H. sapiens* and *M. musculus* would be referenced to create a predicted sequence for *X. tropicalis.* The gene structure/sequence data generated by cloning of cDNAs would allow analysis of CIC isoform spatial expression by *in situ* and further essential experimental analysis of the CIC isoform by temporal expression by RT-PCR and qPCR gene expression. Once the predicted exon structure was generated, PCR primers would be designed to test the accuracy of the prediction and amplify CIC from cDNA to confirm exon junctions, therefore determine the sequence of CIC mRNAs expressed in early *Xenopus* development. Cloning of the full-length CIC cDNA would allow the expression and functional analysis of the CIC isoforms. Defining cDNA sequence would allow a prediction of the peptide sequence of CIC isoforms expressed in early development. Predicted protein

sequence would then allow comparison of domain structure with what is known in other models, *M. musculus*, *Drosophila* etc.

The aims of this chapter are to:

- Create a predicted exon structure of *X. tropicalis CIC* by alignment of the *M. musculus CIC* amino acid sequence to the *X. tropicalis* genome.
- Test the predicted exon structure using PCR amplification.
- Sequence the entire ORF using the predicted exon structure.
- Clone the full-length cDNA sequences of CIC-L and CIC-S isoforms of *X. tropicalis* for overexpression and functional studies.

3.2 Results

3.2.1 The creation of the *Xenopus CIC* predicted exon structure from the conserved amino acid sequence of *Mus musculus*.

To create a predicted exon structure for *X. tropicalis*, the *M. musculus* amino acid sequence was aligned to the *X. tropicalis* genome at chromosome 7 using Xenbase scaffold version 8 (Karimi et al., 2018). The alignment was performed using NCBI TBLASTN tool (Gish and States, 1993). The amino acid sequences, instead of the DNA sequence, was aligned to the *X. tropicalis* genome due to redundant nature of the genetic code. The alignment of *Mus musculus* amino acid sequence produced 52 consensus regions which mapped onto the *X. tropicalis* gene locus. Mapped consensus sequences ranged from 6-28 amino acids in length. This data provided potential locations for *CIC* exons within chromosome 7.

3.2.2 Analysing the prediction and identifying the exon structure of *X. tropicalis CIC*.

Mapping of consensus amino acid sequences enabled a prediction of the exon structure, but further analysis was required to test the validity of the exon junctions of the prediction by amplification of the cDNA by PCR. The *CIC* gene codes for 2 prominent isoforms, CIC-L (Long) and CIC-S (Short) in *H. sapiens, M. musculus* and *D. melanogaster* homologs (Lam et al., 2006). The mRNA differs in size due to 5' alternative splicing of unique exons, exons 1 and 2. The CIC-L contains the largest most 5' exon, exon 1 and CIC-S contains the smallest exon, exon 2. Exons 3-22 are common amongst both predominant isoforms of CIC. This data would be essential to

clone the full-length ORF of the prominent isoforms for overexpression and functional studies.

3.2.3 Identification of the CIC-L specific exon

To determine if the predicted model was accurate and identify the CIC-L specific exon, initial analysis to decipher the *X. tropicalis* gene *CIC* focused on the 5' region of the proposed gene. In *H. sapiens* exon 1 is 2,793 bp and the largest of the *CIC* exons (fig 9). Like the *H. sapiens* exon 1, the *M. musculus* exon 1 is the largest of the CIC homolog exons, which is 2,787 bp.
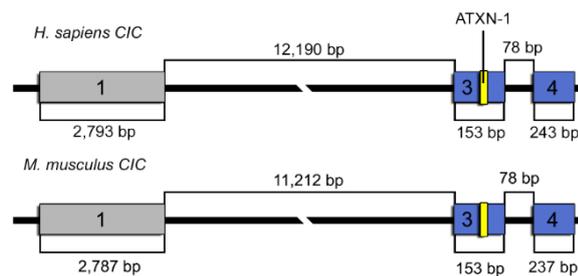
Figure 9, schematic of the 5' end of the *CIC* gene structure in *H. sapiens* and *M. musculus*. Exons 1 (grey), 3 and 4 (blue), ATXN-1 domain (yellow).

Analysis of the gap between exons 1 and 3 revealed it was 12,190 bp in length in *H. sapiens* (fig 9). Consistent with *H. sapiens*, the gap between exons 1 and 3 in *M. musculus* was 11,212 bp. This analysis suggested there would be a large intron between the CIC-L specific exon 1 and exon 3 in *X. tropicalis*. To resolve the location of exons 1 and 3, primers were designed using the predicted exon structure to produce an amplicon of expected size from cDNA, which would include the highly conserved 3' region of exon 1, exon 3 and the 5' region of exon 4. Due to the predicted large distance between exons 1 and 3 (45,924 bp), no amplicon would be generated from genomic DNA contamination.
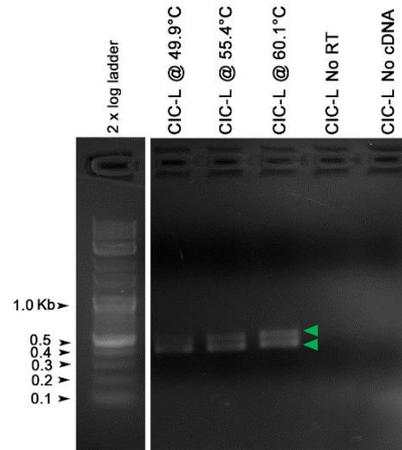
Figure 10, two amplicons were produced in the PCR reaction to establish the exons 1, 3 and 4 locations in the *CIC* locus (green arrows). Primers were designed to produce an amplicon of 366 bp in length.

The PCR reaction created 2 amplicons products (fig 10). Sequencing analysis showed the larger amplicon was 408 bp in length and the smaller amplicon was as predicted, 366 bp in length. Analysis of the smaller amplicon confirmed the accuracy of the predicted exon structure and that the CIC-L specific exon was 45,924 bp upstream of exon 3 as expected (fig 12). Analysis of the larger amplicon sequence data suggested that it contained the intron between exons 3 and 4. Translation of this ORF would produce a protein with a premature stop codon (fig 11).

```
                          10        20        30        40        50        60        70        80        90       100
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Predicted      GATAATCATGGGGAGACCAGGCACAGTCTGGACGAATGTGGAGCCGCGATCTGTGGCAGTGTTTCCCTGGCACTCGCTGGTCCCTTTTCTGGCTCCAAGT

+ intron       GATAATCATGGGGAGACCAGGCACAGTCTGGACGAATGTGGAGCCGCGATCTGTGGCAGTGTTTCCCTGGCACTCGCTGGTCCCTTTTCTGGCTCCAAGT

- intron       GATAATCATGGGGAGACCAGGCACAGTCTGGACGAATGTGGAGCCGCGATCTGTGGCAGTGTTTCCCTGGCACTCGCTGGTCCCTTTTCTGGCTCCAAGT

                         110       120       130       140       150       160       170       180       190       200
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Predicted      CAACCAGATTCATCCGTCCAACCCTCAGAGGGCCAGCAGCCTGTCAATCATCCCGGAGCCTCAAATCAAAGCAAAG--------------------

+ intron       CAACCAGATTCATCCGTCCAACCCTCAGAGGGCCAGCAGCCTGTCAATCATCCCGGAGCCTCAAATCAAAGCAAAGGTAAGTCTGTTGCTGTCATTCTCT

- intron       CAACCAGATTCATCCGTCCAACCCTCAGAGGGCCAGCAGCCTGTCAATCATCCCGGAGCCTCAAATCAAAGCAAAG--------------------

                         210       220       230       240       250       260       270       280       290       300
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Predicted      -------------------------------------------------------AACCCCCAGAATCTGCATCAGTTGCACATGATGCCATGCCAGT

+ intron       GAACCAGTTGGTCTTCAGTGTTCAGTAGTCACCCAGATTTTTATCTGGCCTCTGCAGAACCCCCAGAATCTGCATCAGTTGCACATGATGCCATGCCAGT

- intron       -------------------------------------------------------AACCCCCAGAATCTGCATCAGTTGCACATGATGCCATGCCAGT

                         310       320       330       340
               ....|....|....|....|....|....|....|....|....|.
Predicted      GGCAAGCATTGAAGATGAGCGGTGTGCAGCTCCCCGCACTGATAAT

+ intron       GGCAAGCATTGAAGATGAGCGGTGTGCAGTTCCCCGCACTGATAAT

- intron       GGCAAGCATTGAAGATGAGCGGTGTGCAGCTCCCCGCACTGATAAT
```

Figure 11, the predicted sequence aligned with the large amplicon which contains the intron (+ intron) between exons 3 and 4 (408 bp) and small (366 bp) amplicon of predicted size which does not contain the intron (- intron). The smaller product (- intron) aligns perfectly with the predicted sequence. (Red) exon 1, (purple) exon 3, (blue) exon 4 and (black) intron.

Figure 12, the exon structure of the 5' end of the *CIC* gene. Sequencing data has established that there is a 45,924 bp intron between the the 3' end of the CIC-L specifi exon (exon 1) and the 5' region of the first common exon (exon 3). Primers location (red), Exons 1 (grey), 3 and 4 (blue), ATXN-1 domain coding region (yellow).

3.2.4 Identification of the CIC-S exon in Xenopus tropicalis

Recent analysis of the *D. melanogaster CIC* homolog revealed a small CIC-S specific exon in the CIC-S isoform not previously detected (Fores et al., 2015). The CIC-S specific exon was found upstream of exon 3, the most 5' of the common exons shared between the predominant isoforms, CIC-L and CIC-S. To identify if the CIC-S specific exon existed in *X. tropicalis,* the amino acid sequence for both prominent CIC isoforms of *H. sapiens* were aligned to each other. Alignment of the CIC-L protein sequences to the CIC-S protein highlighted a unique amino acid sequences (**MYSAHRPLMPASSAASRGLGMF**) only found in the CIC-S isoform (fig 13). This suggested that the sequence was likely to be unique exon. The amino acid sequence coded for a 66 bp exon, exon 2. Exon 2 was 10,128 bp downstream of the CIC-L specific exon 1 and 1,996 bp upstream of the first common exon, exon 3 (fig 14).

To reassess if these findings were also true in the *M. musculus* model, the *M. musculus* amino acid sequence for both CIC-L and CIC-S isoforms were aligned. The alignment confirmed once again, that a unique amino acid sequence (**MFSAERPLMPASSAASRGLGMF**) was in the CIC-S isoform, but not in the CIC-L. The amino acid sequence coded for a 66 bp exon, exon 2. The *M. musculus* exon 2 was 9,217 bp, downstream of exon 1 and 1,929 bp upstream of the common exon, exon 3 (fig 14). The unique CIC-S amino acids sequence was combined to create a consensus sequence (**M-SA-RPLMPASSAASRGLGMF**) (fig 13).

Figure 13, the alignment of the *H. sapiens*, *M. musculus*, *X. tropicalis* amino acid sequences with a consensus sequence. A predicted X. *tropicalis* amino acid sequence was created by mapping the *H. sapiens* and *M. musculus* amino acid sequences on to the *X. tropicalis* genome sequence.

The consensus sequence was aligned to the *X. tropicalis* genome in between exons 1 and 3 to determine if the CIC-S specific exon existed. A conserved sequence was identified (**MFSSERPAPPCGLSMF**) upstream of exon 3 (fig 13-14). The amino acid sequence coded for a 48 bp exon. The *X. tropicalis* CIC-S specific exon 2 was 39,452 bp downstream of exon 1 and 6,423 bp upstream of the common exon, exon 3 (fig 14).
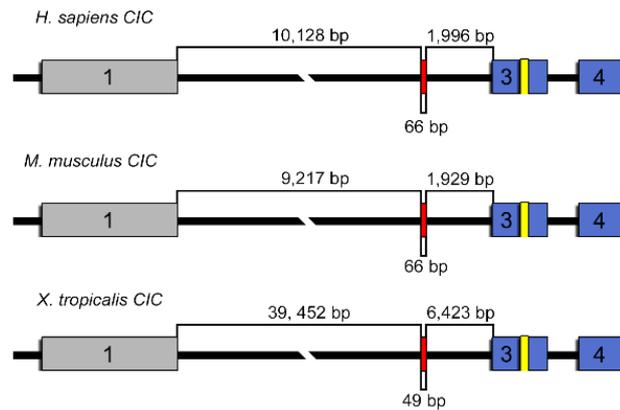
Figure 14 schematic of the CIC-S exon structure in *H. sapiens*, *M. musculus and X. tropicalis*. Exons 1 (grey), CIC-S specific exon (red) 3 and 4 (blue), ATXN-1 domain (yellow).

Primers were designed to evaluate if the conserved amino acid sequence generated from the alignment was accurate and to provide additional information on the 5'- UTR length of the CIC-S isoform. A series of forward primers were created to generate amplicons from cDNA, increasing in size by 50 bp to access the size of the UTR of the CIC-S exon. The reverse primers were designed to span the exon junction of the predicted CIC-S specific exon 2 and the common exon 3. If predictions of the exon structure were incorrect, or the forward primer was not within the 5' UTR, no amplicon would be amplified from the cDNA due to the large intron down stream of proposed exon 2. The reverse exon spanning primer would also prevent amplification from genomic DNA.
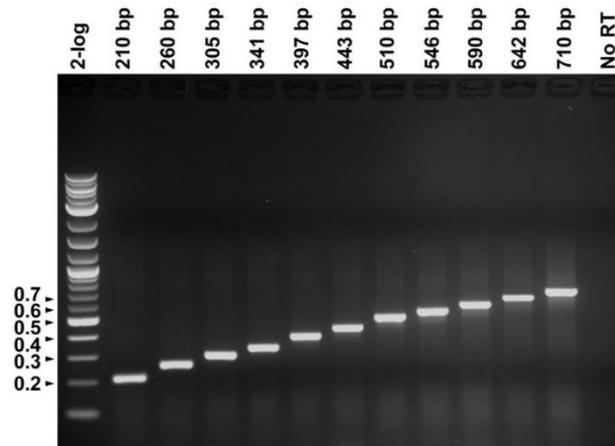
Figure 15, the PCR products generated when analysing the existence of the CIC-S specific exon.

All PCR reactions created amplicons of predicted size (fig 15). This confirmed that there was a CIC-S exon upstream of the exon 3 in the *X. tropicalis CIC*. Sequence data again confirmed that the exon structure prediction was correct (fig 15) (for DNA sequence see Accompanying Material CD) and that there was a CIC-S specific exon (exon 2). A large 39,452 bp intron was confirmed between CIC-S specific exon 1 and the CIC-S specific exon 2 and an intron of 6,424 bp downstream to exon 3.

3.2.5 Generating the open reading frame of the *X. tropicalis CIC* homolog from cDNA fragments

Once the exon structure for the 5' region of *CIC* was deciphered, the next objective of the project was to investigate the common exons and 3' region of the predicted *CIC* exon structure. The analysis of the ORF would confirm exon/intron size and exon junctions. To verify the full structure of *CIC*, rather than attempting to clone the full ORF of the *CIC* isoforms from cDNA into a plasmid vector, which had initially proven to be difficult to clone due to large size of the cDNA amplicons, the cDNAs were cloned into fragments. The entire ORF was cloned in 3 separate amplicons, once sequenced the data was realigned to build a contig, which would give full coverage of the exon structure of *CIC.* Primers were designed to generate 3 amplicons ranging from 1.6 to 2 Kb in size. Primers were designed to be exon-exon junction spanning to prevent genomic DNA contamination.

Sequencing data obtained from CIC-S specific exon allowed the design of the most 5' amplicon, amplicon 1 was created using a forward primer positioned on the exon-exon junction of the CIC-S specific exon 2 and exon 3. The reverse primer was designed to be positioned within an area which was highly conserved when aligning the *M. musculus* amino acid sequence to the *X. tropicalis* genome. The sequence

corresponded to the predicted exon 11, which contained the C2 domain coding region. The predicted product size for this amplicon was 1,975 bp (fig 16). Amplicon 2 was designed to be within the body of the gene and was generated by designing a forward primer, positioned on a proposed exon-exon junction of exon 9 and 10. The exon-exon junction of exon 9 and 10 was an area of high conservation when aligning the *M. musculus* amino acid sequence to the *X. tropicalis* genome. The reverse primer was designed to exon-exon spanning and positioned on the exon 13-14 junction, another area of high conservation. The predicted amplicon product for this amplicon was 1,672 bp (fig 16). Amplicon 3, the most 3' of the fragment was created by designing a forward primer, which was positioned on the conserved exon-exon junction of predicted exons 13-14. The reverse primer was positioned to be within the highly conserved C1 coding region found in exon 22 of *M. musculus* homolog. The predicted amplicon product for this amplicon was 1,655 bp (fig 16). cDNA was created from 10 stage 25 *X. tropicalis* embryos. Amplicons were generated using GoTaq Long master mix (Promega), due to being able to generate large amplicons and its high-fidelity proof reading ability.
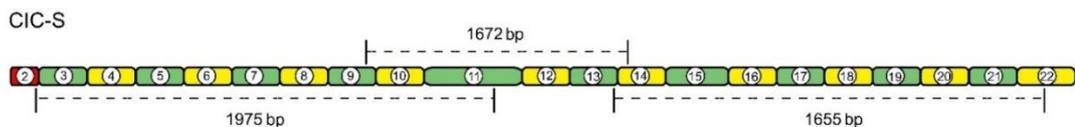
Figure 16 shows the CIC-S cDNA structure, positioning and length of the amplicon fragments for generating the full ORF contig. Each primer pair had an exon junction spanning primer.
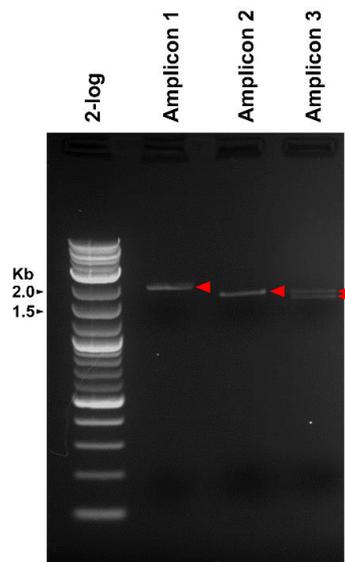
Figure 17, the amplicons created to generate the full exon structure of *CIC*. Amplicons 1-3 were of expected size, but an additional product was observed in the amplicon 3 reaction (see red arrows).

Amplicons 1, 2 and 3 were of expected size (fig 17), which confirmed the predicted exon model was accurate. The *CIC* contig was constructed by sequencing 28 fragments of the full length of the gene in both directions (5'-3' and 3'-5' strands) (for gene structure see the Accompanying Material CD). The complete ORF of the aligned sequence data was analysed to find the correct ORF. The amino acid sequence generated from the full cDNA sequence was tested to see if there would be a full ORF and no introduction of a premature stop codon. A splice-site analysis tools (SplicePort) was used to find the correct splice sites when aligning the cDNA sequence to genomic DNA sequence (Dogan et al., 2007).

The *X. tropicalis* gene homolog consists of 22 exons (table 2), which spans over 62,086 bp. The exons contain 7,593 bp. In comparison the *H. sapiens* homolog spans 23,162 bp and the exons contain 7,617 bp. The *M. musculus* homolog spans 22,576 bp and the exons contain 7596 bp.

Exon number	Size	Additional information
1	2,812 bp	CIC-L specific exon, N1 domain
2	49 bp	CIC-S specific exon
3	150 bp	ATXN-1 domain
4	208 bp	14.-3- motif
5	130 bp	
6	186 bp	HMG-box
7	166 bp	HMG-box
8	200 bp	
9	229 bp	
10	101 bp	
11	1,165 bp	C2 domain
12	209 bp	
13	173 bp	
14	167 bp	
15	300 bp	

16	190 bp	
17	247 bp	
18	221 bp	
19	155 bp	
20	132 bp	NLS
21	132 bp	
22	372 bp	C1

Table 2, exon size, locations and region of encoded domain/motifs, based on the *X. tropicalis CIC* gene.

Whilst cloning the 3' fragment for the full ORF of *CIC,* two products were amplified in the PCR reaction (fig 17). The larger amplicon was cloned and sequenced. Sequencing data revealed alternative splice isoform which lacked exon 21 (for sequence see Accompanying Material CD). If mRNA was translated it would produce a truncated protein.

3.2.6 Generating the full length endogenous isoforms of *CIC* from cDNA.

Over expressing not only of the CIC-S isoform, but CIC-L *X. tropicalis* isoform would be of great interest in establishing the function of the isoforms. To be able create full length cDNAs, a protocol was developed and optimised to enable large cDNA's for CIC-L (7,545 bp) and CIC-S (4,782 bp).

Total RNA was extracted from 10x stage 25 embryos, a time in development which both isoforms are known to be highly expressed. cDNAs were generated with a gene specific primer, which was designed to bind within the 3' UTR of CIC. The gene specific primer strategy for generating cDNA allowed the production of a greater concentration of CIC cDNAs. The predicted exon structure model was tested by creating primers which would create amplicons of a predicted size from cDNA. Amplification from genomic DNA would be easily detectable due to the size of the amplicon with the addition of introns in comparison to cDNA without introns. All forward primers were paired with the same reverse primer which was in the 3' UTR.

The first experiments adopted Superscript II (Invitrogen, Cat no: 10328062) for reverse transcriptase of cDNA. PCR master mix (Promega, Cat no: M7502) was utilized to create amplicons from the cDNA, this Taq mix generates amplicons of up

to 2.5kb in length, although this would be a limiting factor when attempting to create larger amplicons. Amplicons were generated at expected sizes (fig 18).
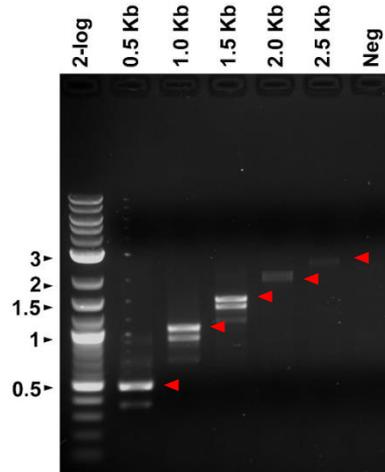
Figure 18, the PCR reaction for 0.5, 1, 1.5, 2 and 2.5 Kb products generated from cDNA (see red arrows). The largest amplicon was 2.5 kb. Additional bands can be seen likely due to amplification from genomic DNA contamination which would amplify introns.

The next approach was to create larger amplicons and to do this a superscript III was used due its ability to create longer cDNA due to reduced RNase H activity, increased half-life, and improved thermal stability. SuperScript III reverse transcriptase also produces higher cDNA yields. The PCR reaction was performed with Phusion DNA Polymerases (Thermo Scientific) due to its long-range PCR (up to 20 kb) and high-fidelity amplification (fig 19-20).
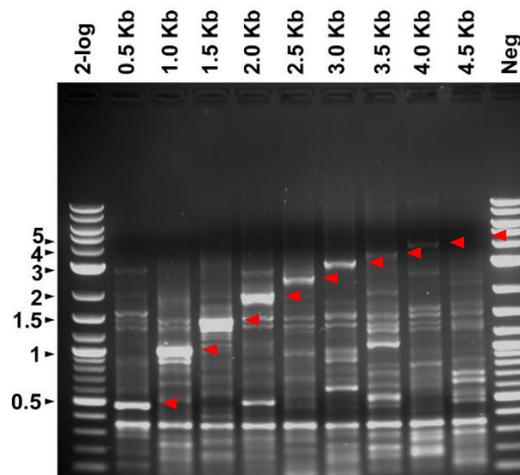
Figure 19, the PCR reaction for 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4 and 4.5 kb primers and products generated from cDNA (see red arrows). The largest amplicon generated was 4 kb.
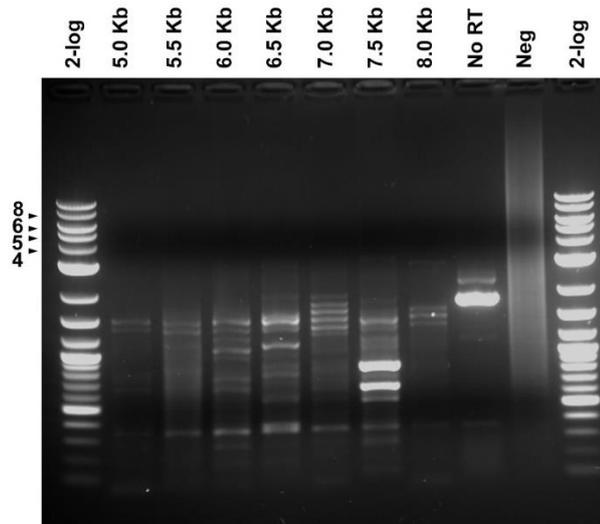
Figure 20 the PCR reaction for 5, 5.5, 6, 6.5, 7, 7.5, and 8 kb primers and products generated from cDNA. No amplicons were created of expected size.

Although Phusion can generate amplicons of up to 20kb, the largest amplicon was only 4kb generated from the cDNA from Superscript III reverse transcriptase after 3 attempts. An alternative to Phusion was tested, GoTaq long (Promega) which could also generate large amplicons of 20kb and had high fidelity function. GoTaq Long also had the benefit of leaving adenine over hangs, unlike Phusion which leaves blunt ends in on its PCR amplicons. This would enable TA-cloning.
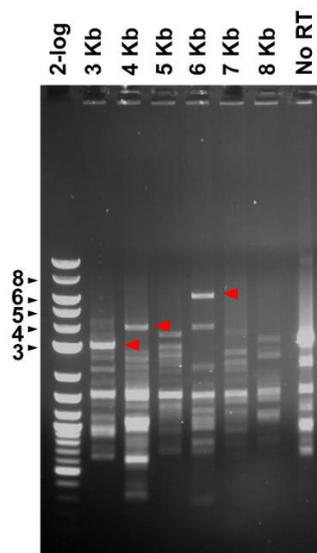
Figure 21, the PCR reaction for 3, 4, 5 and 6 kb primer products generated from cDNA (see red arrows). The largest amplicon generated was 6 kb. No product was generated at appropriate size for the 5kb reaction. Additional bands can be seen likely due to off target priming on genomic DNA. They are not CIC products with introns because these would be bigger than the predicted cDNA product.

GoTaq Long PCR reaction could generate (6 kb) amplicons products which were 2 kb larger than Phusion's 4 Kb amplicon (fig 21). To optimise the PCR reaction and reduce genomic DNA contamination total mRNA was treated with DNAse I (Promega), an endonuclease that non-specifically cleaves DNA. The DNase I enzyme was inactivated by phenol/chloroform purification. cDNA was created by Superscript IV reverse transcriptase which creates higher yields of cDNA than Superscript III reverse transcriptase. This cDNA was treated with RNase H. RNase H is an endoribonuclease that specifically cleaves RNA bound to DNA but does not digest single or double-stranded DNA. RNase H treatment prevents inhibition from RNA bound to DNA when DNA polymerase generates large amplicons (>5 Kb) from cDNA.
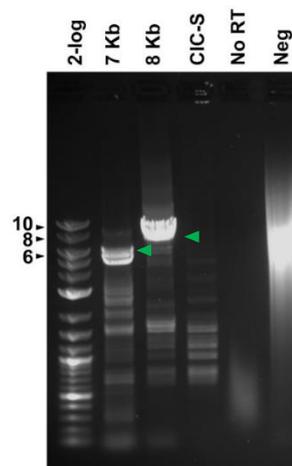
Figure 22, the PCR reaction for, 7,8 kb and CIC-S specific primers and products generated from cDNA (see green arrows). The largest amplicon generated was 8 kb. Additional bands can be seen likely due to amplification from genomic DNA contamination which would amplify introns.

The modified approach generated a large concentration of full length CIC-L corresponding to the precited length of 8 Kb (fig 22). The PCR reaction was also able to generate a 5 Kb product using primers designed to amplify CIC-S. This reaction would need further optimisation.

No CIC-L full length PCR product was cloned into the pGEM T-Easy vector by TA-cloning. This was likely due to the degradation of the 5'/3' ends of the PCR products due to UV light, or due to poor efficiency of the GoTaq Long Taq producing adenine overhangs. To overcome this problem, PCR could be incubated at 72°C with a Taq polymerase and adenosine triphosphate (ATP). This reaction would add adenine residue to the 3′ ends of the DNA molecules making it suitable for TA cloning.

As an alternative approach to TA cloning of the cDNA, primers were designed to amplify the full ORFs of CIC-L and CIC-S isoforms (fig 23-24). These primers contained sites for restriction enzyme cloning. The forward primers contained *ClaI* site and reverse primers contained the *XbaI* site at the 5' ends. The PCR reaction created amplicons of expected sizes for both isoforms, CIC-L (7621 bp) and CIC-S (4,906 bp). PCR products were digested and ligated into a linearised pCS2+ vector which was digested by ClaI and XbaI at the multiple cloning site (fig 23-24).
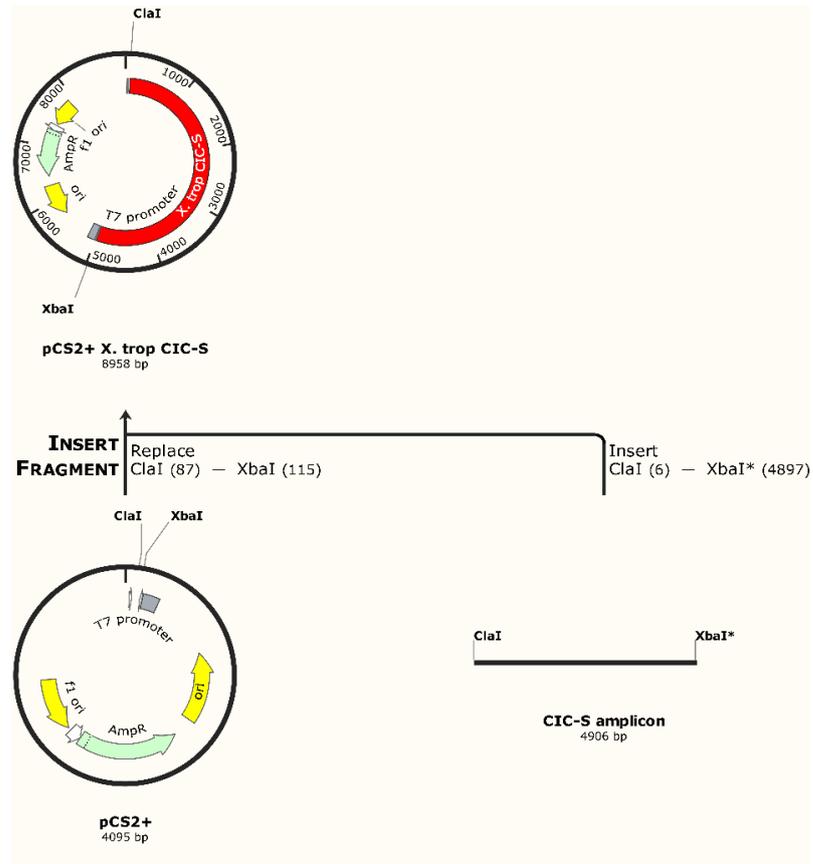
Figure 23, the strategy for cloning CIC-S into pCS2+. Primers were designed with restriction sites (*ClaI* & *XbaI*) to amplify the CIC-S ORF. The PCR amplicons digests and were ligated into the pCS2+ *ClaI* & *XbaI* linearised plasmid.
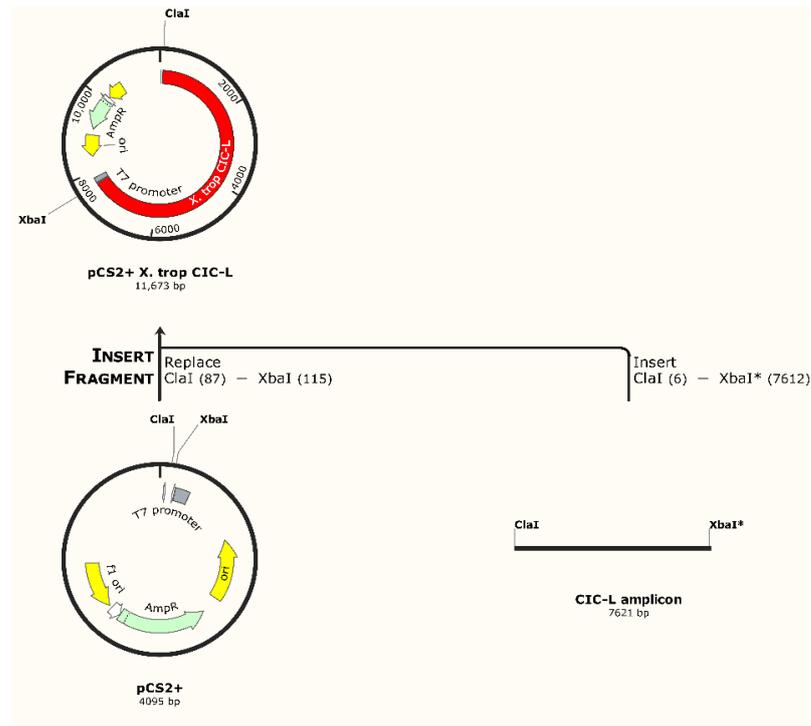
Figure 24, the strategy for cloning CIC-L into pCS2+. Primers were designed with restriction sites (ClaI & XbaI) to amplify the CIC-S ORF. The PCR amplicons digests and were ligated into the pCS2+ ClaI & XbaI linearised plasmid.

Successful cloning was confirmed by colony PCR and 4 clones were sequenced for CIC-S cDNA cloning. Sequencing data from 2 clones reconfirmed the previous ORF (Appendix 1) (for pCS2+ XCIC-S plasmid DNA sequence see Accompanying Material CD). Although the cloning strategy was successful for cloning the CIC-S isoform, CIC-L was not cloned due to shortages of time towards the end of the project.

3.2.7 Analysis of the *X. tropicalis* CIC amino acid sequence based upon cDNA

Once the ORF of the prominent isoforms were established, the ORF from the cloned cDNA was used to generate a predicted amino acid sequence (fig 25-26). This was aligned to the *M. musculus* and *H. sapiens* amino acid sequences. A consensus sequence was also generated (CIC-L, fig 25 & CIC-S, fig 26).

H. sapiens CIC-L protein consists of 2517 amino acid residues with a molecular weight of 258.7 kDa. The shorter isoform CIC-S consists of 1608 residues with a protein weight of 163.9 kDa. The *M. musculus* CIC-L protein isoform consists of 2510 residues with a molecular weight of 258.2 kDa. The shorter isoform CIC-S consists of 1604 residues with a protein weight of 163.8 KDa. The *X. tropicalis ORF* was calculated to produce a protein which would contain 2510 amino acids residues and molecular weight of 269.6 kDa. The CIC-S cDNA sequence would produce a protein that contained 1608 amino acid residues and molecular weight of 163.85 kDA.

Analysis of the *X. tropicalis* CIC-L and CIC-S amino acid sequences reveals that they share approximately 51% sequence identity to *H. sapiens* and *M. musculus amino* acid sequences (table 3). Whilst *D. rerio* has 38.2% sequence identity for the CIC-L amino acid sequence and 42.8% for CIC-S when compared to *X. tropicalis* amino acid sequences. In *D. melanogaster*, CIC-L has 25.9% sequence identity to *X. tropicalis*, whilst CIC-S has 25.3% sequence identity (table 3).

```
                                10        20        30        40        50        60        70        80
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
M. musculus CIC-L      MKPMKKACPGLAGSASGSKSPPATRAKALRRRGAGEGDKPEEE--EEAQPQEQAG------------------PEEAEEG
H. sapiens CIC-L       MKPMKKACTGLSGPGSGSKSPPATRAKALRRRGAGEGDKPEEEDDEAQQPQPQSG------------------PEEAEEG
X. tropicalis CIC-L    MKSVKKATSTTHSKASNSARSKKLRRKMEDGRQEEEEEQEEEAAEEEMEEEADEDSGHPLAHVDTKTIPVKMEPAEQEVT
Consensus              MK  KKA        S S     R K     R   E     EE   E                          P E E

                                90       100       110       120       130       140       150       160
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
M. musculus CIC-L      E----EEEAERDPGAEGTHPELQPNDP---TPGLTEDPKGDGEAG--------RWEPSLSRKTATFKSRAPKKKYVEEHG
H. sapiens CIC-L       E----EEEAERGPGAEGPPLELHPGDP---APGPAEDPKGDGEAG--------RWEPSLSRKTATFKSRAPKKKYVEEHG
X. tropicalis CIC-L    ESQVKDQEQATQSAEQVTGIEVEEGLPRTVKLEASMEASSDGQASNKAPYDLGRWEMPSSRKTATFKSKAPKKKYIEGHG
Consensus              E       E         E       P          DG A       RWE    SRKTATFKS APKKKY E HG

                               170       180       190       200       210       220       230       240
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
M. musculus CIC-L      TGN-VGVVGAPEERERTPEDASALG------VPPRPPTSTRSSSTDTASEHSADLEDEPPEACGPGPWPSTGTSEGYDLR
H. sapiens CIC-L       AGS-SGVAGAPEERVRTPEEASGLG------VPPRPPTSTRSSSTDTASEHSADLEDEPAEACGPGPWPPGSTSGSYDLR
X. tropicalis CIC-L    SQEGQEDSSSSSEGPATLQDCSGEQPVTSVIVSEERCHSVRSSSTDTASEHSADLEEEAEKTR-IDIAPSSGHCTDLGFH
Consensus                      E   T   S        V     S RSSSTDTASEHSADLE E           P
```

```
                              250       260       270       280       290       300       310       320
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L         QLRSQRVLARRGDGLFLPAVVRQVRRSQDLGVQFPGDRALTFYEGVPGGGVDVVLDVTPPPGALMVGTAVCTCVEPGVAA

H. sapiens CIC-L          QLRSQRVLARRGDGLFLPAVVRQVRRSQDLGVQFPGDRALTFYEGVPGAGVDVVLDATPPPGALVVGTAVCTCVEPGVAA

X. tropicalis CIC-L       HLRQQRILVRR-EESFQLCQVKQFRQNQ-VGVQFPGEQPLTFVDITQH---LVLLDKIPQLGEVEVGVPVCFCMNPDDTL

Consensus                  LR QR L RR    F   V Q R  Q  GVQFPG   LTF         V LD  P  G   VG  VC C   P


                              330       340       350       360       370       380       390       400
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L         YREGVVVEVATKPAAYKVRLSPGPSSHAGPPGTLPQAQQTLHREPEEAVWVTRSSLRLLRPPWEPGALLRKHPAGPEEEQ

H. sapiens CIC-L          YREGVVVEVATKPAAYKVRLSPGPSSQPGLPGSLPQPPQPLHREPEEAVWVARSSLRLLRPPWEPETMLRKPPTGPEEEQ

X. tropicalis CIC-L       FREGIVVEVVQNPLSYNVCIQQSPG----------------RREREYRLVPHSCIRLLRSPWCKESVLATPQNAEEKKQ

Consensus                  REG VVEV    P Y V     P                  R  E  V S RLLR PW      L       E  Q


                              410       420       430       440       450       460       470       480
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L         AEPGPALPPCPSSVEPKQPEDAEVSNISFGSNLGTRCEEGEEKHPPSLGTPVLLPLPPPQLLSPPPKSPAF-GGPGRPSE

H. sapiens CIC-L          AEPGATLPPCPAALDPKQPEDAEVSKISFGGNLGTHCEEGEEKHPPALGTPALLPLPPPQLLSPPPKSPAF-VGPGRPGE

X. tropicalis CIC-L       QAAGRFATALPLDQASGSSMELPVS----PRGTEQKHLEDAEVSKISFSMPEEKLSSIQQRILSPPKSPAFGIMLGRLPE

Consensus                   G       P           VS        E E       P        Q    PPKSPAF   GR E
```

```
                              490       500       510       520       530       540       550       560
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L        QPSPCQEGSQGGSRSSSVASLEKGAAPAARARTPLTAAQQKYKKGDVVCTPNGIRKKFNGKQWRRLCSRDGCMKESQRRG

H. sapiens CIC-L         QPSPCQEGSQGGSRSSSVASLEKGTAPAARARTPLTAAQQKYKKGDVVCTPSGIRKKFNGKQWRRLCSRDGCMKESQRRG

X. tropicalis CIC-L      QPSPLLSPDAGSRSSCSSVSLDKCSTPGSRSRTPLTAAQQKYKKGDVVCTPNGIRKKFNGKQWRRLCSRDSCMKESQRRG

Consensus                QPSP     G   S S  SL K   P  R RTPLTAAQQKYKKGDVVCTP GIRKKFNGKQWRRLCSRD CMKESQRRG


                              570       580       590       600       610       620       630       640
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L        YCSRHLSMRTKEMEGLADSGPGGTGRPAGVAAREGSTEFDWGDETSRDSEASSVAARGDSRPRLVAPADLSRFEFDECEA

H. sapiens CIC-L         YCSRHLSMRTKEMEGLADSGPGGAGRPAAVAAREGSTEFDWGDETSRDSEASSVAARGDSRPRLVAPADLSRFEFDECEA

X. tropicalis CIC-L      YCSRHLSMRTKEMESMSEGRGG----------REGSAEFEW-DDTSRDSEVSSIRT--DSRPRLVAPTDLSRFDFDECEA

Consensus                YCSRHLSMRTKEME          G          REGS EF W D TSRDSE SS     DSRPRLVAP DLSRF FDECEA


                              650       660       670       680       690       700       710       720
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L        AVMLVSLGSSRSGTPSFSPVSTQSPFSPAPSPSPSPLFGFRPANFSPINASPVIQRTAVRSRHLSASTPKAGVLTPPDLG

H. sapiens CIC-L         AVMLVSLGSSRSGTPSFSPVSTQSPFSPAPSPSPSPLFGFRPANFSPINASPVIQRTAVRSRHLSASTPKAGVLTPPDLG

X. tropicalis CIC-L      ANMLVSLGSSRSGTPSFSPVSNQSPFSPTPSPSPSPLFGFRPANFSPINASPVIQR--ARSRHVSASTPKGGTVLTPEML

Consensus                A MLVSLGSSRSGTPSFSPVS QSPFSP PSPSPSPLFGFRPANFSPINASPVIQR   RSRH SASTPK G    P
```

```
                              730       740       750       760       770       780       790       800
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L    PHPPPPAPRERHSSGILPTFQTNLTFTVPISPGRRKTELLPHPGTLGASGAGGGGAAPDFPKSDSLDSGVDSVSHTPTPS

H. sapiens CIC-L     PHPPPPAPRERHSSGILPTFQTNLTFTVPISPGRRKTELLPHPGALGAPGAGGGGAAPDFPKSDSLDSGVDSVSHTPTPS

X. tropicalis CIC-L  HHS---HHRERQAVGILPSFQTNLTFTVPMSPSKRKPD----------SHHGSASSAADFQKSDSMDSGVDSVSHTPTPS

Consensus              H     RER   GILP FQTNLTFTVP SP  RK           G    A DF KSDS DSGVDSVSHTPTPS


                              810       820       830       840       850       860       870       880
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L    TPAGFRAVSPAVPFSRSRQPSPLLLLPPPAGLTSDPGPSVRRVPAVQRDSPVIVRNPDVPLPSKFPGEVGTAGEARAGGP

H. sapiens CIC-L     TPAGFRAVSPAVPFSRSRQPSPLLLLPPPAGLTSDPGPSVRRVPAVQRDSPVIVRNPDVPLPSKFPGEVGTAGEVRAGGP

X. tropicalis CIC-L  TPAGFRSVSPAPIFSRSQQASPSQLLSPPAGLTSDPSPSVRRVPAVQRDSPVIVRNPDVPLPSKFVEKLSDGASRNIGGA

Consensus            TPAGFR VSPA  FSRS Q SP  LL PPAGLTSDP PSVRRVPAVQRDSPVIVRNPDVPLPSKF              GG


                              890       900       910       920       930       940       950       960
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L    GRSCRETPVPPGVASGKPGLPPPLPAPVPITVPPAAPTAVAQP-------------------MPTLGLASSPFQPVAFH

H. sapiens CIC-L     GRGCRETPVPPGVASGKPGLPPPLPAPVPITVPPAAPTAVAQP------------------MPAFGLASSPFQPVAFH

X. tropicalis CIC-L  ASGGSSPRRLSKVCSKDHPSKTQLQIPVPVPINATQMQSAARTLPPGQQSSGEPAACTSAGSTEHPPPVFSISSPFQPVAFH

Consensus                         V S     L  PVPI      A   P               P     SSPFQPVAFH
```

```
                         970       980       990      1000      1010      1020      1030      1040
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   PSPAALLPVLVPSSYPSHPAPKKEVIMGRPGTVWTNVEPRSVAVFPWHSLVPFLAPSQPDPSVQPSEAQQPASHPVASNQ

H. sapiens CIC-L    PSPAALLPVLVPSSYTSHPAPKKEVIMGRPGTVWTNVEPRSVAVFPWHSLVPFLAPSQPDPSVQPSEAQQPASHPVASNQ

X. tropicalis CIC-L PSPAALLPVIVPSDYASHPVPKKEIIMGRPGTVWTNVEPRSVAVFPWHSLVPFLAPSQPDSSVQPSEGQQPVNHPGASNQ

Consensus           PSPAALLPV VPS Y SHP PKKE IMGRPGTVWTNVEPRSVAVFPWHSLVPFLAPSQPD SVQPSE QQP  HP ASNQ


                        1050      1060      1070      1080      1090      1100      1110      1120
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   SKEPAESAAVAHEQPPGGTGGADPGRPPGAVCPESPGPGPPLTLGGVDPGKSLPPTTEEEAPGPPGEPRLDSETESDHDD

H. sapiens CIC-L    SKEPAESAAVAHERPPGGTGSADPERPPGATCPESPGPGPPHPLGVVESGKGPPPTTEEEASGPPGEPRLDSETESDHDD

X. tropicalis CIC-L SKEPPESASVAHDAMP--VASIEDER---CAVPRTDNVSQPP----TEEPAKQLQHQESMTQGTTGDQRGDSETESDHDD

Consensus           SKEP ESA VAH    P          R       P       P            E       G  G  R DSETESDHDD

                        1130      1140      1150      1160      1170      1180      1190      1200
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   AFLSIMSPEIQLPLPPGKRRTQSLSALPKERDSSSEKDGRSPNK-REKDHIRRPMNAFMIFSKRHRALVHQRHPNQDNRT

H. sapiens CIC-L    AFLSIMSPEIQLPLPPGKRRTQSLSALPKERDSSSEKDGRSPNK-REKDHIRRPMNAFMIFSKRHRALVHQRHPNQDNRT

X. tropicalis CIC-L AFFPIGPPDLQLPMHGGKRRTQSLSALPKERDSSSEKDGRSPNKQREKDHIRRPMNAFMIFSKRHRALVHQRHPNQDNRT

Consensus           AF  I  P  QLP   GKRRTQSLSALPKERDSSSEKDGRSPNK REKDHIRRPMNAFMIFSKRHRALVHQRHPNQDNRT
```

```
                          1210      1220      1230      1240      1250      1260      1270      1280
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   VSKILGEWWYALGPKEKQKYHDLAFQVKEAHFKAHPDWKWCNKDRKKSSSEAKPASLGLAGGHKETRERSMSETGTAAAP

H. sapiens CIC-L    VSKILGEWWYALGPKEKQKYHDLAFQVKEAHFKAHPDWKWCNKDRKKSSSEAKPTSLGLAGGHKETRERSMSETGTAAAP

X. tropicalis CIC-L VSKILGEWWYALGPKEKQKYHDLAFQVKEAHFKAHPDWKWCNKDRKKSSSDVKQVMPGPWGVPKEMRERSMSETGTMAAA

Consensus           VSKILGEWWYALGPKEKQKYHDLAFQVKEAHFKAHPDWKWCNKDRKKSSS  K      G  G   KE RERSMSETGT AA


                           1290      1300      1310      1320      1330      1340      1350      1360
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   GVSSELLSVAAQTLLSSDTKVPGSGPCGAERLHAVGAPGS--ARPRAFSHSGVHSLDGGEVDSQALQELTQMVSGPASYS

H. sapiens CIC-L    GVSSELLSVAAQTLLSSDTKAPGSSSCGAERLHTVGGPGS--ARPRAFSHSGVHSLDGGEVDSQALQELTQMVSGPASYS

X. tropicalis CIC-L GTAAEVP--PSGLLLGTESKLGASTSTLGPGLPPSSSSSTQISRPRAFSHSAVQSIEHRE-DSQALQELKQMCSNRSPYP

Consensus           G   E       LL    K  S        L           RPRAFSHS V S   E DSQALQEL QM S     Y

                           1370      1380      1390      1400      1410      1420      1430      1440
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   GPKPSPQYGAPGSFAAPG--EGGTLATSGRPPLLPSRASRSQRAASEDMTSDEERMVICEEEGDDDVIADDSFGTTDIDL

H. sapiens CIC-L    GPKPSTQYGAPGPFAAPG--EGGALAATGRPPLLPTRASRSQRAASEDMTSDEERMVICEEEGDDDVIADDGFGTTDIDL

X. tropicalis CIC-L VQKAAFGGPDSSSFPVPAHSEAPTCPASRACSSLPG-SYHPQRTASEDMTSDEERMVICEEDGDDDVIADDGFGCPDIDL

Consensus            K       F  P  E          LP        QR ASEDMTSDEERMVICEE GDDDVIADD FG  DIDL
```

```
                        1450      1460      1470      1480      1490      1500      1510      1520
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   KCKERVTDSESGDSSGEDPEGNKGFGRNVFSPVIRSSFTHCRPTLDPEPPGPPDPPAAFSKGYGPTPSS-SSSPASTSVS

H. sapiens CIC-L    KCKERVTDSESGDSSGEDPEGNKGFGRNVFSPVIRSSFTHCRPPLDPEPPGPPDPPVAFGKGYGSAPSSSASSPASSSAS

X. tropicalis CIC-L KCKERVTDSDSDGRSGDEVEG-KVQPQRVFSPVIRASSLSGTFPVSNEDLKS-ERLQNMAQSPKPSDHPQPSYPRGYPPL

Consensus           KCKERVTDS S  SG  EG K    VFSPVIR S         E                        S P


                        1530      1540      1550      1560      1570      1580      1590      1600
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   VSTSFSLGSGTFKTQESGQGSTAVPLRPPPPGAGGPATPSKATRFPPTDSATFRRKRPESVGSLEAPGPSVIAAPPSGGG

H. sapiens CIC-L    AATSFSLGSGTFKAQESGQGSTAGPLRPPPPGAGGPATPSKATRFLPMDPATFRRKRPESVGGLEPPGPSVIAAPPSGGG

X. tropicalis CIC-L PPAKPPPRP-TTQHPTKRNPEVRFTMVDPSSSYKRRSVDGGQKAVGEGAAPTGQYSTQSVISSGHSQPLVLPSLPSNHP

Consensus                     T           P              A        SV      P V    PS


                        1610      1620      1630      1640      1650      1660      1670      1680
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   NLLQTLVLPPSKEDREG--TRVPSAPAP--PLAYGAP----AAPLCRPAATMVTNVVRPVSSTPVPIASKPFPTSGRAEA

H. sapiens CIC-L    NILQTLVLPPNKEEQEGGGARVPSAPAP--SLAYGAP----AAPLSRPAATMVTNVVRPVSSTPVPIASKPFPTSGRAEA

X. tropicalis CIC-L SSSMYSSAPKSYEGHRGISDSRPPVPTVGTVLVSGPQGQGILPPGLRPASTAVTNVVRPVSSTPVPIASKPLPVS--RSS

Consensus                   P  E   G     P P    L  G         P  RPA T VTNVVRPVSSTPVPIASKP P S
```

```
                              1690      1700      1710      1720      1730      1740      1750      1760
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L        SSNDIAGARTEMGTGSRVPGGSPMGVSLVYS-DKKSAAAATSPAPHLVAGPLLGTVGKAPATVTNLLVGTPGYGAPASPA

H. sapiens CIC-L         SPNDTAGARTEMGTGSRVPGGSPLGVSLVYS-DKK-SAAATSPAPHLVAGPLLGTVGKAPATVTNLLVGTPGYGAPAPPA

X. tropicalis CIC-L      SISDLS--------ESRIPG-SPVGIGVLYAHDRKPMQQLTSPANSPHS-----QSKPAGGIVTNLVVGAPSYGQPAPTT

Consensus                  S  D          SR PG SP G   Y  D K    TSPA              A   VTNL VG P YG PA


                              1770      1780      1790      1800      1810      1820      1830      1840
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L        VQFIAQGAPGSATPAGSGASTGSGPNGPVPLGILQPGALG-----KAGGITQVQYILPTLPQQLQVAPAPAPAPGTKAAA

H. sapiens CIC-L         VQFIAQGAPGGGTTAGSGAGAGSGPNGPVPLGILQPGALG-----KAGGITQVQYILPTLPQQLQVAPAPAPAPGTKAAA

X. tropicalis CIC-L      GNGTVP-APVTALQFITQNSAGN-SNGAVPLRILQPQELLPATASKRGSITQVQYILPTIPQQLPGGTAASIHFTLPPAN

Consensus                       AP        G   NG VPL ILQP  L        K G ITQVQYILPT PQQL   A       A


                              1850      1860      1870      1880      1890      1900      1910      1920
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L        PSGPAPTTSIRFTLP-PGTSTNGKVLAATAPTAGIPILQSVPSAPPPKAQSVSPVQATPSGGSAQLLPGKVLVPLAAPSM

H. sapiens CIC-L         PSGPAPTTSIRFTLP-PGTSTNGKVLAATAPTPGIPILQSVPSAPPPKAQSVSPVQAPPPGGSAQLLPGKVLVPLAAPSM

X. tropicalis CIC-L      AKMIATPQALPIVQPGPAASPSLGIVSSTG------KAQSVSPAPSS-GSSTQLLSSPGILSPNAQVQGKMLVPMATPHV

Consensus                  A          P P  S       T       QSV  AP    S            GK LVP A P
```

```
                         1930      1940      1950      1960      1970      1980      1990      2000
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   SVRGGGAGQPLPLVSSPFSVPVQNGAQQPSKIIQLTPVPVSTP-----SGLVPP------------LSPATMPGPTSQPQ

H. sapiens CIC-L    SVRGGGAGQPLPLVSPPFSVPVQNGAQPPSKIIQLTPVPVSTP-----SGLVPP------------LSPATLPGPTSQPQ

X. tropicalis CIC-L TVRTG-SAAPLPLVTPPF--PVQNGAQPTNKIIQITPMPVVHSPAPAQSALVPPGSSPLHNAIPVTMATAAVMAASSQPQ

Consensus            VR G     PLPLV PF  PVQNGAQ  KIIQ TP PV     S LVPP                 A      SQPQ


                         2010      2020      2030      2040      2050      2060      2070      2080
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   KVLLPSSTRITYVQSAGGHTLPLGTSSACSQTGTVTSYGPTSSVALGFTSLGPSGPAFVQPLLSG-QAPLLAPGQVGVSP

H. sapiens CIC-L    KVLLPSSTRITYVQSAGGHALPLGTSPASSQAGTVTSYGPTSSVALGFTSLGPSGPAFVQPLLSAGQAPLLAPGQVGVSP

X. tropicalis CIC-L KVVLPSSTRITYVPSAGVPTVPLVTSSSHSQTAPSGSAPCVQSMALGFTAIGPNGQTIVQPLITG-QSQLLAQGQVAVSP

Consensus           KV LPSSTRITYV SAG    PL TS  SQ    S    S ALGFT  GP G  VQPL   Q  LLA GQV VSP


                         2090      2100      2110      2120      2130      2140      2150      2160
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L   VPSPQLPPACTASGGPVITAFYPGSPAPTS-APLGPPSQAPPSLVYTVATSTTPP--AATILPKGPPASATATPAPT---

H. sapiens CIC-L    VPSPQLPPACAAPGGPVITAFYSGSPAPTSSAPLAQPSQAPPSLVYTVATSTTPP--AATILPKGPPAPATATPAPT---

X. tropicalis CIC-L ASTPPVPQTCSG---QVLTAIYPSVGSNTLTVPAPP---AAQSVVYTVASTNTISTVPATILPKAITSPQTISSPPAPCI

Consensus             P  P  C     V TA Y    T  P    A  S VYTVA  T    ATILPK      T    P
```

```
                    2170      2180      2190      2200      2210      2220      2230      2240
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L    ----SPFPSATG-SMTYSLVAPKAQRPSPKAPQKVKAAIASIPVGSFESG-------TTGRPGSTP--RQSSDSGVAREP
H. sapiens CIC-L     ----SPFPSATAGSMTYSLVAPKAQRPSPKAPQKVKAAIASIPVGSFEAG-------ASGRPGPAP--RQPLEPGPVREP
X. tropicalis CIC-L  AGSSSQTTPATAASVTFSLVTSKSPRNLPKPPQKVKATIASIPVGSFEGTSPSPCVRNTARPSPLPPCEETPEPRYIGES
Consensus                S    AT S T SLV  K  R  PK PQKVKA IASIPVGSFE           RP    P          E


                    2250      2260      2270      2280      2290      2300      2310      2320
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L    AAP----ESELEGQPTPPAPPPPT------------------ETWPPTARS-SPPPPLPAEERPGTKGPE-TASKFPSS
H. sapiens CIC-L     TAP----ESELEGQPTPPAPPPLP------------------ETWTPTARS-SPPLPPPAEERTSAKGPETMASKFPSS
X. tropicalis CIC-L  RMSSIGREPSTESHSVSPFPPAHAREAAQAKQRLPIAEDRASREIWVKNMTDDSSASPLPDDRREPPHSPS--IKKEPGC
Consensus              E   E      P PP                 E W         S   P P  R      P      K P


                    2330      2340      2350      2360      2370      2380      2390      2400
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L    SSDWRVPGLGLESRGEPPTPPSPAPATGPSGSSSGSSEGSSGRAAGDTPERKEVTSSGKKMKVRPPPLKKTFDSVD-KVL
H. sapiens CIC-L     SSDWRVPGQGLENRGEPPTPPSPAPA--PAVAPGGSSESSSGRAAGDTPERKEAAGTGKKVKVRPPPLKKTFDSVDNRVL
X. tropicalis CIC-L  SSEFKVTETRLDSNPPTGLPAPSSPVTLPTTQSG--EAAKAGTSLSEAPERK--DGPVKVKVRPPPLKKTFDSVDNRVL
Consensus            SS  V  L          P    P  P             G      PERK    KK KVRPPPLKKTFDSVD  VL
```

	2410	2420	2430	2440	2450	2460	2470	2480

....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L SEVDFEERFAELPEFRPEEVLPSPTLQSLATSPRAILGSYRKKRKNSTDLDSAPEDPTSPKRKMRRRSSCSSEPNTPKSA

H. sapiens CIC-L   SEVDFEERFAELPEFRPEEVLPSPTLQSLATSPRAILGSYRKKRKNSTDLDSAPEDPTSPKRKMRRRSSCSSEPNTPKSA

X. tropicalis CIC-L SEVDFEERFAELPEFKPEEVLPSPTLQSLATSPRAILGSYRKKRKNSTDLDSSTEDPVSPKRKMRRRSSCSSEPNTPKSA

Consensus   SEVDFEERFAELPEF PEEVLPSPTLQSLATSPRAILGSYRKKRKNSTDLDS EDP SPKRKMRRRSSCSSEPNTPKSA

	2490	2500	2510	2520	2530	2540	2550	2560

....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L KCEGDIFTFDRTGTETEDVLGELEYEKVPYSSLRRTLDQRRALVMQLFQDHGFFPSAQATAAFQARYADIFPSKVCLQLK

H. sapiens CIC-L   KCEGDIFTFDRTGTEAEDVLGELEYDKVPYSSLRRTLDQRRALVMQLFQDHGFFPSAQATAAFQARYADIFPSKVCLQLK

X. tropicalis CIC-L KCEGDIFTFERTGNEAEDLLGEMEYDKAPYSSLRRTLDQRRALVMQLFQEHGFFPSSQSTAAFQSRYSDIFPTKVCLQLK

Consensus   KCEGDIFTF RTG E ED LGE EY K PYSSLRRTLDQRRALVMQLFQ HGFFPS Q TAAFQ RY DIFP KVCLQLK

	2570	2580	2590	2600	2610	2620	2630	2640

....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

M. musculus CIC-L IREVRQKIMQAATPTEQPPGAEAPLPGPPPTGMAATPVPTPSPAGGPDPTSPGSDSGTAQVAPPLPPPPEPGPGQPGWEG

H. sapiens CIC-L   IREVRQKIMQAATPTEQPPGAEAPLPVPPPTGTAAAPAPTPSPAGGPDPTSPSSDSGTAQAAPPLPPPPESGPGQPGWEG

X. tropicalis CIC-L IREVRQKIMQAATPTES--------------LFAEHPSPSTSEAGPSDMQPPQD-----------PALRSSEPAETAWEE

Consensus   IREVRQKIMQAATPTE   A PP S AG D P   P   P WE

```
                                    2650
                         ....|....|....|...
    M. musculus CIC-L    APQPSPPPSGPSTAATGR
    H. sapiens CIC-L     APQPSPPPPGPSTAATGR
    X. tropicalis CIC-L  GQEP------PETSRSR-
    Consensus              P       P T
```

Figure 25, CIC-L amino acid residue alignment of *H. sapiens, M. musculus, X. tropicalis* and consensus sequence generated from this alignment. The complete amino acid sequenced was generated from the cDNA ORF of the CIC-S isoform. Each colour represents a different amino acid.

```
                          10        20        30        40        50        60        70        80
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
H. sapiens CIC-S     MYSAHRPLMPASSAASRGLGMFVWTNVEPRSVAVFPWHSLVPFLAPSQPDPSVQPSEAQQPASHPVASNQSKEPAESAAV
M. musculus  CIC-S   MYSAHRPLIPASGAASRGLGMFVWTNVEPRSVAVFPWHSLVPFLAPSQPDPSVQPSEAQQPASHPVASNQSKEPAESAAV
X. tropicalis CIC-S  MFSSERPAPPCG------LSMFVWTNVEPRSVAVFPWHSLVPFLAPSQPDSSVQPSEGQQPVNHPGASNQSKEPPESASV
Consensus            M S  RP  P      L MFVWTNVEPRSVAVFPWHSLVPFLAPSQPD SVQPSE QQP  HP ASNQSKEP ESA V

                          90       100       110       120       130       140       150       160
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
H. sapiens CIC-S     AHERPPGGTGSADPERPPGATCPESPGPGPPHPLGVVESGKGPPPTTEEEASGPPGEPRLDSETESDHDDAFLSIMSPEI
M. musculus  CIC-S   AHEQPPGGTGGADPGRPPGAVCPESPGPGPPLTLGGVDPGKSLPPTTEEEAPGPPGEPRLDSETESDHDDAFLSIMSPEI
X. tropicalis CIC-S  AHDAMP--VASIEDER---CAVPRTDNVSQPP----TEEPAKQLQHQESMTQGTTGDQRGDSETESDHDDAFFPIGPPDL
Consensus            AH  P       R     P     P              E    G   G   R DSETESDHDDAF  I  P
```

```
                        170       180       190       200       210       220       230       240
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S    QLPLPPGKRRTQSLSALPKERDSSSEKDGRSPNK-REKDHIRRPMNAFMIFSKRHRALVHQRHPNQDNRTVSKILGEWWY

M. musculus  CIC-S  QLPLPPGKRRTQSLSALPKERDSSSEKDGRSPNK-REKDHIRRPMNAFMIFSKRHRALVHQRHPNQDNRTVSKILGEWWY

X. tropicalis CIC-S QLPMHGGKRRTQSLSALPKERDSSSEKDGRSPNQREKDHIRRPMNAFMIFSKRHRALVHQRHPNQDNRTVSKILGEWWY

Consensus           QLP  GKRRTQSLSALPKERDSSSEKDGRSPN  REKDHIRRPMNAFMIFSKRHRALVHQRHPNQDNRTVSKILGEWWY


                        250       260       270       280       290       300       310       320
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S    ALGPKEKQKYHDLAFQVKEAHFKAHPDWKWCNKDRKSSSEAKPTSLGLAGGHKETRERSMSETGTAAAPGVSSELLSVA

M. musculus  CIC-S  ALGPKEKQKYHDLAFQVKEAHFKAHPDWKWCNKDRKKSSSEAKPASLGLAGGHKETRERSMSETGTAAAPGVSSELLSVA

X. tropicalis CIC-S ALGPKEKQKYHDLAFQVKEAHFKAHPDWKWCNKDRKKSSSDVKQVMPGPWGVPKEMRERSMSETGTMAAAGTAAEVP--P

Consensus           ALGPKEKQKYHDLAFQVKEAHFKAHPDWKWCNKDRKSSS K   G  G   KE RERSMSETGT AA G   E


                        330       340       350       360       370       380       390       400
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S    AQTLLSSDTKAPGSSSCGAERLHTVGGPGS--ARPRAFSHSGVHSLDGGEVDSQALQELTQMVSGPASYSGPKPSTQYGA

M. musculus  CIC-S  AQTLLSSDTKVPGSGPCGAERLHAVGAPGS--ARPRAFSHSGVHSLDGGEVDSQALQELTQMVSGPASYSGPKPSPQYGA

X. tropicalis CIC-S SGLLLGTESKLGASTSTLGPGLPPSSSSSTQISRPRAFSHSAVQSIEHRE-DSQALQELKQMCSNRSPYPVQKAAFGGPD

Consensus              LL  K  S     L           RPRAFSHS V S  E DSQALQEL QM S   Y  K
```

```
                              410       420       430       440       450       460       470       480
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S        PGPFAAPG--EGGALAATGRPPLLPTRASRSQRAASEDMTSDEERMVICEEEGDDDVIADDGFGTTDIDLKCKERVTDSE

M. musculus  CIC-S      PGSFAAPG--EGGTLATSGRPPLLPSRASRSQRAASEDMTSDEERMVICEEEGDDDVIADDSFGTTDIDLKCKERVTDSE

X. tropicalis CIC-S     SSSFPVPAHSEAPTCPASRACSSLPG-SYHPQRTASEDMTSDEERMVICEEDGDDDVIADDGFGCPDIDLKCKERVTDSD

Consensus                   F  P   E         LP      QR ASEDMTSDEERMVICEE GDDDVIADD FG  DIDLKCKERVTDS


                              490       500       510       520       530       540       550       560
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S        SGDSSGEDPEGNKGFGRKVFSPVIRSS-----FTHCRPPLDPE-------PPGPPDPP-VAFGKGYGSAPSSSASSPASS

M. musculus  CIC-S      SGDSSGEDPEGNKGFGRKVFSPVIRSS-----FTHCRPTLDPE-------PPGPPDPP-AAFSKGYGPTPSS-SSSPAST

X. tropicalis CIC-S     SDGRSGDEVEG-KVQPQRVFSPVIRASSLSGTFPVSNEDLKSERLQNMAQSPKPSDHPQPSYPRGYPPLPPAKPPPRPTT

Consensus               S   SG   EG K      VFSPVIR S     F       L  E        P P D P      GY   P


                              570       580       590       600       610       620       630       640
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S        SASAATSFSLGSGTFKAQESGQGSTAGPLRPPPPGAGGPATPSKATRFLPMDPATFRRKRPESVGGLEPPGPSVIAAPPS

M. musculus  CIC-S      SVSVSTSFSLGSGTFKTQESGQGSTAVPLRPPPPGAGGPATPSKATRFPPTDSATFRRKRPESVGSLEAPGPSVIAAPPS

X. tropicalis CIC-S     QHPTKRNPEVRFTMVDPSSSYKRKRSVDGGQKAVGEGAAPTGQYSTQSVISSGHSQPLVLP-SLPSNHPSSSMYSSAPKS

Consensus                               S         G G  T    T           P S        AP S
```

```
                              650       660       670       680       690       700       710       720
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S      GGGNILQTLVLPPNKEEQEGGGARVPSAPAPSLAYGAPAAPLSRPAATMVTNVVRPVSSTPVPIASKPFPTSGRAEASPN

M. musculus  CIC-S    GGGNLLQTLVLPPSKEDRE--GTRVPSAPAPPLAYGAPAAPLCRPAATMVTNVVRPVSSTPVPIASKPFPTSGRAEASSN

X. tropicalis CIC-S   YEGHRGISDSRPPVP--------TVGTVLVSGPQGQGILPPGLRPASTAVTNVVRPVSSTPVPIASKPLPVS--RSSSIS

Consensus                G        PP           V            P  RPA T VTNVVRPVSSTPVPIASKP P S      S


                              730       740       750       760       770       780       790       800
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S      DTAGARTEMGTGSRVPGGSPLGVSLVYS-DKK-SAAATSPAPHLVAGPLLGTVGKAPATVTNLLVGTPGYGAPAPPAVQF

M. musculus  CIC-S    DIAGARTEMGTGSRVPGGSPMGVSLVYS-DKKSAAAATSPAPHLVAGPLLGTVGKAPATVTNLLVGTPGYGAPASPAVQF

X. tropicalis CIC-S   DLS--------ESRIPG-SPVGIGVLYAHDRKPMQQLTSPANSPHS-----QSKPAGGIVTNLVVGAPSYGQPAPTTGNG

Consensus             D            SR PG SP G    Y D K    TSPA            A    VTNL VG P YG PA


                              810       820       830       840       850       860       870       880
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S      IAQGAPGGGTTAGSGAGAGSGPNGPVPLGILQPGALG-----KAGGITQVQYILPTLPQQLQVAPAPAPAPGTKAAAPSG

M. musculus  CIC-S    IAQGAPGSATPAGSGASTGSGPNGPVPLGILQPGALG-----KAGGITQVQYILPTLPQQLQVAPAPAPAPGTKAAAPSG

X. tropicalis CIC-S   TVP-APVTALQFITQNSAGN-SNGAVPLRILQPQELLPATASKRGSITQVQYILPTIPQQLPGGTAASIHFTLPPANAKM

Consensus                  AP          G    NG VPL ILQP  L        K G ITQVQYILPT PQQL    A           A
```

```
                              890       900       910       920       930       940       950       960
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S         PAPTTSIRFTLP-PGTSTNGKVLAATAPTPGIPILQSVPSAPPPKAQSVSPVQAPPPGGSAQLLPGKVLVPLAAPSMSVR

M. musculus  CIC-S       PAPTTSIRFTLP-PGTSTNGKVLAATAPTAGIPILQSVPSAPPPKAQSVSPVQATPSGGSAQLLPGKVLVPLAAPSMSVR

X. tropicalis CIC-S      IATPQALPIVQPGPAASPSLGIVSSTG------KAQSVSPAPSS-GSSTQLLSSPGILSPNAQVQGKMLVPMATPHVTVR

Consensus                   A         P P S       T          QSV  AP      S              GK LVP A P   VR


                              970       980       990       1000      1010      1020      1030      1040
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S         GGGAGQPLPLVSPPFSVPVQNGAQPPSKIIQLTPVPVSTP-----SGLVPP------------LSPATLPGPTSQPQKVL

M. musculus  CIC-S       GGGAGQPLPLVSSPFSVPVQNGAQQPSKIIQLTPVPVSTP-----SGLVPP------------LSPATMPGPTSQPQKVL

X. tropicalis CIC-S      TG-SAAPLPLVTPPF--PVQNGAQPTNKIIQITPMPVVHSPAPAQSALVPPGSSPLHNAIPVTMATAAVMAASSQPQKVV

Consensus                   G    PLPLV  PF  PVQNGAQ    KIIQ TP PV       S LVPP                A      SQPQKV


                              1050      1060      1070      1080      1090      1100      1110      1120
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S         LPSSTRITYVQSAGGHALPLGTSPASSQAGTVTSYGPTSSVALGFTSLGPSGPAFVQPLLSAGQAPLLAPGQVGVSPVPS

M. musculus  CIC-S       LPSSTRITYVQSAGGHTLPLGTSSACSQTGTVTSYGPTSSVALGFTSLGPSGPAFVQPLLSG-QAPLLAPGQVGVSPVPS

X. tropicalis CIC-S      LPSSTRITYVPSAGVPTVPLVTSSSHSQTAPSGSAPCVQSMALGFTAIGPNGQTIVQPLITG-QSQLLAQGQVAVSPAST

Consensus                LPSSTRITYV SAG    PL TS  SQ      S       S ALGFT  GP G   VQPL      Q  LLA GQV VSP
```

```
                              1130      1140      1150      1160      1170      1180      1190      1200
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S     PQLPPACAAPGGPVITAFYSGSPAPTSSAPLAQPSQAPPSLVYTVATSTTPP--AATILPKGPPAPATATPAPT------

M. musculus  CIC-S   PQLPPACTASGGPVITAFYPGSPAPTS-APLGPPSQAPPSLVYTVATSTTPP--AATILPKGPPASATATPAPT------

X. tropicalis CIC-S  PPVPQTCSG---QVLTAIYPSVGSNTLTVPAPP---AAQSVVYTVASTNTISTVPATILPKAITSPQTISSPPAPCIAGS

Consensus               P  P  C     V TA Y    T  P     A  S VYTVA  T    ATILPK     T   P


                              1210      1220      1230      1240      1250      1260      1270      1280
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S     -SPFPSATAGSMTYSLVAPKAQRPSPKAPQKVKAAIASIPVGSFEAG-------ASGRPGPAP--RQPLEPGPVREPTAP

M. musculus  CIC-S   -SPFPSATG-SMTYSLVAPKAQRPSPKAPQKVKAAIASIPVGSFESG-------TTGRPGSTP--RQSSDSGVAREPAAP

X. tropicalis CIC-S  SSQTTPATAASVTFSLVTSKSPRNLPKPPQKVKATIASIPVGSFEGTSPSPCVRNTARPSPLPPCEETPEPRYIGESRMS

Consensus              S    AT  S T SLV   K  R  PK PQKVKA IASIPVGSFE          RP  P            E


                              1290      1300      1310      1320      1330      1340      1350      1360
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S     ----ESELEGQPTPPAPPPLP------------------ETWTPTARS-SPPLPPPAEERTSAKGPETMASKFPSSSSD

M. musculus  CIC-S   ----ESELEGQPTPPAPPPPT------------------ETWPPTARS-SPPPPLPAEERPGTKGPE-TASKFPSSSSD

X. tropicalis CIC-S  SIGREPSTESHSVSPFPPAHAREAAQAKQRLPIAEDRASREIWVKNMTDDSSASPLPDDRREPPHSPS--IKKEPGCSSE

Consensus               E   E   P PP               E W     S   P P  R     P   K P SS
```

```
                        1370      1380      1390      1400      1410      1420      1430      1440
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S    WRVPGQGLENRGEPPTPPSPAPA--PAVAPGGSSESSSGRAAGDTPERKEAAGTGKKVKVRPPPLKKTFDSVDNRVLSEV

M. musculus  CIC-S  WRVPGLGLESRGEPPTPPSPAPATGPSGSSSGSSEGSSGRAAGDTPERKEVTSSGKKMKVRPPPLKKTFDSVDNRVLSEV

X. tropicalis CIC-S FKVTETRLDSNPPTGLPAPSSPVTLPTTQSG--EAAKAGTSLSEAPERK--DGPVKKVKVRPPPLKKTFDSVDNRVLSEV

Consensus              V   L      P    P   P          G      PERK    KK KVRPPPLKKTFDSVDNRVLSEV


                        1450      1460      1470      1480      1490      1500      1510      1520
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S    DFEERFAELPEFRPEEVLPSPTLQSLATSPRAILGSYRKKRKNSTDLDSAPEDPTSPKRKMRRSSCSSEPNTPKSAKCE

M. musculus  CIC-S  DFEERFAELPEFRPEEVLPSPTLQSLATSPRAILGSYRKKRKNSTDLDSAPEDPTSPKRKMRRSSCSSEPNTPKSAKCE

X. tropicalis CIC-S DFEERFAELPEFKPEEVLPSPTLQSLATSPRAILGSYRKKRKNSTDLDSSTEDPVSPKRKMRRSSCSSEPNTPKSAKCE

Consensus           DFEERFAELPEF PEEVLPSPTLQSLATSPRAILGSYRKKRKNSTDLDS  EDP SPKRKMRRSSCSSEPNTPKSAKCE


                        1530      1540      1550      1560      1570      1580      1590      1600
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

H. sapiens CIC-S    GDIFTFDRTGTEAEDVLGELEYDKVPYSSLRRTLDQRRALVMQLFQDHGFFPSAQATAAFQARYADIFPSKVCLQLKIRE

M. musculus  CIC-S  GDIFTFDRTGTETEDVLGELEYEKVPYSSLRRTLDQRRALVMQLFQDHGFFPSAQATAAFQARYADIFPSKVCLQLKIRE

X. tropicalis CIC-S GDIFTFERTGNEAEDLLGEMEYDKAPYSSLRRTLDQRRALVMQLFQEHGFFPSSQSTAAFQSRYSDIFPTKVCLQLKIRE

Consensus           GDIFTF RTG E ED LGE EY K PYSSLRRTLDQRRALVMQLFQ HGFFPS Q TAAFQ RY DIFP KVCLQLKIRE
```

```
                              1610      1620      1630      1640      1650      1660      1670      1680
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

    H. sapiens CIC-S     VRQKIMQAATPTEQPPGAEAPLPVPPPTGTAAAPAPTPSPAGGPDPTSPSSDSGTAQAAPPLPPPPESGPGQPGWEGAPQ

    M. musculus  CIC-S   VRQKIMQAATPTEQPPGAEAPLPGPPPTGMAATPVPTPSPAGGPDPTSPGSDSGTAQVAPPLPPPPEPGPGQPGWEGAPQ

    X. tropicalis CIC-S  VRQKIMQAATPTES--------------LFAEHPSPSTSEAGPSDMQPPQD-----------PALRSSEPAETAWEEGQE

    Consensus            VRQKIMQAATPTE                A  P P  S AG  D   P          P       P    WE GA


                              1690
                         ....|....|....|

    H. sapiens CIC-S     PSPPPPGPSTAATGR

    M. musculus  CIC-S   PSPPPSGPSTAATGR

    X. tropicalis CIC-S  P------PETSRSR-

    Consensus            P       P T
```

Figure 26, CIC-S amino acid residue alignment of *H. sapiens*, *M. musculus*, *X. tropicalis* and consensus sequence generated from this alignment. The complete amino acid sequenced was generated from the cDNA ORF of the CIC-S isoform. Each colour represents a different amino acid.

	Sequence identity to *X. tropicalis* proteins	
	CIC-L	CIC-S
H. sapiens	51.3%	50.9%
M. musculus	50.6%	51%
D. rerio	38.2%	42.8%
D. melanogaster	25.9%	25.3%

Table 3, the amino acid sequence identity of CIC-L and CIC-S in *X. tropicalis* in comparison to *H. sapiens* (CIC-L NM_001304815.1 & CIC-S NM_015125.4)*, M. musculus* (CIC-L NM_001302811.1 & CIC-S NM_027882.4)*, D.* rerio (CIC-L XM_003200533.5 & CIC-S XM_005173512.4) and *D. melanogaster* (CIC-L NM_001275826.1 & CIC-S NM_080253.4) generated with the William Pearson's lalign program (https://embnet.vital-it.ch/software/LALIGN_form.html).

3.2.8 The HMG-box, ATXN-1 and C1 amino acid sequences are highly conserved

The highly conserved amino acid of important functional domains of H. *sapiens*, *M. musculus, D. melanogaster* and *Danio rerio* domains were aligned to the *X. tropicalis* amino acid sequence and analysed (fig 28-30).

With exception to the ATXN1 domain (fig 28-30), all the *X. tropicalis* domains (14-3-3, HMG-Box, C2, NLS & C1) amino acid sequences had a 100% identity to the *H. sapiens* and *M. musculus* homologs. The *X. tropicalis* ATXN1 amino acid sequence had only 1 substitution of the 26-amino acid, with a 96.2% similarity to the *H. sapiens* and *M. musculus* amino acid sequence (fig 21). Both the NLS (**KRKMRR**) and 14-3-3 (**KRRTQS**) amino acids are contained within the *X. tropicalis* locus (Dissanayake et al., 2011).
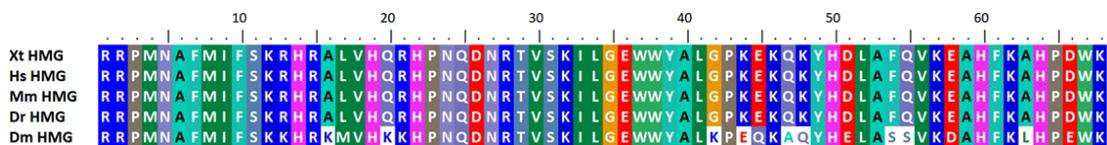
Figure 27, the alignment of HMG-Box amino acid sequences of *Xenopus tropicalis (Xt), Homo sapiens (Hs), Mus musculus (Mm), Danio rerio (Dr)* and *Drosophila melanogaster (Dm)* species which highlights the high level of conservation. The *H. sapiens, M. musculus and X. tropicalis* amino sequences are identical. Although many amino acids have changed in *D. melanogaster*, the amino acids are conserved substitutions.
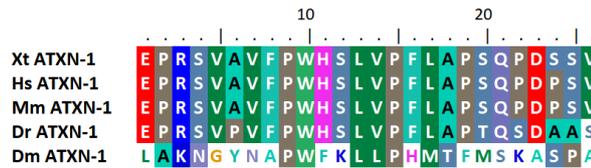
Figure 28, the alignment of the ATXN-1 domain amino acid sequences of *X. tropicalis (Xt), H. sapiens (Hs), M. musculus (Mm), D. rerio (Dr) and D. melanogaster (Dm)* species which highlights the high level of conservation. *The H. sapiens, M. musculus and X. tropicalis* amino sequences are identical.
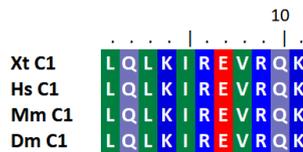
Figure 29, the alignment of the C1 domain amino acid sequences of *X. tropicalis (Xt), H. sapiens (Hs), M. musculus (Mm) and D. melanogaster (Dm) species which* highlights the high level of conservation. The *H. sapiens, M. musculus, X. tropicalis and D. melanogaster* amino sequences are identical.

3.3 Discussion

The alignment of the highly conserved *CIC* homologs of *Homo sapiens* and *M. musculus* to the *X. tropicalis* has shown that *CIC* (fig 25-29), like other vertebrate orthologues, is highly conserved in *X. tropicalis. X. tropicalis* contains all the domains found in other species (fig 27-29). High conservation of the domains provided additional information to target likely areas of the *X. tropicalis* genome to search for exons. The *H. sapiens CIC* gene is located at chromosome 19 q13.2, (fig 30) (Lee et al., 2002; Yang et al., 2017; Yip et al., 2012).

Figure 30, *H. sapiens CIC* gene is in chromosome 19 q13.2 (Ensembl).

The *M. musculus CIC* gene homolog is located at chromosome 7, A3 (fig 31) (Lam et al., 2006; Lee et al., 2002; Lu et al., 2017b; Simon-Carrasco et al., 2017).

Figure 31*, M. musculus CIC* gene homolog is located in chromosome 7, A3 (Ensembl).

Although, the *X. tropicalis* sequencing data highlighted that there was an extremely large intron between exons 1 and 3, which is 45,924 bp in length (fig 12), this intron is much larger than that of *H. sapiens* and *M. musculus* first intron. This intron increases the gene span of the *X. tropicalis CIC* homolog (62, 086 bp) by almost 3-fold compared to the *H. sapiens* (23,162 bp) and *M. musculus* (23,421 bp) homologs.

The published protein sequence of *H. sapiens* CIC-L (NM_001304815.1) and *M. musculus* CIC-L (NM_001302811.1) were aligned and showed a sequence identify of 91% (2300/2520 amino acids). The amino acid sequence of *M. musculus* CIC-S (NM_027882.4) and *H. sapiens* CIC-S (NM_015125.4) had a sequence identify of 91% (1468/1611 amino acids) (for alignment sequence see Accompanying Material CD). Very little is known of differences in function in vertebrates between the isoforms of CIC. The isolation and cloning of the isoforms would give some insight into the function.

ATXN-1 (Lam et al., 2006; Lee et al., 2011), 14-3-3 , HMG-box (Fores et al., 2017; Lee et al., 2002), C2, NLS (Dissanayake et al., 2011) and C1 (Astigarraga et al., 2007a) domain regions of CIC were all shown to be highly conserved between *H. sapiens*, *M. musculus* and *X. tropicalis* (fig 25-29). The high sequence identity of the functional domains and motifs only highlights their importance within the CIC protein. For example, the amino acid sequence for the HMG-box and C1 motif (fig 27 & 29) are almost identical across the species which would be expected given CICs role as a transcriptional repressor. When looking at ATXN-1 domain amino acid sequence alignment in the vertebrate species (fig 28), *H. sapiens, M. musculus*, *D. rerio* and *X. tropicalis* all appear to have a higher sequence identify than in the non-vertebrate, *D. melanogaster*. This could be explained by research which has revealed that in the *D. melanogaster* CIC protein homolog, CIC is reliant on Groucho as a binding partner for transcriptional repression, rather than ATXN-1 (Fores et al., 2015; Jimenez et al., 2000). Alternately, ATXN-1 plays a much more important role in vertebrates acting as a binding partner for transcriptional repression (Lam et al., 2006) which could suggest its importance in the vertebrate species and a reason for the higher sequence identify in vertebrate, rather than non-vertebrates.

3.3.1 Characterisation of the *X. tropicalis CIC* gene
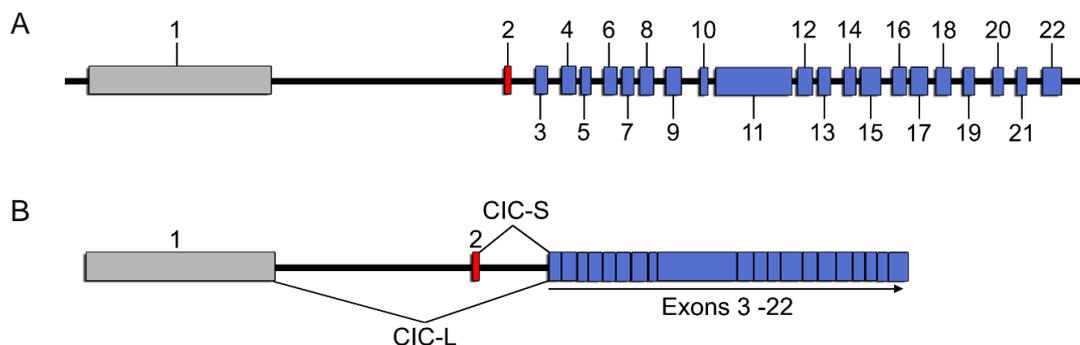
Figure 32, (A) the exon structure of *CIC* in *X. tropicalis*. Exons are in proportion to each other. (B) A diagram of the alternative 5' splicing of the two prominent isoforms of *CIC*, CIC-L and CIC-S. (Grey) CIC-L specific exon, (red) CIC-S specific exon and (blue) the constitutive exons of CIC.

At the start of this project, very little was known about the differences between the CIC-L and CIC-S isoform structures other than that CIC-L contained the large CIC-L specific exon, exon 1, whilst CIC-S lacked the large exon 1. Consistent with the finding that there was a CIC-S specific exon in *D. melanogaster* (Fores et al., 2015), a CIC-S specific exon was found in *X. tropicalis* (fig 13-14). The finding of a unique CIC-S specific exon has allowed spatial and temporal expression studies for both isoforms of *CIC.* The *CIC* gene codes for 2 prominent isoforms (CIC-L and CIC-S), the isoforms differ in size and in their N-terminal regions. CIC-L which contains the largest of the exon of *CIC* (2,812 bp), the N-terminal exon 1 (fig 34b). This first exon is unique to CIC-L and is not contained in the other prominent isoform, CIC-S (fig 32b). CIC-S instead contains exon 2 (49 bp) which is unique to CIC-S. All other exons, exons 3-22, are common between both CIC-L and CIC-S in *H. sapiens* and *M. musculus*.

3.3.2 Alternative isoforms of *CIC*

The addition of an intron discovered in the CIC-L cDNA (fig 10-11) could suggest an alternative splice form (Kondrashov and Koonin, 2003) of *CIC*, which could produce a truncated protein due to premature stop codon. The truncated protein would still contain the exons coding for the ATAXN1 domain (exon 2) and N1 terminal domain (exon1) found in CIC-L (fig 10-11) and may still retain function. It is important to note that no other protein sequences could be found in the NCBI protein database with the additional intron insert in other species. This would need further analysis to establish if the intron is inserted due to error or some other additional function of *CIC*.

In addition, when cloning amplicon 3 (fig 16) for the full ORF of *CIC*, cDNA sequence was identified which lacked exon 21 (for gene structure see the Accompanying Material CD). Like the above example, this could be due to splicing error or could lead to another function of *CIC.* Although exon 21 was an area where no functional domain coding region was present in the DNA sequence, an alternative isoform lacking exon 21 could lead to changes in protein function. This will be required to be investigated further (expression analysis of the exon 21 isoform is carried out in chapter 4).

Chapter 4: Temporal and spatial expression analysis of *Xenopus tropicalis* CIC.

4.1 Introduction

FGF expression in *Xenopus* has been firmly established, with well-defined spatial and temporal patterns during development, both in *X. tropicalis* and *X. laevis* (Christen and Slack, 1997b; Isaacs et al., 1994; Isaacs et al., 1995b; Isaacs et al., 1992; Lea et al., 2009). The Isaacs lab is interested in FGF4 and FGF8 due to their potent mesodermal induction properties (Amaya et al., 1993; Christen and Slack, 1997b; Fletcher et al., 2006; Isaacs et al., 1994; Isaacs et al., 1995b; Isaacs et al., 1992). FGF4 first becomes expressed during gastrulation in the marginal zone (Isaacs et al., 1995b).

The marginal zone in *Xenopus* development is the region of the embryo which contains the Spemann-Mangold organizer (Spemann and Mangold, 2001). This is also known as the Spemann organizer which is a region of cells that induces the development of the central nervous system (Spemann and Mangold, 2001). FGF4 is a potent mesoderm inducer (fig 33) and regulator of anterior-posterior specification (Christen and Slack, 1997b; Deimling and Drysdale, 2011; Isaacs et al., 1994). Like FGF4, FGF8 is essential in mesoderm induction and posterior neural tissue induction and becomes expressed during gastrulation in the marginal zone (Christen and Slack, 1997b; Fletcher et al., 2006). FGF8 has two splice forms, FGF8a and FGF8b, whilst FGF8b is a potent mesoderm inducer, FGF8a has less effect on mesoderm induction.
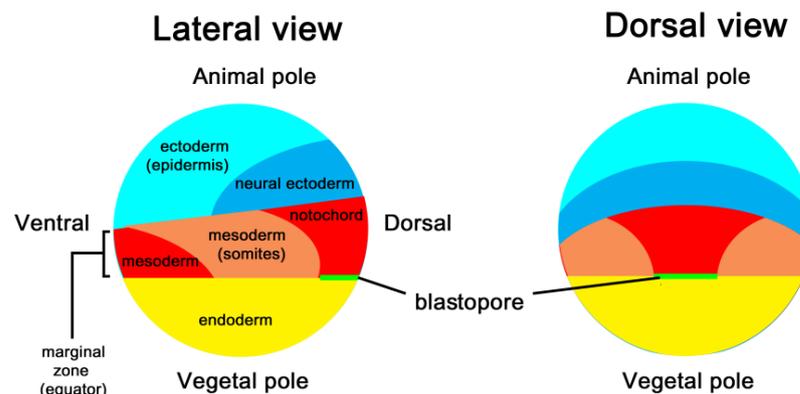
Figure 33, a fate map of a Xenopus gastrula stage embryo (Kumano and Smith, 2002). There are 3 germ layers, ectoderm, mesoderm and endoderm. Ectoderm tissue can become epidermis and nerve tissue. Mesoderm leads

to the formation of muscle, bone, connective tissue, cartilage, notochord, lymphoid tissue, blood, bone marrow and epithelia tissue. Endoderm is the precursor for the linings of the digestive and respiratory system.

The aim of this study is to uncover if there is a relationship between FGF signal transduction and CIC repressional activity during *X. tropicalis* embryonic development. In doing so it will also establish if the FGFR RTK post-transcriptionally regulates CIC in a similar fashion to other RTK receptors, such as Torso (Jimenez et al., 2000) or EGFR, (Goff et al., 2001; Roch et al., 2002) as previously seen in the *D. melanogaster* animal model. Although research in *D. melanogaster* has provided much insight into how CIC is regulated by the Torso and EGFR RTKs, (Goff et al., 2001; Jimenez et al., 2000; Roch et al., 2002), there are still gaps in our knowledge as to when and where CIC is expressed in embryonic development in vertebrates. However, we do know that both receptors utilise the MAPK transduction to post-translationally regulate CIC.

Very little is known about CIC's role in development outside of the invertebrate *D. melanogaster* model, with much of the focus in vertebrates, coming from *H. sapiens* research in the cancer field. In the early *D. melanogaster* pupae, *CIC* was shown to be expressed at the poles of the embryo and is responsible for anterior and posterior patterning (Jimenez et al., 2000). However, studies in *CIC* mutant *M. musculus* in early development show that *CIC* is expressed in the developing lung and is essential for the correct alveolarization (Lee et al., 2011).

To begin to get a greater understanding of CICs role in embryonic development, analysis of the *CIC* isoforms will provide information as to when and where *CIC* is expressed. This data will give an indication to its function in vertebrate developmental context. If FGF does regulate *CIC* expression, it will be expressed at the same temporal and spatial points in development. If FGF regulates CIC post-translationally during gastrulation, it will be required to be expressed in the marginal zone of the embryo. The data from this study will be compared to the well-established expression profiles of FGF. In addition, because no previous attempt at this type of analysis has been performed in vertebrate animal models, it is important to establish if there is an alternative expression pattern between the two prominent isoforms. Identification of distinct expression patterns of CIC isoforms, CIC-L and CIC-S, will give insight into alternative functions of the isoforms.

To address gaps in knowledge regarding *CIC* expression, whole mount *in situ* hybridisation analysis will be utilised to establish spatial expression profiles of the *CIC* isoforms and reverse transcriptase PCR and quantitative PCR analysis will be

employed to establish temporal expression profiles. A detailed analysis of the temporal and spatial expression of the CIC-L and CIC-S isoforms of CIC during early *X. tropicalis* will be completed across eight embryonic developmental stages (8 cell, blastula, early gastrulation, mid gastrulation, late gastrulation, neurula, early tail bud and late tail bud). Sequencing data and characterisation of the *CIC X. tropicalis* gene from this project (chapter 3) has enabled the design of reagents for expression analysis of the prominent isoforms of *CIC*.

To address these questions the following aims of this chapter are to:

- Establish the temporal expression profiles of the *X. tropicalis CIC* isoforms in early development.
- Determine the spatial expression profiles of the *X. tropicalis CIC* isoforms in early development.
- Discover if CIC and FGF share similar temporal and spatial expression during early embryonic development in *X. tropicalis*.
- Identify if the prominent isoforms of CIC have alternative expression profiles.

4.2 Results

4.2.1 Temporal expression of *CIC*

4.2.1.1 Reverse transcriptase-PCR analysis of the prominent isoforms of CIC.

To analyse expression patterns of the *CIC* isoforms, semi-quantitative RT-PCR analysis was performed on cDNA, created from total RNA, collected from sibling embryos from 8-cell (stage 4), through to late tailbud (stage 40) to establish at what point during *X. tropicalis* embryonic development *CIC* becomes expressed. Stages where selected in each of the *Xenopus* developmental landmarks (cleavage, blastula, gastrula, neurula, early and late tailbud) (fig 34). This provided a broad overview of expression and allowed interpretation of CIC function throughout development. For the cleavage phase of development, stage 4 was chosen as this as a mid-point in the cleavage stages and signified embryos which have undergone 3 synchronous rounds of division, forming an 8-cell embryo. Stage 8 chosen because it specifies the period in development when Mid-blastrula transition occurs, an important point in development. All gastrula stages (stages 10-12) were analysed. Emphasis was placed on gastrula stages of development due to it being an important time point in development when mesoderm induction takes place under regulation by FGF (Amaya et al., 1993; Cooke, 1989; Isaacs et al., 1994; Isaacs et al., 1995b; Isaacs et al.,

1992). Stage 18 is the mid-point of the neurula phase of development when the nervous system is formed. Stages 25 was selected as the mid-point of the early tailbud a phase of development when neurulation is completed, and tail formation begins. Stage 35 was chosen when as the mid-point to late tailbud phase, a period in development prior to the start of metamorphization of the tadpole.
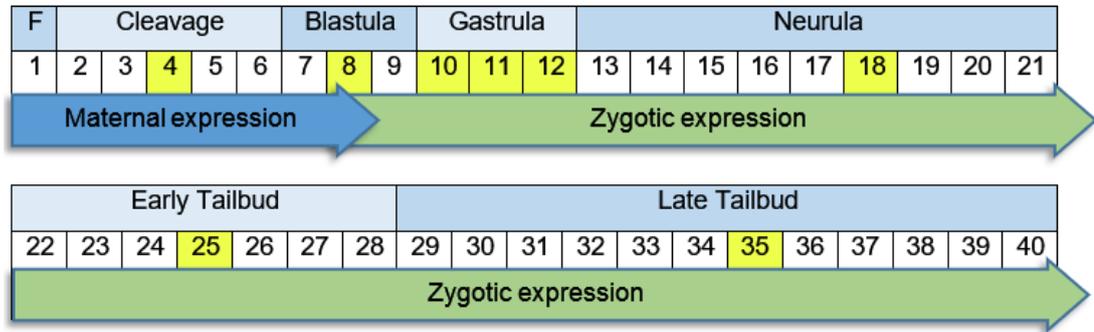
Figure 34, shows the embryonic stages in the developmental landmarks chosen for the reverse transcriptase-PCR and *in situ* hybridisation analysis. One stage was chosen for each of the landmarks except for the gastrula stages which gives rise to the mesodermal tissue.

Initial temporal expression analysis was performed using RT-PCR, the two prominent isoforms (CIC-L and CIC-S) (fig 35a) and an isoform lacking exon 21 (fig 35b). PCR products were generated from cDNA and sampled at 30 cycles. The ribosomal protein L8 and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) genes were used as housekeeping controls (fig 36).
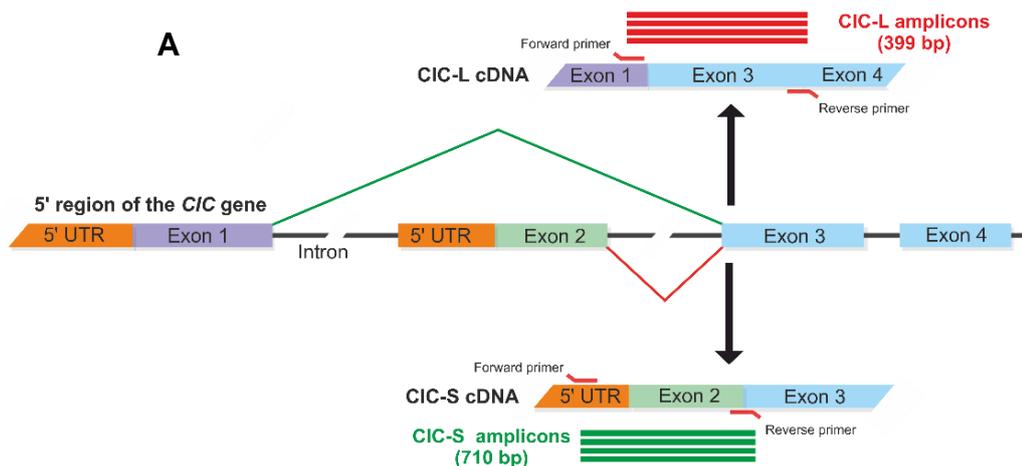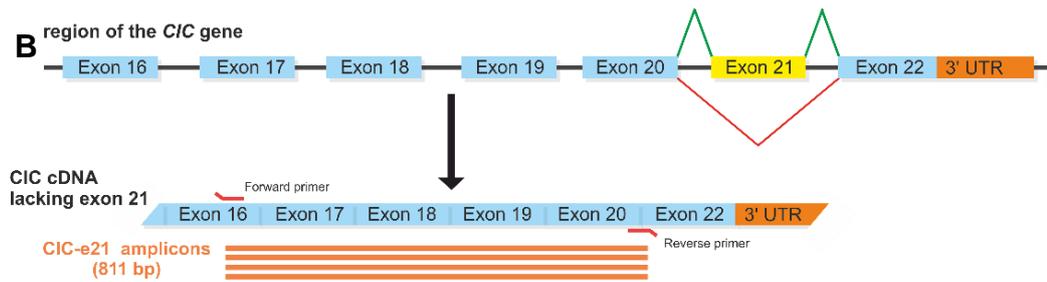
Figure 35, the primer design strategy for amplification of the 5' specific isoforms of (A) CIC-L and CIC-S from cDNA for RT-PCR expression analysis. The CIC-L forward primer was designed to bind to the CIC-L specific exon 1 and reverse primer was designed to bind exon junction of exons 3 and 4. The CIC-S forward primer was designed to bind to the 5'UTR/intron upstream of the CIC-S specific exon 2 and the reverse primer was designed to bind to the exon 2 and 3 junction. (B) The design strategy for amplification of the isoform which lacks exon 21 (CIC-e21). The CIC-e21 forward primer was designed to bind to exon 21 and the reverse primer was designed to bind to the exon junction of exons 20-22, which would only be possible if exon 21 was no present.
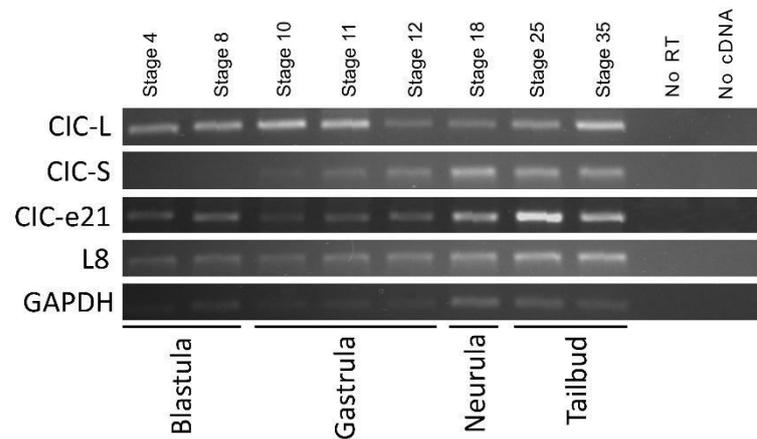
Figure 36, the 1.5% agarose gel displaying the expression of CIC-L, CIC-S and isoform lacking exon 21 (CIC-e21) for RT-PCR analysis. Amplicons were amplified from total mRNA extracted from *X. tropicalis* embryos (n=10). RT-PCR samples were collected from on 8 developmental stages, with particular interest to stages 10-12 (gastrulation). Analysis of the CIC-L and CIC-S expression was repeated 3 times and this figure is representative of those results. Amplicons where of expected size for each amplicon; CIC-L 399bp, CIC-S 710bp, CIC-e21 811bp, L8 436bp and GAPDH 353bp.

RT-PCR analysis reveals expression of *CIC-S* cannot be detected at stage 4 and 8 prior to Mid-blastula transition (MBT) (fig 36) (Newport and Kirschner, 1982). CIC-S first becomes expressed in the early gastrulation (stage 10), suggesting that *CIC-S* is zygotically expressed. *CIC-S* expression increases through mid and late gastrulation (stage 11-12) with peak expression at the neurula phase of development. Expression of *CIC-S* is maintained in early and late tailbud phase of development although expression levels decrease in comparison to the earlier neurula expression levels.

CIC-L expression is detectable from the 8-cell stage (stage 4) embryo, implying the larger isoform of *CIC* is maternally expressed (fig 36). Expression increases in the blastula stage embryo (stage 8), up until early gastrulation, where it peaks. Expression decreases at the later stages of gastrulation to resembling expression levels previously seen in *CIC-S* at this phase in development. *CIC-L* maintains similar levels of expression from late gastrulation through to the neurula phase with increases of expression in early to late tailbud stages peaking for a second time at late tailbud stage.

The isoform of *CIC* lacking exon 21 is expressed prior to MBT suggesting it is maternally expressed and can be seen at the stage 4 of development and increases expression at stage 8 (fig 36). Expression decreases from stage 8 to the point of expression in the stage series at early gastrulation (stage 10). Expression increases from early gastrulation gradually up until peak expression at the early tailbud stage and expression decreases at the late tailbud stage.

4.2.1.2 Temporal qPCR analysis of the common isoforms.

To further determine the accuracy of expression profiles of the prominent isoforms previously analysed using RT-PCR, CIC-L and CIC-S expression analyse was repeated using qPCR, a more quantitative measure of expression. Temporal analysis was performed on cDNA generated from total RNA collected from sibling embryos utilising the same stage series of development previously seen in the RT-PCR experiment with emphasis on gastrulation.
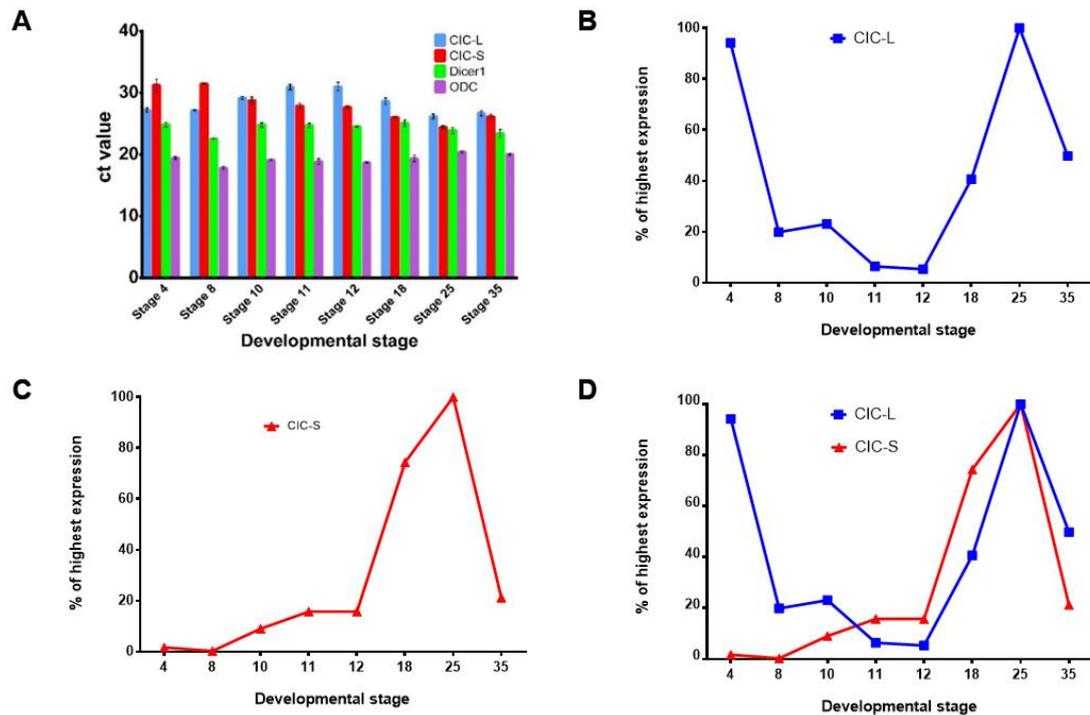
Figure 37, results from the expression qPCR analysis of CIC-L and CIC-S. qPCR samples were analysed across 8 developmental stages. (A) The CT values of CIC-L, CIC-S, Dicer1 and ODC (threshold 0.379 $^{\Delta}$Rn) (StepOne v2.3 software). (B) CIC-L expression normalised to Dicer1 ($2^{-\Delta\Delta CT}$). (C) CIC-L expression normalised to Dicer1 (2-ΔΔCT). (D) Combined data of CIC-L and CIC-S normalised to Dicer1.

CIC-S expression analysis by qPCR is in keeping with previous RT-PCR results (fig 37). CIC-S has very low expression in the stage 4 embryo and is undetectable at stage 8 (fig 37c). CIC-S expression increases after stage 8 of development, when zygotic genes are activated after MBT. Expression of CIC-S increases from stage 8 until mid-gastrulation where expression is maintained until late gastrulation. At the end of gastrulation (stage 12), expression of CIC-S sharply increases through neurula stages (stage 18) to peak expression during early tailbud phases of development and then decreases at the late tailbud stage (stage 35). In keeping with the RT-PCR data, qPCR analysis showed that CIC-L is highly expressed at stages 4 -8 and CIC-L is expressed before MBT, suggesting that it is maternally expressed (fig 37b). CIC-L expression sharply decreases in expression from stage 4 to stage 8. Expression increases slightly at early gastrulation before decreasing as embryonic development moves from early to late gastrulation. Like CIC-S, CIC-L expression sharply increases through neurulation to peak expression during early tailbud phases of development (fig 37d). Both CIC-L and CIC-S show peak expression in the early tailbud stage. After the early tailbud phase of development, in the late tailbud stage (stage 35) expression of CIC-L decreases (fig 37d).

4.3 Spatial expression of *CIC*

4.3.1 Spatial analysis of CIC isoform expression by whole mount *in situ* hybridisation.

To determine the spatial expression patterns of the CIC isoforms, anti-sense RNA digoxigenin-labelled probes were created from linear plasmid template created from DNA or cDNA. The CIC-S template was created from cDNA generated from reverse transcriptase of mRNA from the *CIC-S* isoform. The template for CIC-S was amplified from a 710bp fragment. The forward primer was designed to bind within the 5' UTR/intron1, upstream of exon 2 and the reverse primer was designed to bind at the exon 2-exon 3 junction (fig 38).
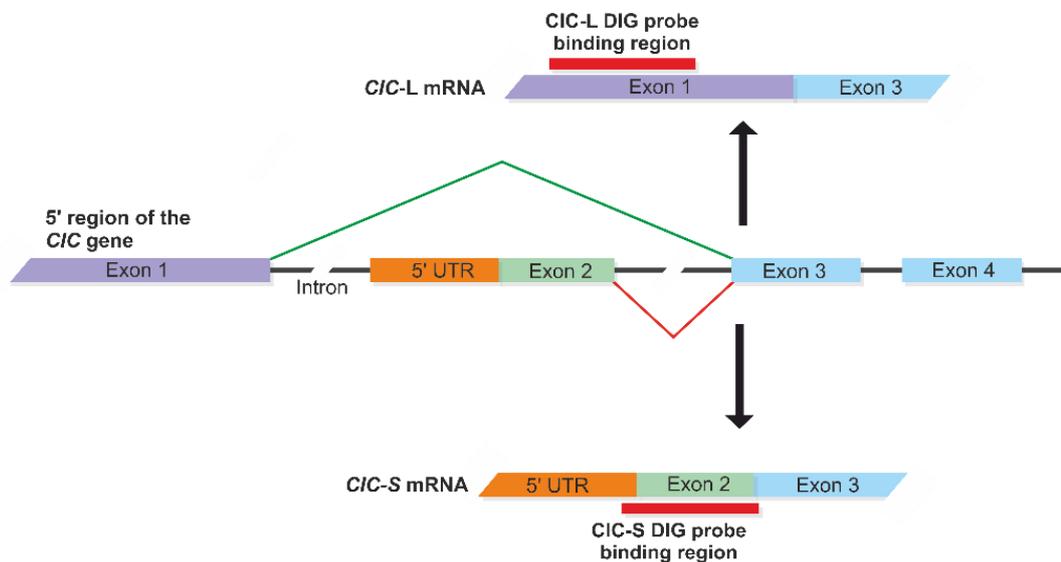
Figure 38 the probe design strategy for *in situ* hybridisation analysis of CIC-L and CIC-S in *X. tropicalis*. The CIC-S probe was designed to bind to the CIC-S specific exon 1. The CIC-S probe was designed to bind in the 5'UTR, exon 2 and the exon 2-3 junctions.

Figure 39, results for CIC-L *in situ* hybridisation. Stages 4-8 (animal view), stage 10-12 (vegetal view) and stages 25-35 (lateral view, anterior to the left). CIC-L has enriched expression at the blastula stages 4 and 8. Expression is found around the blastopore and more heavily near the dorsal blastopore lip (dbl) at the start of gastrulation (stage 10). Expression becomes more widespread at the tailbud stages (25-35).
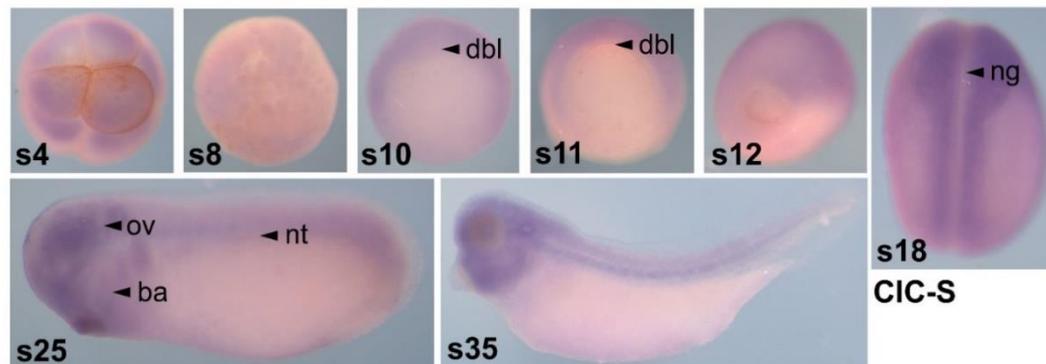
Figure 40, results for CIC-*S in situ* hybridisation. Stages 4-8 (animal view), stage 10-12 (vegetal view) and stages 25-35 (lateral view). CIC-S has little to no expression at stages 4-8. Expression in CIC-S is found around the blastopore and more heavily near the dorsal blastopore lip (dbl) at the start of gastrulation (stage 10). Expression becomes more widespread at the tailbud stages (25-35).

4.3.1.1 CIC-L spatial expression

Whole mount *in situ* hybridisation analysis reveal CIC-L is highly expressed in 8-cell embryos (stage 4), validating maternal expression of the CIC-L isoform (fig 39). CIC-L expression is enriched at the animal hemisphere through the blastula stages (stage 8). At gastrulation, CIC-L expression becomes localised in the equator of the embryo at the marginal zone, with heavier expression at the dorsal marginal region, adjacent to the dorsal blastopore lip. At the neurula stages of development, CIC-L is localised along the neural plate although it has little to no expression within the neural groove, which gives the appearances of two bands of expression alone the neural plate (fig 39). CIC-L has broad expression at the anterior cranial end of the neural plate which will later form the brain. Expression of CIC-L during early and late tailbud stages is widespread at the anterior region and dorsal regions of the embryo. CIC-L is also expressed at the neural tube, notochord, pronephros and somite regions in the dorsal region. CIC-L expression although widespread in these regions can be seen to be more intense at the forebrain, midbrain, hindbrain, otic vesicle, eye and branchial arches (fig 39).

4.3.1.2 CIC-S expression

Spatial expression data from whole *in situ* hybridisation, show CIC-S has very low expression at stages 4 and 8, prior to MBT (fig 40). This validates previous results seen in the temporal expression analysis, highlighting CIC-S is zygotically expressed. Throughout gastrulation, CIC-S is expressed in the marginal zone whilst being enriched at the dorsal region near the dorsal blastopore lip. During neurulation CIC-S shows a similar spatial expression profile to CIC-L. CIC-S expression is contained along the length of the neural plate, with broader expression at the anterior cranial end of the neural plate (fig 40). Like CIC-L, CIC-S appears to also have low expression within the neural groove. At the tailbud stages, CIC-S is expressed in the head region at the forebrain, midbrain, hindbrain, otic vesicle, eye and branchial arches. In the ventral regions of the embryo, CIC-S is expressed at the neural tube and notochord. CIC-S appears to have a striated pattern of expression, suggesting expression within the somites (fig 40).

4.4 **Discussion**

4.4.1 CIC-L and CIC-S have different temporal expression profiles.

4.4.1.1 Expression of CIC-L during the maternal phase of development.

RT-PCR (fig 36) and qPCR (fig 37) analyse of *CIC*s temporal expression has revealed that the prominent isoforms, CIC-L and CIC-S, have different temporal expression profiles in *X. tropicalis*. CIC-L is detected early in development, in the cleavage (stages 4) and blastula staged embryos (stage 8), demonstrating that it is maternally expressed. In addition, temporal analysis was performed on an isoform lacking exon 21, identified during cloning of the CIC-S cDNA (fig 36). Temporal expression analysis of the isoform lacking exon 21 appeared to have similar expression profile early in development as CIC-L, displaying maternal expression. These early alternative maternal/zygotic temporal expression profiles may indicate that the *CIC* isoforms have alternative functions in early development and this will be an important area for further study. Due to limitations of time and the main focus of this project in establishing if CIC operates downstream of FGF signalling the exon 21 containing isoform was considered to be out of the remit of this project.

Whole mount *in situ* hybridisation analysis of the CIC isoforms in *X. tropicalis* validate the RT-PCR and qPCR temporal results, like previous qPCR data, spatial expression of CIC-S is at very low levels and located at the animal pole of the 8-cell embryo (stage 4) (fig 40). Although previous RT-PCR data indicate CIC-S is undetectable at this stage of development, *in situ* data is more comparable to the qPCR data with slight/low concentrations of CIC-S in comparison to the enriched CIC-L expression at the animal pole of the embryo (fig 36-37 & 39-40). In blastula stage embryos (stage 10), CIC-L remains enriched in the animal hemisphere, whilst CIC-S expression in the animal pole appears depleted in comparison (fig 39-40).

Recent spatial and temporal analysis utilising the *D. rerio* model, found that *CIC* is highly expressed maternally (Chen et al., 2014). Unlike other vertebrate models, *D. rerio* has two copies of the *CIC* gene (*cica* and *cicb*), likely due to a genome duplication event (Taylor et al., 2003). In the early maternal phase of development, from the 4-cell to 1K-cell stage, both copies of the gene, *cica* and *cicb*, have high expression levels. Expression of *cica* and *cicb* rapidly decreases in the shield stage and continues to be weakly expressed throughout early development and in

adulthood. Although *cica* and *cicb* are weakly expressed, *in situ* hybridisation data indicates *cicb* is highly enriched in the hindbrain of the 5-6 somite staged embryo.

Interestingly, in the *D. melanogaster* model, CIC-L has a suggested specific role in oogenesis (Goff et al., 2001; Rittenhouse and Berg, 1995). Although no evidence has been found yet to suggest CIC-L is involved in oogenesis in vertebrate animal models it could suggest a reason for its presence and high expression during the maternal phase of development. CIC-L maternal expression occurs at the earliest phase of embryogenesis, a period in development when the embryo is dependent on the genotype of the mother. The early embryo relies on the maternally deposited mRNA and proteins in the oocyte. Expression of the zygotic genome through the maternal phase of development is silenced by several factors (Edgar and Schubiger, 1986; Schier, 2007; Shermoen and O'Farrell, 1991). Some of the examples seen are epigenetic histone modifications, which regulates gene expression in chromatin/heterochromatin (Prioleau et al., 1994) and lack of transcription machinery, due to shortened cell cycles, in the very early stages of development (Shermoen and O'Farrell, 1991).

CIC has an essential role during the maternal phase of development at least in *D. melanogaster*, regulating maternally acting genes (*tll* and *hkb*) downstream of the RTK, *tor* (Jimenez et al., 2000) . Studies in the *D. melanogaster* model identified a novel recessive loss-of-function mutation in the *CIC* gene (*cic[1]*), which influences the anterior-posterior patterning during the maternal phase of development (Jimenez et al., 2000). Embryos with the *cic[1]* mutation lacked large regions of the segmented trunk seen in wild type larva, losing large numbers of the 14 denticle belts. Denticle belts are the external stripes indicating the segments found along the trunk of *D. melanogaster* larva. Remarkably, the phenotype resembled that of embryos produced from females carrying a *tor* dominant gain-of-function mutations, caused constitutive tor RTK signalling throughout the entire embryo (Klingler et al., 1988). Constitutive tor RTK signalling leads to ectopic expression of *tll* and *hkb* blocking repressional activity of CIC. The wildtype CIC protein usually functions to repress *tll* and *hkb* in the body of the embryo (Jimenez et al., 2000). This highlights, at least in the *D. melanogaster* model that CIC has an important function in the maternal phase of development and that CICs repressonal activity is regulated by MAPK transduction, during the maternal phase. These findings also highlight that CIC is required for normal development of the anterior-posterior axis. Further analysis will be required to access if CIC plays as an important role in the *Xenopus* maternal phase of development, as it does in *Drosophila* maternal development.

4.4.1.2 CIC-S becomes expressed during the zygotic phase of development.

As the embryo moves through embryonic development, it transitions from maternal to zygotic expression. The embryo moves from being heavily dependent on the maternal genotype to becoming reliant on the zygotic genome expression. The activation of zygotic transcription is regulated by several factors, one of which is the nucleo-cytoplasmic ratio (Newport and Kirschner, 1982). Newport and Kirschner demonstrated that the egg has a finite concentration of transcriptional repressor which is packaged in the egg during oogenesis. When the *Xenopus* oocyte is fertilised and undergoes cleavage from single cell up until MBT, the cells of the embryo reach a critical ratio of nucleus to cytoplasm (nucleus increases, and cytoplasm decrease with every cell division) and the concentration of maternally deposited transcriptional repressor reaches a critical concentration which can no longer prevent repression of transcription, leading to onset of zygotic transcription. The activation of zygotic transcription coincides with the slowing of the cell cycle, allowing the expression of larger genes (Heyn et al., 2014). It also leads to and increases of cell motility and increases of asynchronous cell division. CIC-L may be one such transcriptional repressor that is involved in the nucleo-cytoplasmic ratio postulated by Newport and Kirschner (Newport and Kirschner, 1982). CIC-L expression is initially high in early cleavage and as the embryo undergoes further cell cleavage, CIC-L concentrations decreases until zygotic expression of the embryo is activated, and zygotic expression of CIC-L occurs. In contrast to the findings of the CIC-L temporal expression profile in RT-PCR analysis, expression of CIC-S is undetectable before gastrulation (stage 10) (fig 36). Although qPCR analysis of CIC-S shows very low concentration in 8-cell (stage 4) embryos, likely due to greater sensitivity of qPCR vs RT-PCR, it is undetectable at MBT (stage 8). As development moves past MBT (stage 8), a period when the zygotic gene expression starts to be switched on (Venkatarama et al., 2010), the CIC-S isoform begins to be expressed in the early gastrulation stage (stage 10).

4.4.1.3 *CIC* is expressed around the blastopore during gastrulation.

In early gastrulation (stage 10), both CIC-S and CIC-L have similar expression levels and appear to be expressed in similar regions around the blastopore of the embryo in the *in situ* hybridisation data (fig 39-40). Both isoforms are expressed around the marginal zone of the embryo, a region where the presumptive mesoderm tissue is found in the developing embryo. Both isoforms are also enriched at dorsal marginal zone (DMZ), CIC-L appearing slightly more so. The DMZ is a key embryonic region

in early gastrulation that comprises of the Spemann-Mangold organizer, also known as the Spemann organizer. The Spemann organizer is a cluster of cells adjacent to the dorsal blastopore lip which induces development of the central nervous system. It gets its name from the landmark experiment in 1921 that established the concept of induction in amphibians. The formation of the vertebrate body axis, notably in the anterior–posterior (AP) axis, is strongly reliant on the induction from the Spemann organizer dorsal signalling centre. Although the presence of *CIC* expression at the DMZ does not guarantee its involvement with vertebrate body axis patterning, given the findings in *D. melanogaster* that CIC is involved with the regulation of the AP patterning (Jimenez et al., 2000), it could suggest CIC retains this role to some degree, but this will need further analysis.

In comparison to temporal expression studies in *X. tropicalis*, the *D. melanogaster* homolog of CIC has been shown to be highly expressed during the maternal phase of development, much like that observed for the CIC-L isoform of *X. tropicalis.* Unlike what is observed in *X. tropicalis, in D. melanogaster CIC* expression becomes undetectable during gastrulation (Jimenez et al., 2000). Although CIC-L expression is at its lowest point of expression during mid to late gastrulation in *X. tropicalis* development, it is still detectable by both RT-PCR and qPCR. In addition, CIC-S expression becomes activated at the start of gastrulation and increases throughout, which conflicts with the findings in *D. melanogaster*. Whilst it must be noted, this early study CIC expression in *D. melanogaster* embryonic development doesn't distinguish between the prominent isoforms of CIC, due to the fact that the CIC-S specific exon of *D. melanogaster* was not discovered until 2015 (Fores et al., 2015).

4.4.1.4 CICs expression is enriched at the neural plate.

As the *X. tropicalis* embryos proceed through development, both prominent isoforms, CIC-L and CIC-S have robust increases of expression as embryos proceed from gastrulation to the neurula phase (stage 18) of development. Embryos at the neurula stage of development undergo a process known as neurulation. Neurulation is the folding process in which the neural plate forms the neural tube (Burnside and Jacobson, 1968). Neurulation begins with the formation of the neural plate triggered by induction of the notochord (Shermoen and O'Farrell, 1991); this signalling causes the ectoderm germ layer to form the thick and flat neural plate. The neural tube later forms the brain and spinal cord structures giving rise to the central nervous system (CNS). Although both CIC-L and CIC-S undergo rapid increases in expression during the neurula stage of development CIC-S appears to have higher expression levels at

this point in development which may suggest CIC-S has an important role during neurulation, but further analysis will be required to prove this.

The CIC isoforms undergo rapid increases of expression during the neural stage of development (fig 36-37), which could suggest CIC repressional activity is at its greatest during neurulation. Although high mRNA expression doesn't guarantee high protein expression. Both CIC-L and CIC-S expression levels returns to levels at the early tailbud stage of development and decrease during the late tailbud stage. Analysis of CIC-L and CIC-S spatial expression during neurulation (stage 18) reveals that both isoforms have heavily enriched expression at the neural plate (fig 36-37). Expression is widespread except at the area of the neural groove. As CIC is known to function as a repressor of proliferation (Jimenez et al., 2002; Roch et al., 2002), if CIC acts in this context on the neural cells in the surrounding neural plate, those cells are likely to have reduced proliferation, whilst cells within the neural grove could be allowed to proliferate due to the lack of presence of CIC. It is important to realise that although there may be high concentrations of CIC mRNA in the neural plate, this does not necessarily lead to translation of CIC protein.

CIC is an important transcription factor in the development of the invertebrate brain in *D. melanogaster*. Studies have shown that CIC is critical for promoting growth and patterning of multiple organs, such as wings and eyes and more importantly, in the formation of the CNS of *D. melanogaster* (Yang et al., 2016). Mnb and Wap are known regulators of CIC repressional activity, functioning by binding and phosphorylating CIC. Mnb functions as a kinase and is an ortholog of mammalian DYRK1A, which acts through its adaptor protein Wap.

Experiments in adult *D. melanogaster* which were targeted by *Mnb* and *Wap* RNAi knockdown had reduced brain volumes, predominantly at the optic lobes, in comparison to the wild-type siblings. This revealed Mnb and Wap are required for normal brain growth. In addition, targeting of *CIC* by RNAi knockdown revealed that adults possessed larger brain volumes in comparison to wild-type siblings due to increased brain tissue growth which was unhindered by CIC repressional activity (Yang et al., 2016). CIC is known to be involved in the regulation of organogenesis in many tissue types. Importantly, the CIC RNAi knockdown was able to recuse the reduced brain volume phenotype previously associated with *Mnb* knockdown siblings, suggesting that down-regulation of CIC by Mnb promotes brain growth. This is not the first study to indicate CICs importance in normal brain development in the CNS.

Numerous studies in *H. sapiens* brain cancer, in particular ODG, highlight that CIC has an important role in CNS development and as putative tumour suppressor (Bettegowda et al., 2011a; Davoli et al., 2013; Yip et al., 2012). RAS/MAPK transduction signalling has repeatedly been shown to be a switch for the transition of cell lineages. One such example is the transition from neurogenesis to gliogenesis in the developing brain, due to the controlled expression of *Etv5* (Ahmad et al., 2018; Li et al., 2014; Li et al., 2012). As previously shown RAS/MAPK plays an important role on the control of CIC repressional activity by phosphorylation of the CIC protein. CIC functions in a 'default state' when no RAS/MAPK transduction signal is present, able to bind to its transcriptional targets, blocking transcription when no signal is received in the cell (Jimenez et al., 2000). Activation of a RTK by binding of ligand leads to the activation of the RAS/MAPK signalling cascade leading to the internalisation of di-phosphorylated MAPK and phosphorylation of CIC, preventing repression.

A recent experiment in *M. musculus*, found that mutations within the region of the *CIC* gene encoding for the HMG-box DNA binding site prevented it functioning in a 'default state' when no RAS/MAPK transduction signal was present in the cell, leading to increased derepression of CIC targets. This results in disruption of neural stem cells fates, resulting in the neural cells following a glial cell lineage instead of neuronal cell lineage, due to the derepression and upregulation of *Etv5* expression (Ahmad et al., 2018).

In ODG, a primary brain tumour which develops from a subset of glial cells, known as an oligodendrocyte. ODG is commonly found in the frontal lobe and less commonly in the temporal lobe. As previously mentioned ODG is caused predominantly by missense mutations (59.3% of cases) arising in the HMG-box and C1 domain coding regions of the CIC gene. Like the aberrant changes seen in normal neural stem cells fates which transition to a glial cell lineage due to misregulation of *Etv5*, CIC mutations in ODG lead to derepression of Etv5 and other PEA3 family of ETS transcription factors (Etv1 & Etv4) promoting cellular proliferation and migration. Indeed, this also corresponds with previous findings in a *M. musculus* study when the ATXN1-CIC complex was disrupted, it results in abnormal maturation and maintenance of upper-layer cortical neurons, leading to impairment of learning, memory and causing multiple behavioural abnormalities (Lu et al., 2017b). In addition, germline heterozygous *CIC* truncating mutations were found in a number of patients suffering from intellectual disability, attention deficit hyperactivity disorder and autism spectrum disorder (Lu et al., 2017b). The findings from these studies highlight CIC is heavily involved with regulation neural patterning, development of the brain and CNS.

Perhaps unsurprisingly the CIC isoforms appear to be highly expressed during the neural phase of development, which further demonstrate the likelihood that CIC is involved in the formation of the CNS in early embryonic vertebrate development. These findings and the increased expression of the prominent isoforms could suggest a greater requirement CIC repressional activity during neurulation in the embryo although this remains to be speculation and although high mRNA expression doesn't guarantee high protein expression.

4.4.2 CIC isoforms are expressed in the temporal and spatial regions as FGF4 and FGF8.

4.4.2.1 Maternal expression and FGF ligand expression.

The findings from this study allow a detailed comparison of *CIC's* temporal and spatial expression to the well-established temporal and spatial expression profiles of FGF. One of the aims of this study is to discover if CIC and FGF share similar temporal and spatial expression during early embryonic development in *X. tropicalis*. In early development, like CIC-L, *FGF1, FGF2, FGF13* and *FGF22* are maternally expressed in *X. tropicalis (Lea et al., 2009)*, although, FGF13 is a member of the intracellular FGFs (iFGFs) (*FGF11, FGF12, and FGF14*).

Activation of the MAPK transduction is entirely under the regulation of FGF signalling in the very early stages of amphibian development (Branney et al., 2009; Christen and Slack, 1999; LaBonne and Whitman, 1997; Shinya et al., 2001). MAPK activation is present at low levels from the early cleavage to blastula stages in *X. tropicalis* (Branney et al., 2009; Christen and Slack, 1999; Shinya et al., 2001). These FGFs are present when CIC-L is expressed in the maternal phase of development. Activated MAPK begins to increase at stage 8.5 and has a large increase by stage 9. FGF8 expression corresponds to the rapid increase of activated MAPK at stage 8.5-9 (Branney et al., 2009). FGF1, FGF2 and FGF22 are all known mediators of MAPK transduction and may be responsible for low level activation of MAPK seen prior to MBT (Branney et al., 2009; Shinya et al., 2001). If FGF does regulate CIC during the maternal phase, it likely to be these FGFs that regulate CIC-L by activation of MAPK transduction.

4.4.2.2 FGF ligand expression during gastrulation.

During gastrulation, *FGF1, FGF2, FGF20* and *FGF22* are expressed at the marginal zone of the *X. tropicalis* embryo (Lea et al., 2009), although FGF22 expression

becomes undetectable during late gastrulation. Importantly, like the prominent isoforms of CIC, FGF4 and FGF8 are expressed during early gastrulation (stage 10), at the marginal zone in the early presumptive mesoderm (Branney et al., 2009; Isaacs et al., 1995b; Lea et al., 2009). FGF4 is expressed in response to Activin (Fisher et al., 2002a). FGF4 and FGF8 maintain expression of a subset of genes which are important for mesoderm induction (Fletcher and Harland, 2008). Both FGF4 and FGF8 are essential for paraxial mesoderm formation (Fisher et al., 2002a; Fletcher et al., 2006; Isaacs et al., 2007; Isaacs et al., 1995b). As this study has shown, both isoforms of CIC are expressed in the same region as the FGFs. If CIC retains some role in anterior-posterior axis formation seen in *D. melanogaster* it is likely to be involved during gastrulation when the anterior-posterior specification begins (Jimenez et al., 2000; Klingler et al., 1988).

4.4.2.3 Expression of FGFs during neurulation.

In the early neurula embryo FGF1, FGF2, FGF3, FGF4, FGF8, FGF19, FGF20, FGF22 are expressed. FGF4 becomes expressed early in neurula stages through to late tailbud stages. FGF4 is expressed in the posterior mesoderm of the neurula embryo, with peak expression at stage 15 (Lea et al., 2009). FGF4 later becomes expressed at the tailbud in the later stages of development (stages 25-40). In addition, FGF4 is weakly expressed at the otic vesicle and midbrain-hindbrain boundary (Lea et al., 2009).

FGF8 is expressed throughout neurulation with peak expression at stage 15. Expression decreases towards the end of neurulation and remains low throughout the early and late tailbud stages of development. Although FGF8 is involved in the regulation of mesoderm induction, it is also plays a role in neural tissue induction (Christen and Slack, 1997b; Fletcher et al., 2006). FGF8 is expressed forebrain, tailbud, otic vesicle, branchial arches somites and midbrain-hindbrain boundary (Lea et al., 2009). FGF8 is important for posterior neural tissue induction (Christen and Slack, 1997b; Fletcher et al., 2006).

Both temporal and spatial expression data demonstrates that CICs expression overlap FGF4 and FGF8 expression. Although these findings suggest CIC encounters FGF within the cell further analysis will be required to establish if FGF signalling interacts and regulates CIC repressional activity. Spatial and temporal analysis combined give a clear picture that FGF and CIC are expressed and present in the same periods and regions in embryonic development. Like FGF 4 and 8, both

long and short isoforms of CIC are expressed at the marginal zone during gastrulation. Although this is not direct evidence that they interact it does provide information that they are at least are in the presence of each other. Further studies analysing the effects of CIC knockout will address this question. Although this project primary focus is to analysis if mesoderm regulating FGF4 and FGF8 interact with CIC during gastrulation. Interaction with one FGF may suggest that is indicative to be involved with others.

Chapter 5: Investigating the function and stability of CIC protein by knockout and overexpression

5.1 Introduction

In the previous chapter of this thesis, we established the alternative expression profiles of the prominent isoforms of *CIC* (CIC-L & CIC-S) in the developing *X. tropicalis* embryo. *CIC* expression was found to be evident in same the spatial and temporal locations of development as *FGF4* and *FGF8* expression (Lea et al., 2009). The next phase to this project is to identify if *CIC* is involved downstream of the FGF pathway. Further analysis would be required to uncover if there is a relationship between FGF signal transduction and CIC repressional activity.

The reverse genetics method for discovering if a gene is involved in a biological pathway is to alter gene expression and observe if any phenotypical changes occur. If expression of a given gene is above or below a critical threshold required for normal biological function within the cell/embryo, it can result in a mutant phenotype. Changes in phenotype due to the alteration of the wildtype expression of a given gene can resemble previously seen phenotypes which can give insight in to which pathways a given gene is involved in eg. tissue or organogenesis.

To address if CIC is involved downstream of the FGF pathway during mesoderm formation; genome modification tools, such as the transcription activator-like effector nucleases (TALENs) (Joung and Sander, 2013; Lei et al., 2012) and mRNA targeting morpholinos (Nasevicius and Ekker, 2000; Summerton, 1999) will be implemented to create *CIC* knockdown of expression. TALENs consist of a TAL effector target DNA-binding domain which is fused to the *FokI* endonuclease which cleaves double stranded DNA, introducing double-strand breaks (DSB) (fig 41). The TAL effector target DNA-binding domain can be designed to target specific regions. Non-homologous end joining (NHEJ) repair (Moore and Haber, 1996) re-joins the DNA from either side of the DSB. The NHEJ repair mechanism introduces errors in the genome by the introduction of either insertions or deletions (indels) of nucleotides, or chromosomal rearrangement. Changes due to introduction of indels may alter the ORF within the gene causing frame shifts rendering the protein products as non-functional or lead to nonsense mediated decay (NMD) of the mRNA.
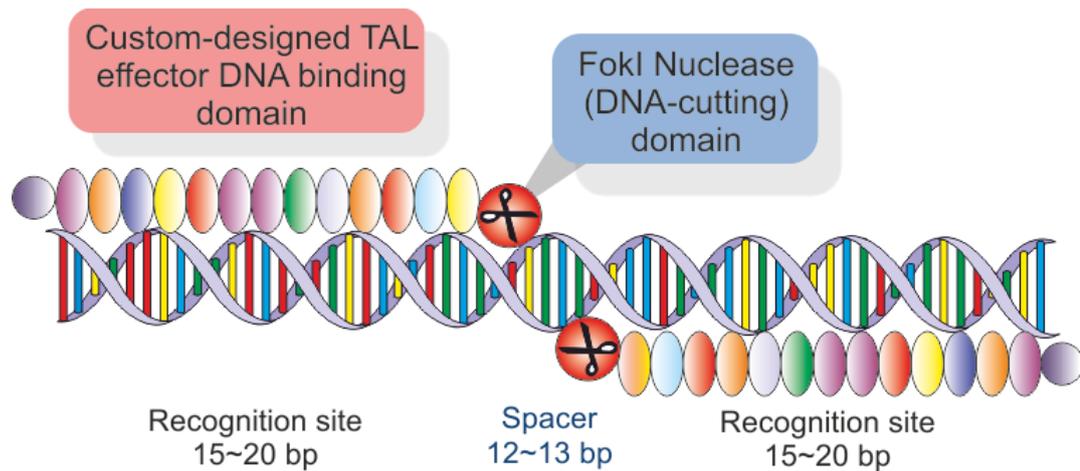
Figure 41, the forward and reverse TALEN are designed to bind to recognition sites upstream and downstream of the gene target. A fok1 endonucleases introduces strand breaks.

An alternative approach to TALENs gene knockdown is the mRNA targeting morpholinos, these are short single-stranded DNA analogue oligomers which contain a phosphorodiamidate backbone. The morpholinos function by binding to target mRNA acting to sterically block the binding of other molecules to their mRNA targets. This can lead to prevention of the initiation of translation of the given protein products, leading degradation of target mRNAs and knockdown of the target gene (Nasevicius and Ekker, 2000; Summerton and Weller, 1997). In addition, morpholinos can also be used to inhibit RNA splicing (Vetrini et al., 2006).

Morpholinos can be used to target a specific isoform of a given gene whilst allowing other isoforms to be expressed. This approach will be used to determine the functions of the prominent isoforms of *CIC* (CIC-L & CIC-S). Morpholinos have an advantage over other knockdown technologies in that they are capable of acting as soon as they are introduced into the cell. Whereas, TALEN mRNA is required to be translated by the cells own translational machinery before becoming active and introducing somatic mutations. This can result in mosaicism due to some cells dividing before being targeted and some cells being targeted but having the introduction of different sized indels. It is of interest to establish if there is an effect on embryo development if morpholinos are used to target the maternally expressed CIC-L given the fact that TALEN cannot target CIC for knockdown at the earliest stages of development. We know that CIC-L is maternally expressed (chapter 4) and targeting this maternally deposited mRNA could give insight to its function in early development in comparison to the zygotically expressed CIC-S isoform.

FGF overexpression and knockdown have distinct phenotypes in the developing *Xenopus* embryos. Overexpression of FGF leads to truncation and loss of anterior structures of the anterior-posterior axis of the embryo whilst knockdown leads to anteriorisation of the embryo (Amaya et al., 1991; Amaya et al., 1993; Fletcher et al., 2006; Isaacs et al., 1994). To access if CIC is involved in the FGF pathway, knockdown and overexpression studies of *CIC* could give an indication if it has a role downstream of FGF (Amaya et al., 1991; Amaya et al., 1993; Fletcher et al., 2006; Isaacs et al., 1994). The *H. sapien*s (Dissanayake et al., 2011) and *M. musculus* (Kim et al., 2013) homologs of *CIC* will be injected as mRNA into the early embryo to look at the effects of overexpression on amphibian development. In simplistic terms, if *CIC* is overexpressed and has a role in the FGF pathway, the phenotype would likely resemble that of *FGF* knockdown, whilst knockdown of *CIC* would likely resemble that of *FGF* overexpression. Once translation of *H. sapiens* CIC homolog is confirmed, western blot analysis will be utilised to determine if and how FGF dependent MAPK phosphorylation of CIC effects its abundance and sub-cellular localisation. The *H. sapiens* homolog of CIC tagged with eGFP will be co-injected with FGF4 and FGF8 mRNA into *X. tropicalis* embryos.

This chapter will address the following aims which are to:

- Determine the effects on *X. tropicalis* development from overexpressing homologs of *CIC* by microinjection of mRNA.
- Determine the effects on *X. tropicalis* development from inhibiting *CIC* by knockdown by TALENs.
- Determine the effects on *X. tropicalis* development from inhibiting prominent isoforms of CIC by morpholino microinjection.
- Determine how FGF dependent MAPK phosphorylation of CIC effects its abundance and sub-cellular localisation.

5.2 Results

5.2.1 Overexpression analysis of CIC homologs

Over expression studies of *CIC* were performed to determine the effects on amphibian development. These experiments aimed to establish if the overexpression of CIC showed a similar phenotype to inhibition of FGF caused by the dominant negative FGFR (dn-FGFR) (Isaacs et al., 1994). This result would be expected if

MAPK signalling downstream of the FGFR could not regulate CIC repression due to its overexpression.

5.2.1.1 No change of phenotype when overexpressing *X. tropicalis* CIC-S
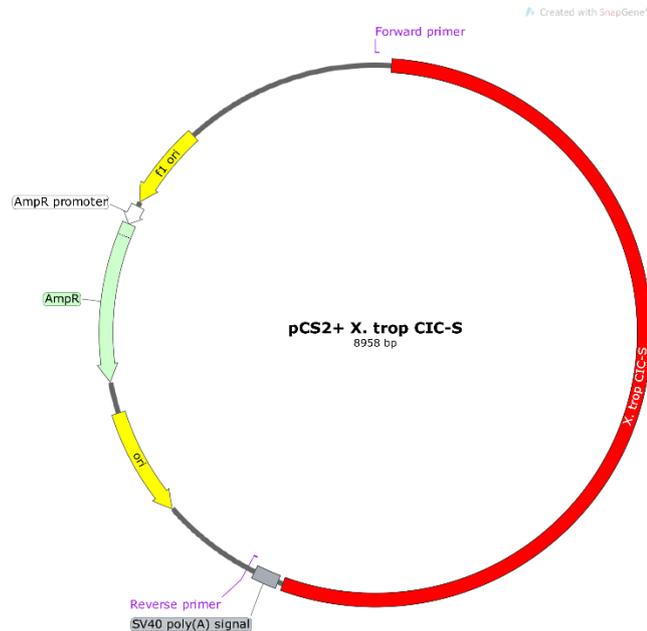
Figure 42, a plasmid map of pCS2+ X. trop CIC-S plasmid containing the ORF of *X. tropicalis* CIC-S homolog.

To overexpress the endogenous *X. tropicalis* CIC-S isoform primers were designed to create PCR based template from the pCS2+ *X. trop* CIC-S plasmid. The 5,185 bp amplicon template contained the SP6 promoter, enabling SP6 transcription, the ORF of CIC-S and SV40 poly(A) signal (fig 42). Embryos were micro-injected with mRNA made from the template at 320 ng/ul.

When overexpressing the *X. tropicalis* homolog no change in phenotype was observed in comparison to wildtype siblings. Although it must be noted that a limitation of the *X. tropicalis* CIC-S overexpression experiment was that it was not possible to validate the translation of the CIC-S protein by western blot because no antibody existed to target the *X. tropicalis* homolog of the CIC protein.

Although as previously mentioned the overexpressed *X. tropicalis* CIC-S isoform has not been validated. Further work will be required to validate this result either by introducing tags (eGFP/Myc) upstream of the ORF of CIC-S in the pCS2+ X. trop CIC-S plasmid or producing an antibody specific for the *X .tropicalis* CIC protein for detection of translation by western blot.

5.2.1.2 No change of phenotype when overexpressing the *M. musculus* CIC-S
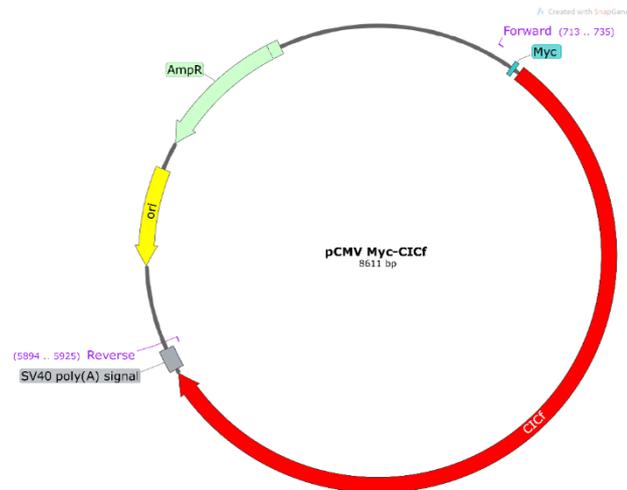
Figure 43, a plasmid map of pCMV Myc-CICf plasmid containing the ORF of *M. musculus* CIC-S homolog (Kim et al., 2013).

Initial overexpression experiments utilised the pCMV Myc-CICf plasmid which contained the ORF frame for a Myc-tagged *M. musculus* homolog of the CIC-S isoform (fig 43). Primers were designed to create PCR based template to generate an amplicon of 5,213 bp, which would contain the Myc-tag, the ORF of the *M. musculus* homolog of CIC-S and the bovine growth hormone polyadenylation signal (bGH poly (A) signal) to produce mature mRNA. The forward primer was designed to contain a SP6 promoter enabling SP6 transcription from the PCR amplicon template. Embryos were micro-injected with mRNA made from the template encoding for the Myc-CIC-S at 320 ng/ul.

The embryos appeared to have no observable difference in phenotype in comparison to the wild-type siblings. To determine if the *M. musculus* Myc-CIC-S mRNA was being appropriately translated within the injected embryos, 10 embryos (stage 15) were collected for western blot analysis. The translated mRNA should produce a 250 kDa protein. The equivalent of 5 embryos were added per lane. Western blot analysis confirmed appropriate translation and the appearance of a 250 kDa protein (fig 44).
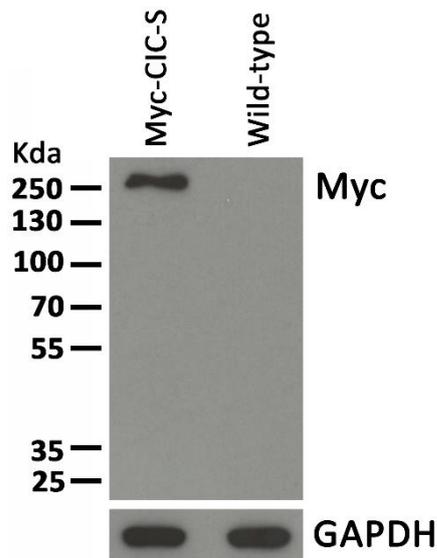
Figure 44, the western blot detecting the *M. musculus* Myc-CIC-S protein (250 kDa, arrow) produced from mRNA injected at 320ng/ul into *X. laevis*. 5 x embryos were collected at the early neurula stage (15). No protein product detected in the un-injected wild-type control sample.

5.2.1.3 No change in phenotype when overexpressing *H. sapiens* CIC-S
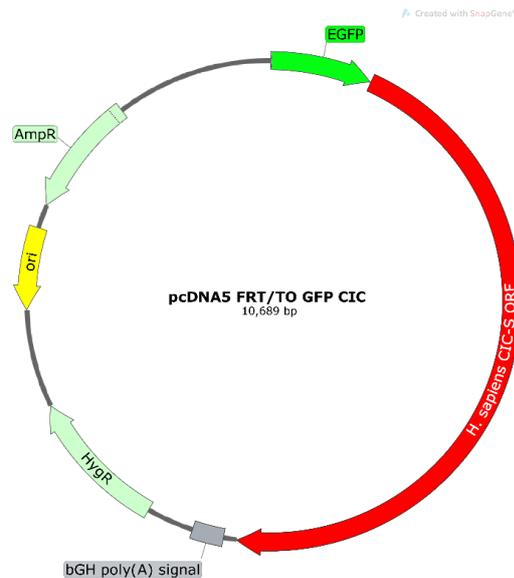
Figure 45, the plasmid map of pcDNA5 FRT/TO GFP CIC plasmid containing the ORF of *H. sapiens* CIC-S homolog (5.5Kb) (Dissanayake et al., 2011).

An alternative approach to detect translation of the CIC protein was to obtain the pcDNA5/FRT/TO GFP-CIC plasmid which contained the ORF of *H. sapiens* CIC-S homolog, tagged with the ORF of eGFP protein (fig 45) (Dissanayake et al., 2011). This approach would allow initial easier detection of appropriate translation of the eGFP-tagged CIC-S protein by fluorescence microscopy. Utilising an alternative CIC-

S homolog would also have the benefit of establishing if the *H. sapiens* homolog of CIC-S would produce a mutant when no phenotype was present for the *M. musculus* homolog of CIC-S. Preliminary micro-injection experiments utilised the pcDNA5/FRT/TO GFP-CIC plasmid, rather than creating mRNA (fig 45) (Dissanayake et al., 2011). If the eGFP signal could be detected in plasmid micro-injections it would confirm that the plasmid could be used to produce a template to create mRNA for the eGFP-CIC-S protein. 300pg total DNA of plasmid was micro-injected in the early cleavage staged *X. laevis* embryos at 2 - 4 cell stages. Sibling embryos were injected with 300pg total DNA of pCMV-GFP plasmid (Matsuda and Cepko, 2004) as a comparison to the eGFP fluorescence signal.
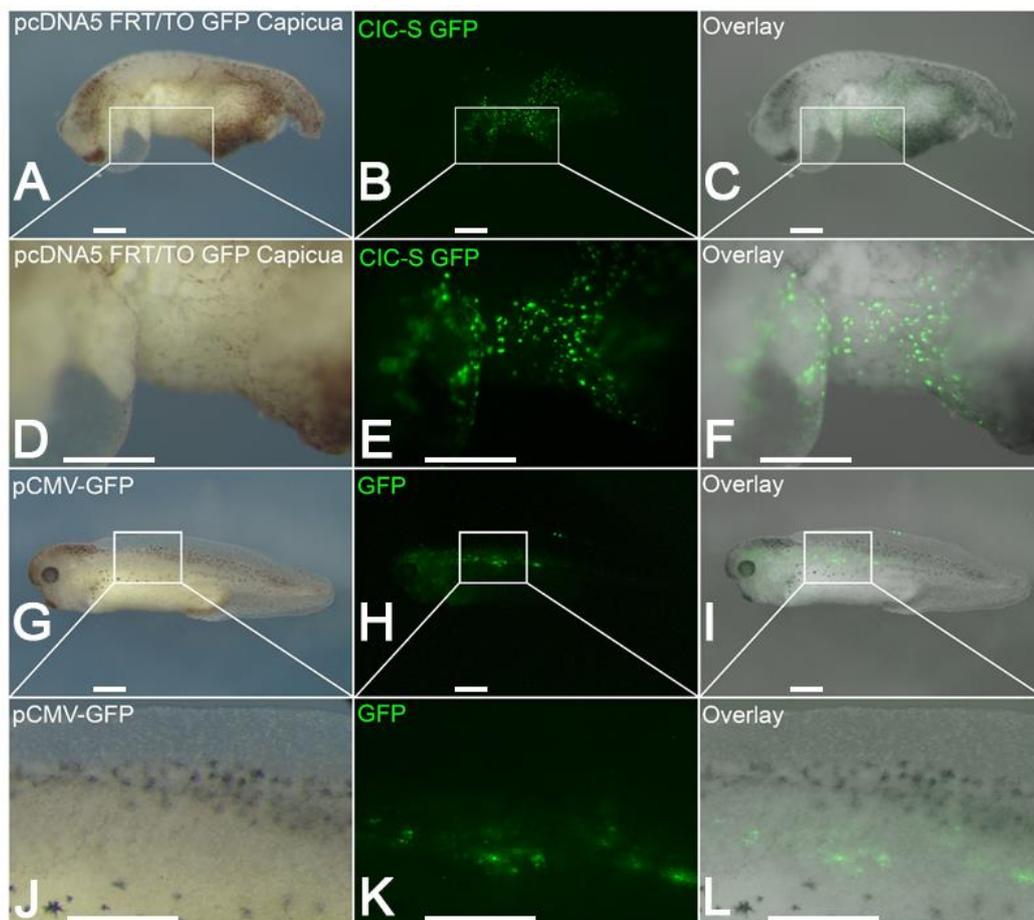
Figure 46, (A-F) shows a pcDNA5 FRT/TO GFP CIC injected embryo and (G-L) a pCMV-GFP injected sibling embryo at late-tailbud stage of development (35-36) (Lateral view). (A, D, G & J) Images taken with brightfield microscopy, (B, E, H & K) images taken with florescence microscopy and (C, F, I & L) images which have been combined showing both brightfield and GFP signal. (D, E, F, J, K, L) Shows images of higher magnification. Scale bars: 250 µm.

pcDNA5/FRT/TO GFP CIC plasmid injected embryos showed a strong eGFP signal which was enriched at the nucleus (fig 46b, c, e & f). pCMV-GFP microinjected embryos showed a strong eGFP signal (fig 46h, i, k & l), the GFP signal was

widespread throughout the cell. pCMV-GFP micro-injected embryos appeared to have a wild-type phenotype (fig 46g).

5.2.2 Optimising the PCR template design for the *H. sapiens* homolog CIC-S mRNA for micro-injection

Once confirmation of an eGFP-CIC-S protein and strong eGFP-CIC signal was detected, the next approach was to use the ORF of the *H. sapiens* CIC-S isoform contained in pcDNA5 FRT/TO GFP CIC plasmid (fig 45) to produce a PCR template for the creation of mRNA for micro-injection. The template and mRNA would be created using the same approach utilised for the *M. musculus* Myc-CICf mRNA production, as the ORF proved difficult to clone into pCS2+ due to the large size of amplicon (fig 43-44) (Dissanayake et al., 2011). To optimise and allow efficient translation of protein from the mRNA, three approaches to producing the mRNA were tested, (i) a PCR based template would be designed which would contain a 6bp linker between the SP6 promoter and the eGFP coding region (fig 45 & 47m), (ii) to produce a template with 126bp linker between the SP6 promoter and eGFP coding region (fig 48n) and (iii) to produce a PCR based template which does not amplify the bGH-poly A signal contained within the plasmid (fig 47o). Instead, a poly-A-tailing kit (Ambion) would be used to create mRNA with a chemically applied poly-A tail.

Template design	Contains:		
	6bp linker	126bp linker	bGH-poly A
(i) (fig 46m)	x		X
(ii) (fig 46n)		x	X
(iii) (fig 46o)	x		

mRNA was injected at a concentration of 320 ng/ul in to 2-4 cell staged *X. tropicalis* embryos (6.4ng total mRNA per embryo).
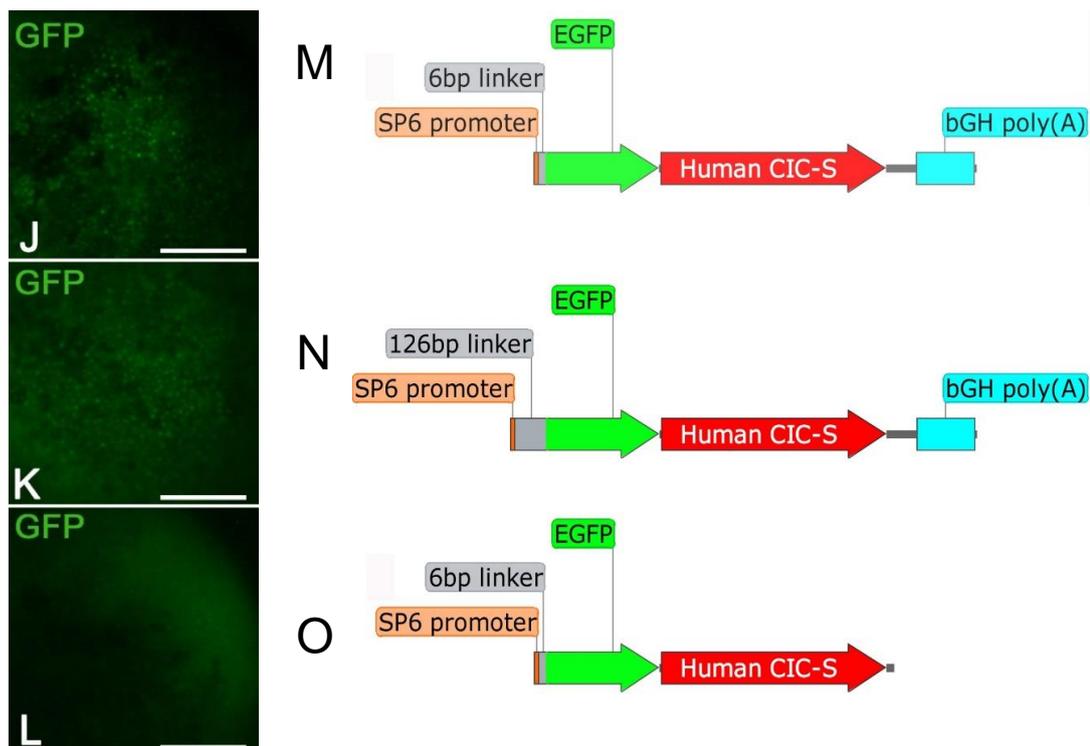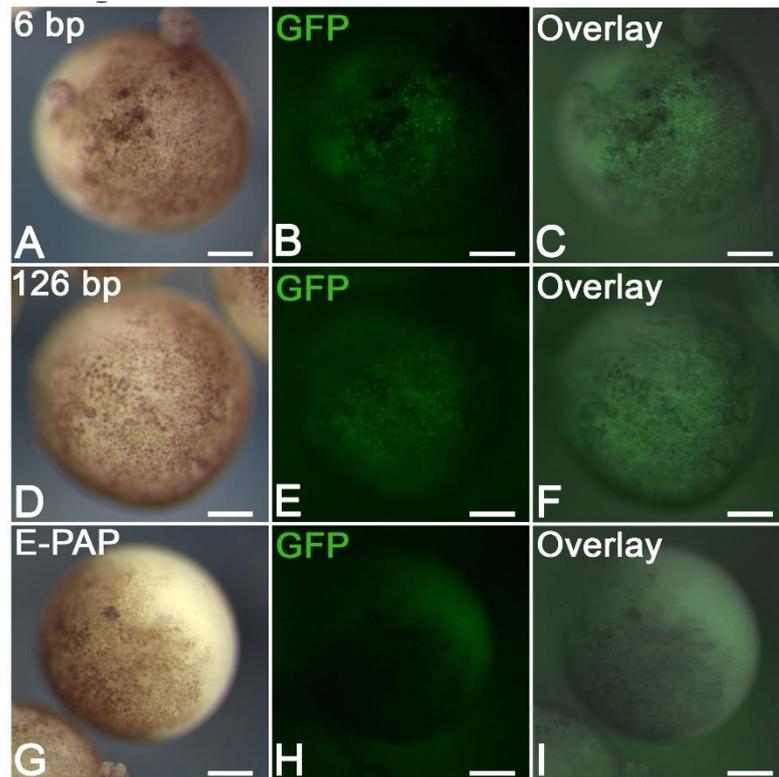
Figure 47, (A-C & J) *X. tropicalis* embryo injected with mRNA created from 6bp linker upstream of the eGFP coding sequence (M). (D-F & K) embryo injected with mRNA created from 126bp linker upstream of the eGFP coding sequence (N) and (G-I & L) embryos injected with mRNA created 6bp linker with chemically attached poly (A) signal (O). All 3 sets of mRNA were injected at 320ng/ul concentrations (6.4ng total per embryo) at the 2-cell stage. (A, D & G) Images taken with brightfield microscopy, (B, E, H and J-L) images taken with florescence microscopy and (C, F & I) images combined with both brightfield and GFP signal (animal view). (M-O) schematics

of mRNA template design. (E, F & K) the mRNA generated using the 126bp linker and bGH poly-A signal appeared to give the strongest nuclear GFP signal. Embryos were imaged at the early gastrula stage 11. Scale bars: 250 µm.

Fluorescence microscopy revealed that mRNA micro-injections produced a eGFP signal which appeared to be nuclear (fig 47j-k). This would be expected as CIC-S is a nuclear protein and demonstrates the eGFP-tagged protein does not hinder the import of the CIC protein into the nucleus. Further analysis of the fluorescence signal from the injection experiment revealed that mRNA from PCR template (ii) produced the strongest signal (fig 47d-e & j), this was the template design which contained the 126 bp linker between the SP6 promotor and the eGFP coding sequence and contained the ORF for bGH poly-A (fig 47n).

mRNA from PCR template design (i) produced a eGFP signal, but was it was not as strong or as widespread throughout the embryo as the eGFP signal from the mRNA from PCR template design (ii) (fig 47a-c & j). PCR template design (i) contained the 6 bp linker between the SP6 promoter and the ORF for bGH poly-A (fig 49m). Both sets of mRNAs, containing the bGH poly-A, produced a nuclear GFP signal (fig 49b, e & j-k, m-o). The strongest GFP signal was found in the mRNA produced with the longer 126 bp linker (fig 47k, n). mRNA from PCR template design (iii) which did not contain the bGH poly-A, did not produce a strong nuclear GFP signal (fig 47l & o).
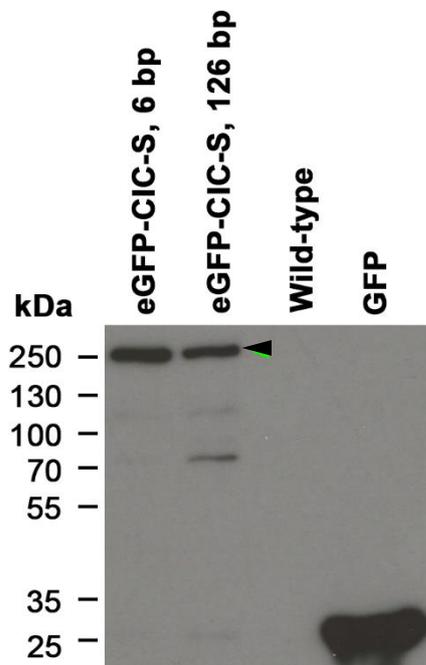
Figure 48, the western blot detecting the *H. sapiens* eGFP-CIC-S protein (250 kDa, arrow) produced from both mRNA designs (126 bp & 6 bp linker) injected at 320ng/ul into *X. laevis*. 5 x embryos were collected at the early neurula stage (14). A GFP protein of 27 kDa was detected (positive control) and no products were detected in the un-injected wild-type control.

The western blot protocol was optimised for larger proteins, 5 embryos were collected per cell lysis. Western blot analysis was performed on embryos micro-injected with mRNA created from 6bp linker (i) (fig 47m) and 126bp linker (ii) (fig 47n) templates. The western blot detected a GFP-tagged 250 kDa protein and the GFP protein control at 27 kDa using GFP-antibody (fig 48). This revealed that the full length eGFP-CIC-S protein was produced from the both (i), (ii) template designs. All further experiments utilised the (ii) PCR template to produce *H. sapiens* GFP-CIC-S mRNA. No change in phenotype was observed in comparison to the wild-type non-injected sibling embryos by the latter stages of development (stage 37-40).

5.2.3 Increased overexpression CIC-S has no change of phenotype in *X. laevis* embryos
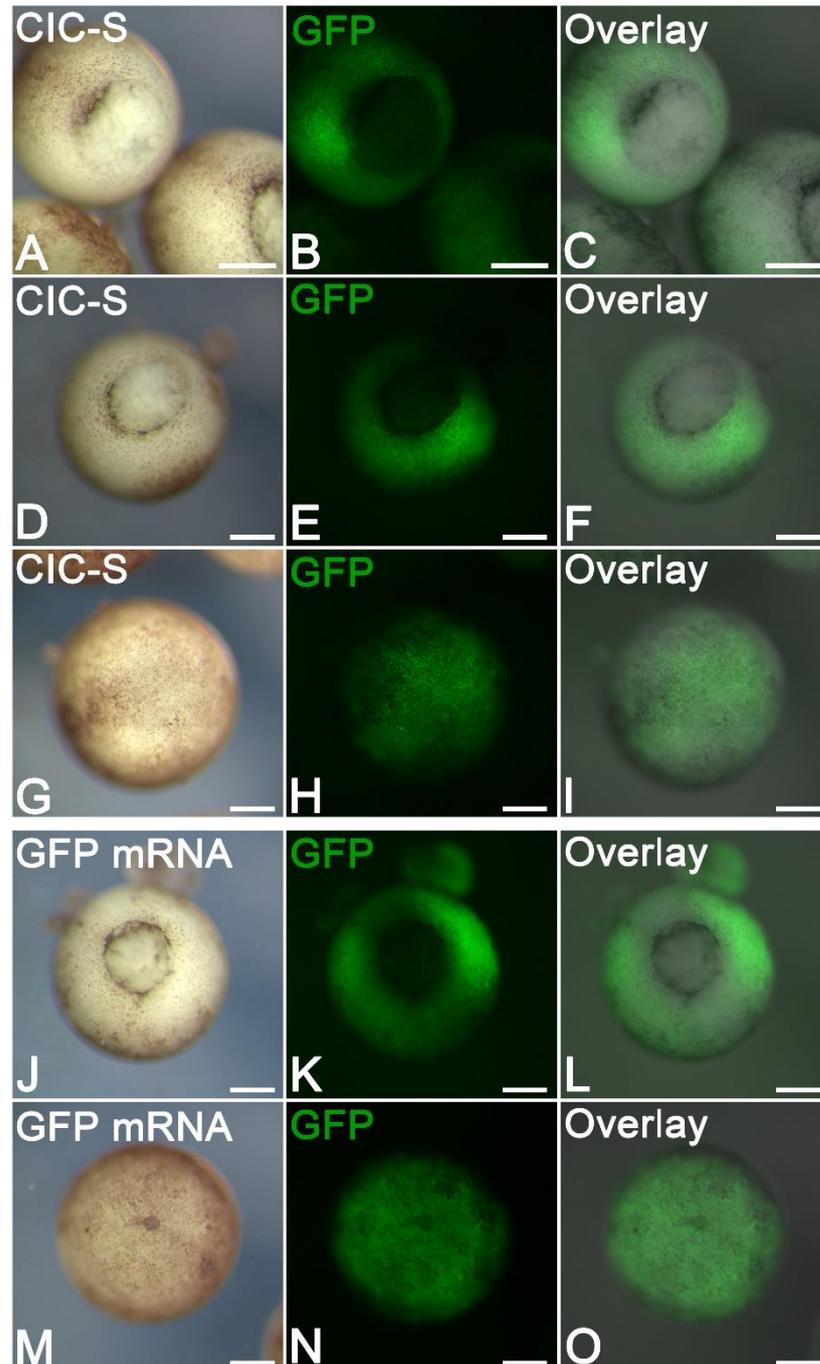
Figure 49, (A-I) *X. laevis* embryos injected with *H. sapiens* GFP-CIC-S mRNA (126bp linker + bGH poly-A) at 2-cell stage with 720 ng/ul (14.4 ng total per embryo). (J-O) embryos injected with GFP mRNA at and 10ng/ul (0.2 ng total per embryo). (A, D, G, J & M) Images taken with brightfield microscopy, (B, E, H, K & N) images taken with florescence microscopy and (C, F, I, L & O) images which have been combined showing both brightfield and GFP signal. Images were taken at early to mid-gastrula stages (10.5-11.5) of animal & vegetal view. Scale bars: 250 µm.

To determine if a change in phenotype could be found in the embryos microinjection concentrations of *H. sapiens* eGFP-CIC-S mRNA were increased to 720 ng/ul (14.4 ng total per embryo) (fig 49a-i) and GFP mRNA was injected at 10ng/ul (0.2 ng total per embryo) (fig 49j-o) as a comparison. Embryos containing *H. sapiens* GFP-CIC-S mRNA showed a widespread, stronger signal than previously seen in the 320ng/ul eGFP signal in the GFP mRNA injected embryos. No noticeable change in phenotype was found in comparison to wild-type siblings in either the eGFP-CIC-S mRNA embryos or the GFP mRNA injected embryos by the latter stages of development.

5.2.4 Treating *X. tropicalis* embryos with FGF leads to the degradation and post-translational modification of CIC-S

To establish if there was a relationship between FGF signal transduction and CIC repressional activity, *X. tropicalis* embryos were co-microinjected with *H. sapiens* GFP-CIC-S mRNA and FGF4 and/or FGF8 mRNA. This would establish if treating embryos injected with FGF would have any effect on CIC. The experiment followed the following design:

Experiment	GFP-CIC-S mRNA	FGF4 mRNA	FGF8 mRNA
A	X		
B	X	X	
C	X		X

Embryos were injected 2 ng total per embryo of *H. sapiens* GFP-CIC-S mRNA, whilst embryos which were injected FGF4 were injected with 6.25 pg total per embryo and embryos injected with FGF8 were injected with 12.25 pg total per embryo. FGF4 was injected with a higher concentration due to its increase potency.
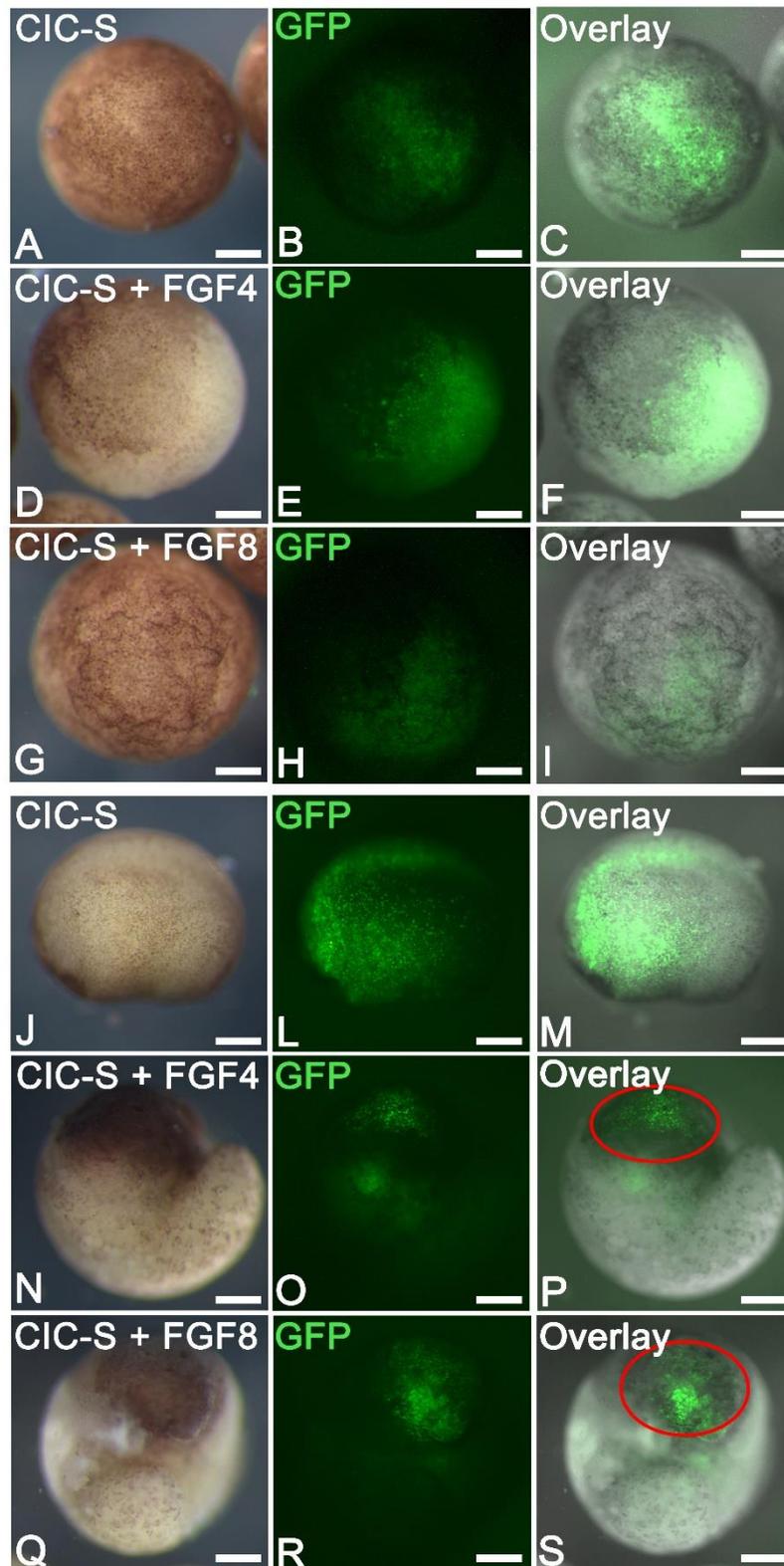
Figure 50, (A, B & C, J, L & M) *X. tropicalis* embryos injected with *H. sapiens* GFP-CIC-S mRNA, (D, E, F, N, O & P) embryos co-injected with GFP-CIC-S mRNA and FGF4 mRNA, (G, H, I, Q, R & S) embryos co-injected GFP-CIC-S mRNA and FGF8 mRNA. (A-I) Images taken of embryos at stage 10 (animal view) and (J-S) images taken at stage 23 (lateral view). (A, D, G, J, N & Q) images taken with brightfield microscopy, (B, E, H, L, O & R) images taken with

A eGFP signal was present in all embryos examined at stages 10 and stages 23 (fig 50a-i). Embryos injected with only *H. sapiens* GFP-CIC-S mRNA appear to have a wild-type appearance throughout development (fig 50a, b & c, j, l & m). The eGFP signal can be seen to be widespread throughout the embryo and nuclear in appearance. Embryos injected with *H. sapiens* GFP-CIC-S and FGF4 mRNA have a wild-type phenotype at stage 10 with a widespread eGFP-signal throughout the embryo (fig 50g-i). By stage 23 of development, embryos have become exogastrulated and the eGFP-signal is concentrated and nuclear in the presumptive animal pole and appears to be dispersed and cytoplasmic in areas outside of the animal pole region (fig 50q-s). This suggesting that the eGFP-CIC-S protein is targeted for degradation at the marginal zone and vegetal pole, whilst being unaffected in the presumptive animal pole (fig 50p & s).

H. sapiens GFP-CIC-S and FGF8 injected embryos, like those co-injected with FGF4, have a wild type phenotype at stage 10 with a widespread GFP-signal throughout the embryo with a GFP-CIC-S signal which appears to be nuclear. At stage 23, *H. sapiens* GFP-CIC-S and FGF8 injected embryos similarly to FGF4 injected embryos are exogastrulated. Again, the eGFP-signal appears to be concentrated and nuclear in presumptive animal pole and the eGFP signal appears to be dispersed and cytoplasmic in the rest of the embryo.
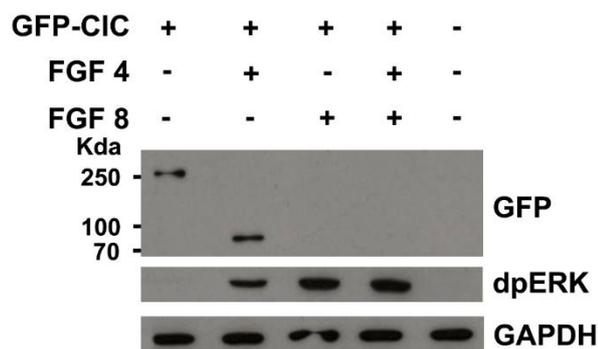
Figure 51, the western blot gel image showing the results of co-injections of GFP CIC-S mRNA, FGF4 or/and FGF8 mRNA. The *H. sapiens* GFP-CIC protein at 250 kDa when not co-injected with FGF4 or FGF8 mRNA. 5 x embryos were collected at the early neurula state (14) in *X. tropicalis*.

5 x embryos were collected for western blot analysis to observe the effects of the protein concentration when co-injecting with FGF4 or/and FGF8 mRNA with *H. sapiens* GFP-CIC-S mRNA (fig 51). All samples injected with either/or FGF4 and FGF8 produced protein product of expected size for dpERK on the western (fig 51).

No observable dpERK band appeared for the GFP-CIC-S only injected sample or the control un-injected sample. All samples produced a GAPDH band suggesting samples were equally loaded onto the western. A 250 kDa protein was detected in embryos injected with *H. sapiens* GFP-CIC-S as previously seen in the GFP-CIC-S mRNA optimisation experiments (fig 47), with no observable dpERK band signal. Due to the difficulty of the experiment this result was only repeated twice.

In embryos injected with GFP-CIC-S mRNA and FGF4 mRNA no 250 KDa protein was detected in the western blot with the GFP-antibody (Clonetech), instead a smaller 80-90 kDa protein was detected (fig 51). This could be a product of protein degradation or post translational modification of the GFP-CIC-S protein. Samples injected with GFP-CIC-S mRNA and FGF8 mRNA no protein product was detected, likely due to total degradation of the protein. Embryos were co-injected with GFP-CIC-S mRNA, FGF4 and FGF8 mRNA, similarly to FGF8 co-injections, did not produce a GFP-CIC-S protein product, suggesting total degradation of the GFP-CIC-S protein. This indicates direct evidence that FGF signalling leads to changes of CIC concentration within the embryo.

5.2.5 Analysis of the *CIC* gene knockdown in early *X. tropicalis* development

All knockout/knockdown experiments were performed using the *X. tropicalis* species, rather than *X. laevis*. The *X. tropicalis* species has the benefit of being a diploidy organism which is a simpler model for genetic studies in comparison to the allotetraploid *X. laevis* species. Because experiments would be performed in the *X. tropicalis* it would reduce the variation of phenotypes because less alleles would require targeting in knockout/knockdown experiments. Being a diploid organism would mean that it would be a closer representation to *H. sapiens* genetics (Hellsten et al., 2010).

5.2.6 Knockdown of *CIC* leads to posteriorisation of the developing *X. tropicalis* embryo

Knockdown of *CIC* in *X. tropicalis* was done using TALENs genome editing technology (fig 42 & 52) (Boch, 2011). In *X. tropicalis, the* HMG-box coding region spans two exons, exon 6 and 7. The TALENs were designed to flank and target exon 6 (fig 52 & 54). Because of the conserved nature of the HMG-box coding region, the design of the TALENs could be applied to both *X. tropicalis* and *X. laevis* species.
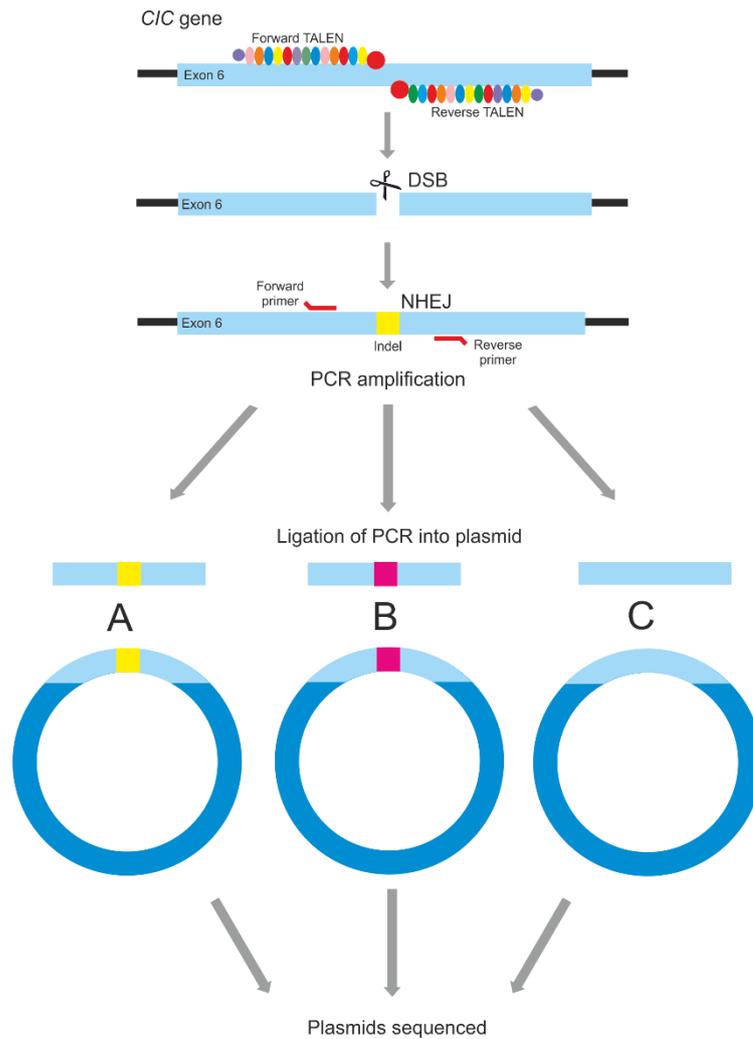
Figure 52, the TALENs experimental pathway. TALENs are injected into the embryo at 1-2 cell stage of development. The TALENs target the exon 6 of the *CIC* allele, the coding region for the HMG-box. Targeting by the TALENs leads to double strand breaks. Nonhomologous end joining repair introduces somatic mutations (indels) which are passed on to progeny of the mutated cell. DNA is extracted, and the targeted region is amplified by PCR. PCR products are ligated into plasmid and then sequenced to confirm, establish efficiency and penetrance of targeting.

To determine optimal concentration of TALENS mRNA to inject in *X. tropicalis*, forward and reverse TALENs mRNA were injected at 1, 2 and 4 cell stage embryos at varying concentrations, 1ng, 0.5ng and 0.25ng total per embryo. Forward only TALEN mRNA was injected at 1ng, 0.5ng and 0.25ng total per embryo as a control. In addition, un-injected embryos were used as control comparisons.

	1ng total (Fert 1)	0.5ng total (Fert 2)	0.25ng total (Fert 3)	1ng total (Fert 1)	0.5ng total (Fert 2)	0.25ng total (Fert 3)	Fert 1	Fert 2	Fert 3
	TALEN forward + reverse			TALEN forward			Un-injected		
■ Alive	81	43	37	68	23	44	107	45	43
■ Dead	33	6	7	11	6	5	9	8	7
■ Wild type appearance	7	15	19	64	20	40	101	42	41
■ Reduction or loss of pigment/ eye	41	13	10	1	2	4	3	3	1
■ Failure to gastrulate	20	9	6	3	1	0	3	0	1
■ Reduction or loss of head	13	6	2	0	0	0	0	0	0

Figure 53, shows collated data from 3 sets of TALENS mRNA injections across 3 fertilisations using the same male and female *X. tropicalis* adults. (Bar chart & table) The data shows phenotypes and death rates of embryos injected with CIC TALENs RNA at concentrations of 1ng, 0.5ng and 0.25ng per embryo at the 2-cell stage. Phenotypes were recorded at the late tailbud stage (40). Injection of only the forward TALEN and un-injected embryos were used as controls. The image shows the range of embryos from the 1ng TALENs injections. Embryos ranged in severity from reduced pigmentation to total head loss. Arrow points to the enlarged proctodaeum.

Initial TALENs injections and targeting of the *CIC* allele exhibited phenotypes which ranged in severity (fig 53). Targeting displayed interruption of development at the anterior of the anterior-posterior axis, which increased in severity with increases of TALENs mRNA concentration. Embryos ranged from slight reduction in head/eye pigmentation to total loss of head structures (fig 53). In addition, most embryos had an enlarged proctodaeum. Phenotypes and death rates were recorded from the injections of TALENs to optimise further experiments (fig 53-54).

In injections of 1ng total mRNA of forward and reverse TALEN mRNA in the first fertilisation, there was a (33/81) 29% death rate, which was higher than that of TALEN forward (1ng total mRNA) only injected embryos at (11/68) 13.9% (fig 53). The sibling un-injected control embryos had a death rate of (9/107) 7.76%. (74/81) 91.8% of the 1ng total mRNA TALENs injected embryos produced a mutant phenotype with the remaining (7/81) 8.2% producing a wild type phenotype. (41/81) 50.8% of the embryos produced a phenotype which either had a reduction in eye size, loss of

pigmentation or had a total loss of one or more eye structure in the embryo. (13/81) 16.1% of embryos either had a reduction or total loss of head structure. (20/81) 24.8% of embryos failed to gastrulate appropriately.

As concentrations of TALENs mRNA were reduced to 0.5ng total mRNA, phenotypes became less prominent and death rates reduced (fig 53). In embryos injected forward and reverse TALEN mRNA in the second fertilisation, there was a (6/43) 12.3% death rate. The forward only TALEN injected control had a death rate of (6/23) 20.7% and the un-injected control embryos in the second fertilisation had a death rate of (8/45) 15.1%. Of the TALENs mRNA injected embryos, (15/43) 34.9% had a wild-type phenotype, whilst (28/43) 65.1% had a mutant phenotype. (13/43) 30.2% of embryos had a loss of eye structure or pigmentation, (9/43) 14% had a reduction or total loss of head structure and (9/43) 20.9% failed to gastrulate.

At the lowest concentration of injections at 0.25ng total mRNA of forward and reverse TALEN mRNA in the third fertilisation, there was a (7/37) 15.9% death rate (fig 53). Forward only TALEN injected controls had a death rate of (5/44) 10.2% and the un-injected control embryos in the third fertilisation had a death rate of (7/43) 14%. (19/37) 51.4% of embryos have a wild type phenotype, (18/37) 48.7% had a mutant phenotype. (10/37) 27% of embryos had a loss of eye structure or pigmentation, (6/37) 16.2% failed to gastrulate and (2/37) 5.4% had a reduction or total loss of head structure. Embryos were collected for targeting confirmation and sequencing data confirmed all embryos were targeted (fig 52).

Figure 54, (A-H) *X. tropicalis* embryos injected with 1ng (total) of CIC TALENs RNA (forward & reverse) at the 2-cell stage and (I) an un-injected wild-type embryo as a comparison (lateral view). (A-H) Phenotypes of embryos treated with CIC TALENs range in severity from total head loss to reduced pigmentation in the eye. Black arrows indicate enlarged proctodaeum observed in some of the embryos. (A-H) All embryos were DNA sequenced for confirmation of targeting by TALENs or for confirmation of (I) wild type sequences. Embryos were imaged, and

DNA extracted at the late tailbud stage (40-41). TALENs targeted sequencing reads show the introduction of indels (red rectangle) adjacent to the TALENS targeting site (black rectangle).
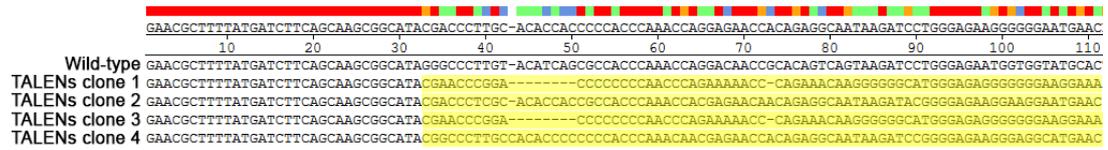
Figure 55, *X. tropicalis* DNA sequencing targeted by CIC TALENs and somatic mutations from 4 clones extracted from one embryo. All clones 1-4 show indels in comparison to the wild-type control embryo (yellow rectangle). In this batch of sequencing TALENs has 100% targeting penetrance.

Figure 54 shows the sequencing data for 8 of the embryos injected with TALENS and an additional wild type control. All TALENS injected embryos were confirmed to be targeted by DNA sequencing indicating a 100% efficiency. Sequencing also confirmed the wild-type embryo was not targeted.

Due to the mosaic nature of TALENs targeting and the introduction of multiple different mutations (indels), sequencing appears to be heterogeneous adjacent to the targeting region (fig 54, red boxes). The heterogeneous sequencing is caused by indels in the form of frame shifts, codon insertion, deletions within a single embryo. To get an understanding of the severity of phenotypes penetrance of targeting was measured by amplifying targeted regions by PCR and T-cloning into plasmid (fig 52 & 54, 55). Screening indicated a high penetrance of 80-100% in targeted embryos. As the severity of phenotype increases, so too should the penetrance of TALENS *CIC* targeting in the mosaic embryo. A range of severity can be seen for the injected embryos from reduction in eye pigmentation, cyclopia to loss of all head structure (fig 53-54).

5.2.7 Knockdown of the *CIC* isoforms leads to difference in phenotype

Whereas the TALENs system is reliant on the embryos translational machinery to produce the TALENs proteins from injected mRNA for *CIC* allele targeting and knockdown. The amount of time required to produce the TALENs protein, the efficiency of the TALENs protein and number of cell divisions which would take place before the TALENs mRNA is translated delays targeting, leading to mosaicism of daughter cells within the embryo. Previous expression profiles analysis (chapter 4) in *X. tropicalis* has revealed that CIC-L isoform is expressed in the maternal phase of development. Current genome editing technologies (TALENs or CRISPR/Cas9) would be unable to target the CIC-L isoform in the maternal phase of development. To overcome this problem, translation blocking antisense morpholino oligos were

designed which can target the maternally deposited mRNA in maternal phase as well as being used to target the zygotically expressed CIC-S isoform. Morpholinos were designed which would bind to sequence around the start codon of mRNA expressed for both the CIC-L and CIC-S isoforms.

Morpholinos were used to establish if there were difference in function between the two isoforms in early development. Morpholinos function by binding to a target sequence within an RNA, sterically blocking proteins that might otherwise interact with the RNA knocking down gene expression. *X. tropicalis* embryos were injected with 3 different combinations of morpholinos, CIC-L + CIC-S, CIC-L + Fluorescein or CIC-S + Fluorescein. An inert fluorescent fluorescein-labelled standard control morpholino would allow detection of successful delivery of morpholino within the cytosol of your cells (fig 57a-c). In addition, 3 different concentrations of morpholinos were injected in to the embryos at 40ng, 20ng or 10ng total morpholino per embryo (fig 56). For the highest concentration, CIC-L, CIC-S or fluorescein-tagged morpholinos were diluted from stock solutions to 10ng/ul in a 2nl volume which when micro-injected would be 20ng total per embryo (fig 56). Each combination of morpholino micro-injected combined gave a total concentration of 40ng total morpholino per embryo. Morpholinos were further diluted to 5ng/ul in a 2nl volume which when micro-injected would be 10 ng total per embryo and when combined with combined gave a total concentration of 20ng total morpholino per embryo. For the lowest concentration morpholinos were diluted to 2.5ng/ul in a 2nl volume which when micro-injected be 5ng total per embryo and when combined with combined gave a total concentration of 10ng total morpholino per embryo. When micro-injecting CIC-L and CIC-S morpholinos at the same time, no fluorescein labelled morpholino was added so that all morpholino micro-injection combination had the same concentration (40ng, 20ng and 10ng) (fig 57e).

	Morpholino	Wild type appearance	Responsiveness	Kinked back	Total embryos
40ng total injected	CIC-L + CIC-S	0	Severe delay in response and movement – twitch.	42	51
	CIC-L + control	0		39	45
	CIC-S + control	20	Severe delay in response and movement – twitch.	7	36
	Un-injected control	29			31

	Morpholino	Wild type appearance	Responsiveness	Kinked back	Total embryos
20ng total injected	CIC-L + CIC-S	9	Slight delay in some	5	32
	CIC-L + control	17		3	31
	CIC-S + control	18	Slight delay in some and decrease of movement	2	27
	Un-injected control	36			47

	Morpholino	Wild type appearance	Responsiveness	Kinked back	Total embryos
10ng total injection	CIC-L + CIC-S	16		0	17
	CIC-L + control	18		2	27
	CIC-S + control	25	Slight delay	3	39
	Un-injected control	55		2	61
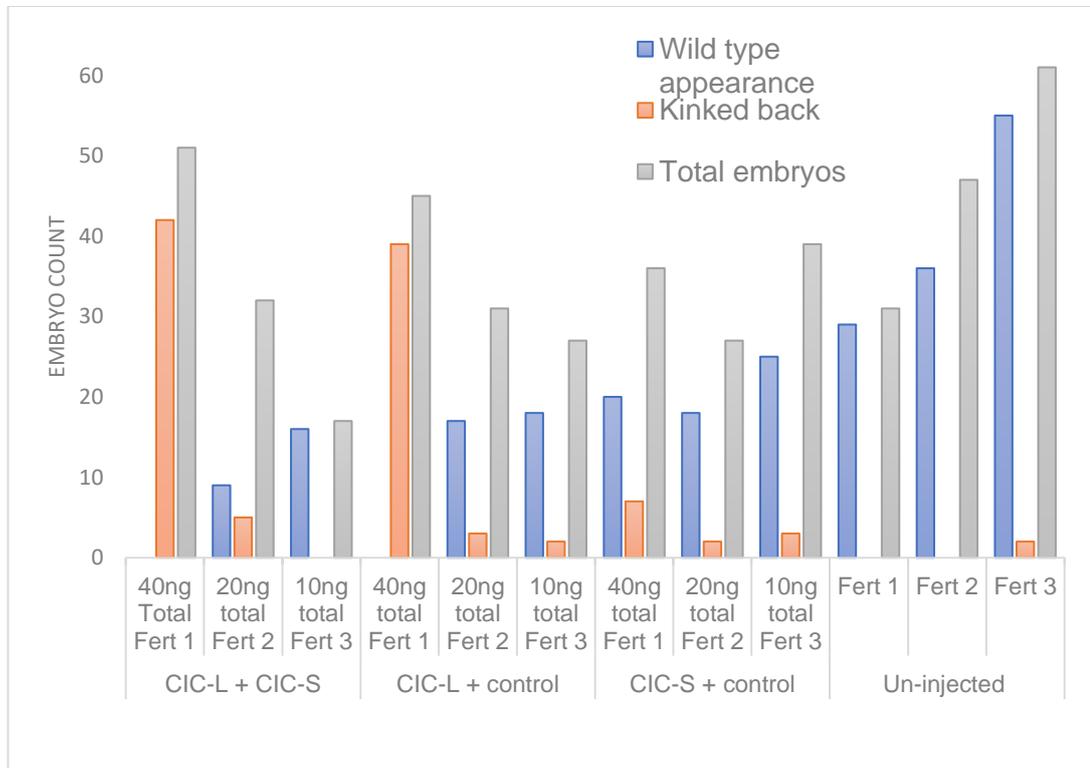
Figure 56, (Table) shows collated data from 3 sets of morpholino injections across 3 fertilisations using the same male and female *X. tropicalis* adults. The table represents the data collected from phenotypes and touch

responsiveness in embryos (stage 40) treated with CIC-L or CIC-S targeting morpholinos along with the fluorescein-tagged at varying concentrations of 40ng, 20ng and 10ng per embryo at the 2-cell stage. Un-injected wild-type embryos are used as a comparison. Phenotypes indicate dose dependant effect. As the CIC-S and CIC-L morpholino concentrations increased both together and individually so did the severity of the specific phenotypes. (Bar chart) The data from the table in bar chart form.

The combination of CIC-L and CIC-S targeting morpholino at 40ng total per embryo produced embryos with a phenotype that had a kinked back along the main body axis likely due to irregular formation of neural tube, somites, or notochord (fig 57e). (42/51) 82% of micro-injected with CIC-L and CIC-S targeting morpholinos embryos had a kinked back phenotype and no embryos had a wild-type appearance (fig 56). In addition, embryos had a severe delay touch response and suffered from prolonged twitching. Some embryos swam in circles and appeared paralyzed when prodded with a pair of forceps.

Injections of the CIC-L and fluorescein control morpholinos injected at 40ng total per embryo produced embryos with kinked backs, but were responsive to touch and no prolonged twitching, unlike those seen in embryos injected with both CIC-L and CIC-S targeting embryos (fig 56 & 57g). Again, embryos appeared to swim in circles.

Injections of the CIC-S and fluorescein control morpholinos at 40ng total per embryo produced embryos which the slight majority (20/36) 55.56% had a wild-type appearance with the remaining embryos being either slightly shorter in length along the anterior posterior axis (fig 57f). Embryos slight reduction in pigmentation/eye size or suffering from oedema in the abdominal cavity (fig 57f). These embryos had a severe delay in touch response and suffered from prolonged twitching and paralysis previously seen in embryos injected with a combination of CIC-L and CIC-S targeting morpholino. This suggests that the severe delay in touch response and prolonged twitching was due to the CIC-S targeting. Although the kinked back phenotype was present in embryos injected with CIC-S it was less severe in CIC-S (7/36)19.44% than in CIC-L (42/51) 86.67% injected embryos (fig 56).

CIC-L and CIC-S targeting morpholino injected at 20ng total per embryo produced fewer kinked back embryos at (5/32) 15.63% of the siblings injected compared to the higher concentrations of CIC-L and CIC-S morpholinos at (42/51) 82% (fig 56). These embryos had a slight delay in responsiveness seen in higher concentrations (40ng total). Embryos injected at this concentration had higher numbers of wild-type embryos (9/32) (28.13%). CIC-L and fluorescein control morpholinos injected at 20ng total per embryo had fewer kinked back embryos (3/31) (9.68%) and an increase in

numbers of embryos with a wild-type phenotype (17/31) (54.84%). Embryos were as responsive to touch as the wild-type control embryos. Injections of the CIC-S and fluorescein control morpholinos injected at 20ng total per embryo produced kinked backed embryos at (2/27) 7.41% and (18/27) 66.67% wild-type embryos. Embryos displayed slight delay to touch response and decrease of movement.

CIC-L and CIC-S targeting morpholino injected at 10ng total per embryo produced embryos which were (16/17) 94.12% wildtype (fig 56). None of these embryos appeared to have the kinked back phenotype previously seen at higher concentrations. The embryos injected with 20ng CIC-L and fluorescein control morpholinos total per embryo lead to (2/27) 7.41% of embryos which had the kinked back phenotype, (18/27) 66.67% had a wild-type phenotype. The embryos injected with 20ng CIC-S and fluorescein control morpholinos produced embryos which had slight delays in touch response. (25/39) 64.1% of embryos had a wild-type appearance and (3/39) 7.69% had the kinked back phenotype.
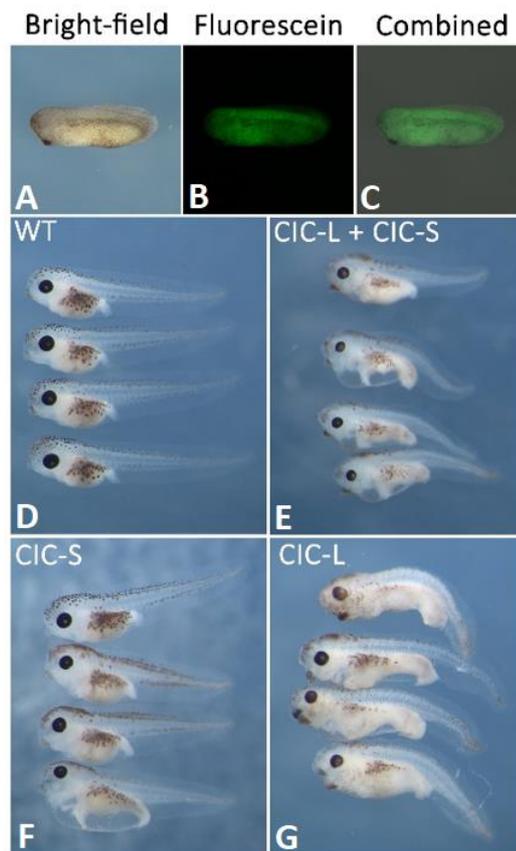
Figure 57, (A-C) shows lateral view of an embryo injected with the fluorescein morpholino control. (E-G) lateral view of morpholino phenotypes after injecting embryos at the 2-cell stage with 40ng of the CIC-L, CIC-S & fluorescein morpholinos (late-tailbud- stage 40). (F) CIC-S morpholino injections produced embryos which looked like the wild type embryos but lacked responsiveness. (G) CIC-L morpholino injections produced embryos with

kinked backs. (E) Combined CIC-L and CIC-S injections had a combination of both phenotypes. (D) Wild type embryos.

5.3 Discussion

5.3.1 Knockdown of *CIC* expression results in a similar phenotype to FGF overexpression

If CIC functions downstream of FGF transduction, acting to repress expression of FGF target genes, relief of CIC repression should result in an increase of expression of a subset of FGF regulated genes. If FGF regulates CIC repressional activity it should lead to reduction of the CIC protein within the nucleus, leading to inhibition of repression likely by MAPK transduction utilised by other RTKs (Tor or EGFR) (fig in introduction) (Astigarraga et al., 2007a; Goff et al., 2001; Jimenez et al., 2000; Klingler et al., 1988; Paroush et al., 1997; Roch et al., 2002; Suisse et al., 2017b).

CIC knockdown by TALENs (fig 53-54) remarkably resembles phenotypes seen in overexpression of FGF4 and FGF8 in early *Xenopus* development (Amaya et al., 1991; Amaya et al., 1993; Fletcher et al., 2006; Isaacs et al., 1994). In keeping with the hypothesis that CIC acts as a transcriptional repressor downstream of the FGF pathway, embryos show a posteriorised phenotype, with a truncation of the anterior of the anterior-posterior axis (fig 53-54) (Isaacs et al., 1994). The embryos range from having a loss or reduction of pigmentation to total loss of the eye (fig 53-54). In more severe examples of TALEN injections, embryos have a reduction or total loss of head structures (forebrain) (fig 54). In addition, embryos appear to have an enlarged proctodaeum at the posterior region seen previously in FGF4 and FGF8 overexpression (fig 53-54) (Isaacs et al., 1994). The range of phenotypes in the posteriorised embryos is likely due to mosaicism of the TALENs targeting.
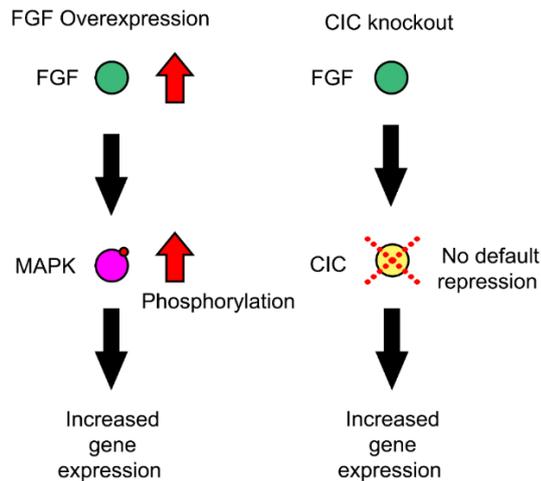
Figure 58, a diagram of the FGF pathway and hypothesis of the FGF-CIC pathway. When FGF is overexpressed, it leads to increases in activation of MAPK, which in turn increases expression of FGF gene targets. When *CIC* is knocked down by TALENs targeting, despite no increase in MAPK transduction a reduction of the CIC protein hinders repressional activity, preventing default repression increasing expression of target genes.

If FGF regulated genes which are normally repressed in a default state by CIC without FGF MAPK signalling and the default repression is removed, FGF regulated genes should be over expressed (fig 58). Studies have shown mis-regulation of the anterior-posterior axis in *Xenopus* by the overexpression of *FGF4*, either by injection of plasmid, mRNA or protein applied to beads leads to increased expression of the Hox genes, *HoxA7*, *HoxB9* and *HoxC6* during gastrula and neural stages of development (Pownall et al., 1996a). Hox genes are a subset of homeobox genes which regulate and direct the development of the anatomical structures by specifying the positional identity of cells (Taniguchi, 2014). First discovered in *D. melanogaster*, Hox genes are known regulators of anterior-posterior specification establishing a body plan during development and are found throughout the animal kingdom (Holland and Garcia-Fernandez, 1996). FGF expression during mesoderm induction plays significant role in the normal anterior-posterior axis during development by the activation of a number of Hox genes (Cho and De Robertis, 1990).

As previously mentioned, FGF expression plays an important role in anterior-posterior formation and is finally balanced. Overexpression of FGF4 and FGF8 and knockdown of CIC are not the only cause of the posteriorisation phenotype seen *Xenopus*. Studies have shown that overexpression of *Xcad3* leads to the loss of anterior structures of the anterior-posterior axis, resulting in loss head structures, including the hindbrain (Isaacs et al., 1998). *Xcad3* overexpression lead to the upregulation of *HoxA7* and *HoxB9*. Both *HoxA7* and *HoxB9* are known to be restricted in their spatial expression to the trunk of the embryo, but when Xcad3 is upregulated both genes

have expanded spatial expression in the anterior axis. In addition, *HoxB1* and *HoxB3* were found to be down regulated in embryos with *Xcad3* overexpression (Isaacs et al., 1998).

The posterisation phenotype of *CIC* knockdown suggests that FGF4 and FGF8 are involved in the same pathway during embryonic development (fig 9 & 58), but further evidence will be required to confirm the relationship. This data still confirms that CIC has a conserved role in the anterior-posterior development in vertebrates. Previous experiments highlight CICs role in the anterior-posterior patterning of the developing *D. melanogaster* embryo. These new findings suggest that CIC appears to retain a conserved role in anterior- posterior formation in vertebrate embryonic development (Goff et al., 2001; Jimenez et al., 2000; Paroush et al., 1997).

5.3.2 Knockdown of *CIC* isoforms produce alternative phenotypes.

While TALENs knockdown of CIC causes posterisation of the *X. tropicalis* embryo, targeting of the *CIC* expression with morpholinos causes a range of different effects. Targeting of CIC-L expression with morpholinos lead to embryos developing a bend along the main body axis with tails facing downwards (stage 40) (fig 57e & g). Embryos were responsive to touch, although, the bent axis caused embryos to swimming in an irregular circular motion. The cause of the CIC-L knockdown phenotype was likely due to be the irregular formation of neural tube, somites, or notochord, but further analysis will be required to confirm this. In *D. melanogaster* development we know that CIC is required for the normal development of the central nervous system (Ajuria et al., 2011; Astigarraga et al., 2007a; Jimenez et al., 2000). We know that CIC acts to repress *ind* expression, which operates downstream of EGFR in the developing *D. melanogaster* embryo (Ajuria et al., 2011). Unlike in *D. melanogaster* which contains only one gene copy of *ind*, the *X. tropicalis* and *H. sapiens* homologs, known as *Genomic Screened Homeobox* (*Gsx*) contain copes 2 of the gene (*Gsx1* and *Gsx2*).

Unlike the phenotype of CIC-L knockdown by morpholino, embryos which have CIC-S knockdown appear to have a wild-type appearance but with a reduction or delay in responsiveness to touch with prolonged twitching. The severity of the phenotype increased when morpholino concentrations were increased. Effects of this phenotype are likely to be caused by neuro-muscular defects, but further analysis would be required.

Studies have shown interruption of the ATAXN1-CIC protein complex lead to abnormal neuro-muscular disease in *H. sapiens* (Lu et al., 2017b). One such condition, the inherited neuromuscular disease SCA1 leads to patients suffering from problems with balance and co-ordination, muscle spasms loss of sensation in the hands and feet (peripheral neuropathy) and muscle stiffness (spasticity). An interruption of the ATAXN1-CIC protein complex could explain the defective neuromuscular CIC-S phenotype, but an important consideration would be that SCA1 itself is caused by the polyglutamine (polyQ) tract expansion of ATAXN1, which binds to the wild-type CIC protein, rather than a mutant form of CIC (Brusse et al., 2006). Although we know interruption of ATAXN1-CIC protein complex can cause neuromuscular symptoms, and cancers such as ODG have highlighted CIC is critical for central nervous system development (Tanaka et al., 2017).

A study in 2006 to screen 202 novel genes expressed during gastrula stages of *X. tropicalis* development were knocked down using morpholinos. Genes were selected by the criteria that they were conserved in *X. tropicalis* and in mammals. Morpholinos were designed to target sequences around the initiating AUG codons of these 202 genes. Of the 202 genes targeted, 59% had abnormal had a mutant phenotype during development of the embryo. The recorded knockdown phenotypes were characterised into "synphenotype groups", these are groups of phenotypes which show similar loss-of-function (Rana et al., 2006). This was done to identify genes which may be involved in the same developmental process. Although gene knockdown phenotypes may share the same synphenotype group, they may not share the same pattern of expression or even function in the same developmental process. This is known as "synexpression group" (Niehrs and Pollet, 1999).

Of the 202 genes, 61 genes were observed to have a bent axis by the tailbud stage of development much like that seen in CIC-L morpholino targeting (fig 57e & g). 5 sub-groups were created which ranged in synphenotype from having a short dorsalised body with upturned tail to having a normal body with a wavy tail (Rana et al., 2006). Of the 5 synphenotype sub-groups created, one group, like that of the morpholino CIC-L targeting phenotype, had 13 genes which when targeted with morpholinos produced a kinked back phenotype (bent axis/bent down tail). The morpholino targeted genes in this bent axis synphenotype group were 14-3-3β.2, CPSF2, EDIL3, Lefty-b, heatr5b (TEgg078i21), calhm2 (TGas083e14.2), c1orf131 (TGas141c24), myo18b (TNeu053k08.2), pced1a (TNeu062k05.2), znf674 (TGas106k21.2), PAR6A, Smad10 and Wnt5b.2 (Rana et al., 2006). Interestingly, of the genes in the bent axis synphenotype group, the 14-3-3 protein is known to

regulate CIC transcriptional activity through MAPK transduction. When the 14-3-3 motif is phosphorylated by MAPK transduction at Ser^{173} in CIC which is adjacent to the HMG-box, it recruits the 14-3-3 protein reducing optimal binding ability of HMG-box to bind to its octameric targets (Dissanayake et al., 2011).

In addition, another synphenotype group was created which contained a group 14 genes which when targeted by morpholinos were thought to produce a phenotype with motility defects (Rana et al., 2006). This group was classified to have embryos which appeared to have normal development but whose motility was abnormal, significantly reduced or absent. Like the CIC-S morpholino targeted phenotype embryos in this group displayed delayed touch response, paralysis and circular swimming. The morpholino targeted genes in this motility defects synphenotype group were CC1, HMG17, HoxC8, REEP4, Tinp1, VHLH, AuroraA, FrzA, myo18b (TNeu053k08), uba5 (TNeu098a04), arfgap2 (TEgg043a17) and baz1b (TEgg058h11) (Rana et al., 2006).

When the CIC-L and CIC-S morpholinos were co-injected the embryos looked to have a combination of both phenotypes seen in the separate (CIC-L/CIC-S) injections of morpholinos, such as the bend along the main body axis with tails facing downwards (fig 56e), reduction in eye size, unresponsiveness and prolonged twitching. It is important to consider that co-injections did not reproduce the headless phenotypes previously seen in the TALENs CIC knockdown experiments. The design of the morpholino experiment had the benefit of enabling targeting of the individual prominent isoforms of *CIC*, CIC-L and CIC-S. Although specific targeting of the isoforms could give insight into their function and potential pathways that they are involved in, due to the large number of exons (22) and the complexity that this brings, there may be other unidentified isoforms which are expressed. Given the fact that this project has identified 2 alternative isoforms, if there are other isoforms which are not being targeted by the morpholinos, there is the potential that they could compensate for knockdown of the prominent isoforms by morpholinos. Although this remains to be speculation, it is still an important consideration for these experiments and would require further analysis and could explain why targeting of CIC-L and CIC-S does not reproduce the same phenotype as the CIC HMG-box targeting TALENs.

5.3.3 Overexpression of the CIC-S homologs does not produce a mutant phenotype

CIC knockdown in *X. tropicalis* has revealed a clear and distinct posterisation phenotype which resembles that of FGF4 or FGF8 overexpression (fig 54). Whilst overexpression of the CIC-S homologs of *H. sapiens*, *M. musculus* and *X. tropicalis* homologs shows no change of phenotype in comparison to wild-type sibling embryos. Given the evidence that CIC requires to bind and form a complex with co-repressors in vertebrates (ATAXN1) and invertebrates (Groucho) to function, overexpression of *CIC* may require the overexpression of a binding co-repressor. Without the upregulation of a co-repessor CIC would be unable to upregulate repressional activity, therefore no change in phenotype would be observed. In *D. melangastor,* the Groucho protein (Gro) is known as a global co-rerpesssor that binds to different families of DNA-binding repressors, functioning by binding to promotors preventing transcription (Cinnamon and Paroush, 2008; Jennings and Ish-Horowicz, 2008; Jimenez et al., 2000; Paroush et al., 1997). Gro acts as auxiliary factor interacting with CIC to form a repressional complex by binding to the N2 domain (Sen and Baltimore, 1986). In contrast to the findings in D. melangastor, mammalian CIC homologs require the formation of a protein complex with ataxin 1 (ATXN1) or ataxin1-like (ATXN1L) by binding of the ATXN1 domain in CIC protein (Lam et al., 2006). No evidence has been found to suggest that this is the same mechanism involved in CIC repressional activity in *D. melangastor* or vice versa gro functioning with mammalian CIC.

Recently new research has been found which shows other DNA-binding co-repressors are required in the CIC-Gro repressional complex. The Nuclear Factor-kappaB (NF-kB) protein was first identified as a DNA-binding protein which specifically binds to the 10 bp kB site of the immunoglobulin k light-chain enhancer of B lymphocytes (Sen and Baltimore, 1986), and has been shown to play a role in many other cell types. NF-kB belongs to a family of proteins which share 300 amino encoding for the Rel homology domain (RHD) (Baeuerle, 1991). Members of the NF-kB family are involved in cellular defence mechanisms and differentiation and these include p50 (NF-kB1), p52 (NF-kB2), p65 (RelA), c-Rel, v-Rrel, RelB, and the *D. melengaster* proteins Dorsal and Dif. A study found that CIC binds and targets Toll/interleukin-1 receptor (TIR) homology domain such as such as *zerknullt* (*zen*), *tolloid* (*tld*), and *decapentaplegic* (*dpp*) downstream targets in the presence of the Dorsal protein.

Dorsal becomes internalised into the nucleus upon activation of TIR. TIR signalling leads to phosphorylation-dependent disassembly of cytoplasmic Cactus (Cact)–Dorsal complex allowing Dorsal to become internalised into the nucleus (fig 59) (Belvin et al., 1995). The internalisation of Dorsal leads a nuclear gradient with high concentrations at ventral regions and progressively lower concentrations at lateral and dorsal regions of the developing *D. melanogaster* (Reeves and Stathopoulos, 2009; Stein and Stevens, 2014). Dorsal requires conserved A/T-rich sites situated close to Dorsal binding sites located in enhancers of target genes, near to CIC octameric DNA binding (T(G/C)AATG(A/G)A) sites.
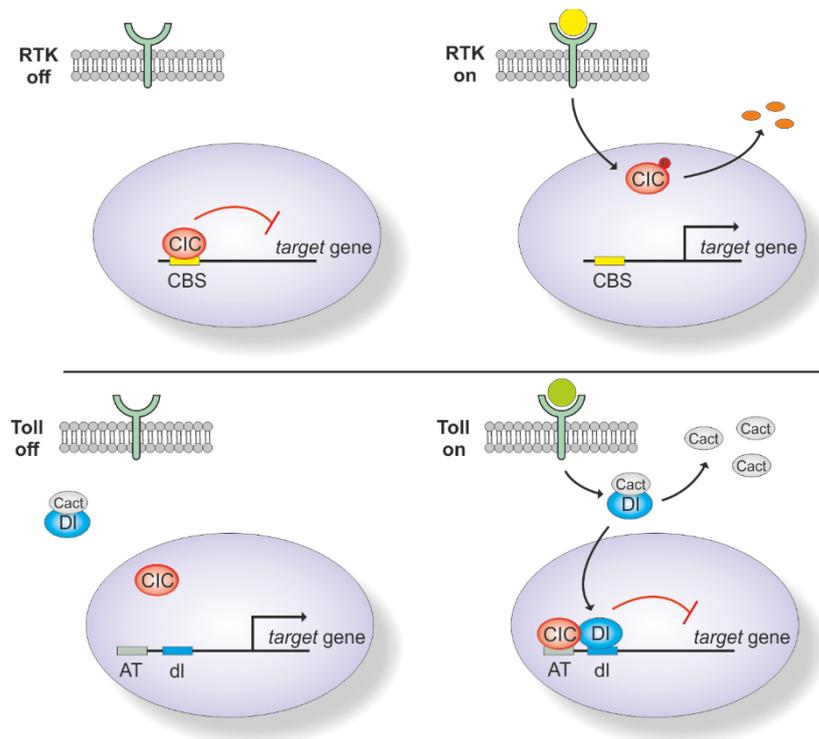
Figure 59, The Toll-CIC signalling pathway. When no RTK signal is present CIC can act as a transcriptional repressor in a 'default state', bound to CIC binding sites (CBS). Upon activation of the RTKs, MAPK transduction of the RTK causes CIC to become phosphorylated leading to relief of repression. In contrast, when Toll signalling is absent, the (Cact)–Dorsal complex remains intact and CIC is unable to bind suboptimal AT sites. Upon activation of Toll signalling the Dorsal is released from (Cact)–Dorsal complex and becomes internalised into the nucleus and enables CIC to bind suboptimal AT sites adjacent to Dorsal binding sites (Papagianni et al., 2018).

Dorsal allows CIC to bind to suboptimal (low-affinity) DNA sites which would not be recognised without Dorsal (fig 59) (Papagianni et al., 2018). When dorsal was not present, CIC was unable to bind to *zen* or *tld* (Sen and Baltimore, 1986). This type of specificity gives CIC another layer of complexity for its transcriptional repressional activity other than the well-established mode of regulation by RTK-MAPK pathways (Dissanayake et al., 2011; Jimenez et al., 2000). CIC binds and represses specific

targets in the cell dependent on which co-repressor partner is present (Papagianni et al., 2018). This suggests another mechanism by which CIC repressional activity is regulated. Overexpression of CIC in *D. melanogaster* leads to mild phenotypical changes such as mild ommatitdial disorganisation, ommatitidium being the units of the *Drosophila* compound eye and loss of veins at the posterior region of the wing. Interestingly the same study found that despite overexpression of the CIC leading to loss of wing veins, the overall integrity of the wing structure was improved (Lam et al., 2006).

5.3.4 Co-injection of FGF leads to reduction of the CIC-S protein within the embryo.

Co-injections of *H. sapiens* eGFP-CIC-S with FGF4 and FGF8 show changes of concentration and localisation of eGFP-CIC-S within the *X. tropicalis* embryo (fig 50-51). Although the florescence microscopy data (fig 50O,P,R,S) appears to contradict the western blot data (fig 51) showing that there is a GFP signalling when looking for florescence GFP signal in the living X. tropicalis embryos. The reason for this is likely due to the increased sensitivity of florescence microscopy in the detection of the GFP signal in comparison to the western blot GFP antibody. Fluorescence microscopy of FGF/GFP-CIC co-injected revealed that the translated GFP-CIC protein did not appear to be concentrated within nucleus in tissue outside of presumptive animal hemisphere region of the exogastrulated embryos (fig 49p & s). Exogastrulation is the interruption of the normal morphogenetic movements of cells and cell layers of the embryo during gastrulation. In Xenopus, cells that represent the mesoderm and endoderm would normally become internalised by several signalling pathways including FGF signalling which regulate of convergent extension (Chung et al., 2005; Tada and Heisenberg, 2012). When the morphogenetic movements are interrupted by the overexpression of FGF it leads to the interruption of the convergent extension pathway, leading to mesoderm and endoderm remaining on outside of the embryo having no contact with the exoderm (Holtfreter, 1933). Cell cycle arrest and or cell death due to overexpression of FGFs could be the cause, but further analyse will be required to establish the mechanism for the redistribution of CIC. Raf is a protein which sits upstream MAPK in the FGF transduction pathway. Studies in *X. laevis* have shown when high concentrations of *Raf* mRNA are introduced into the embryo, it leads to cell cycle arrest leading to embryonic lethality due to exogastrulation (MacNicol et al., 1995). This is likely to be because of inappropriate activation MAPK transduction by Raf-1 (MacNicol et al., 1995). A MAPK activation/expression is a fine

balance which could be easily interrupted. If FGF is shown to be directly involved with the regulation of CIC it would be interesting to assess if MAPK is able to bind and target CIC before entry in to the nucleus.

Western blot analysis of the co-injected embryos clearly indicate that concentrations of CIC are reduced which suggests there is an interaction with CIC either directly or indirectly (fig 51). FGF8 has a greater potency than FGF4, FGF8 co-injection leads to total loss/degradation of the GFP-CIC protein whilst FGF4 co-injections lead to the appearance of a truncated GFP-tagged protein product at 85 kDa. This likely due to a post-translational modification of the CIC protein although the mechanism is unknown. We know that EGFR signalling leads to the redistribution of the nuclear CIC protein from the nucleus into the cytoplasm in the context of the *D. melanogaster* ovarian follicle cells (Goff et al., 2001)

Although we know EGFR can lead to translocation of CIC into the cytoplasm from the nucleus, both EGFR and Tor MAPK transduction in the imaginal wing discs of *D. melanogaster* lead to the degradation of the CIC protein (Ajuria et al., 2011; Roch et al., 2002; Tseng et al., 2007). MAPK phosphorylation targets the CIC protein for ubiquitylation and later proteasomal degradation (Suisse et al., 2017b). Although fluorescence microscopy may suggest that CIC is redistributed from the nucleus to the cytoplasm in the early embryo outside of the animal hemisphere tissue, western blot suggests that CIC protein becomes degraded. This confirms that FGF signalling does influence the abundance and sub-cellular localisation of the CIC-S protein within the *X. tropicalis* embryo.

Chapter 6: Transcriptomic analysis of *FGF4* overexpression and *CIC* knockdown in *X. tropicalis* development

6.1 Introduction

Research in this thesis indicates *CIC* is a transcription factor downstream of FGF signalling. CIC is expressed in the same spatial and temporal regions as FGFs during early development suggesting that they come into contact (chapter 4). In addition, treating CIC with FGF4/FGF8 in the developing *X. tropicalis* embryo leads to the degradation of the CIC protein (chapter 5). The phenotype produced by knockdown of *CIC* in the *X. tropicalis* embryo by TALENs resembles that of phenotype seen in FGF overexpression (chapter 5) (Isaacs et al., 1994). Although evidence points to CIC being downstream of FGF pathway it may only be circumstantial, evidence is still required to establish a direct link between CIC and FGF signalling. One of the ways in which direct evidence could be found is by identifying commonly regulated genes targets between CIC and FGF signalling. FGF4 overexpression leads to upregulation of FGF target genes (Branney et al., 2009) and if CIC acts as an intermediate transcription factor downstream of the FGF pathway, CIC knockdown should lead to the same upregulation of FGF transcriptomic changes in a subset of FGF target genes. CIC could be the unknown intermediate liable repressor postulated by Fisher (2002) to function downstream FGF transduction.

The recently developed RNA-seq technique and next-generation sequencing (Illumina) will be used to establish if there is relationship between FGF and CIC by identifying common transcriptional changes between sibling embryos which have either been targeted for *CIC* knockdown or have *FGF4* overexpressed. Sibling *X. tropicalis* embryos will either be micro-injected with pCSK-eFGF (FGF4) (Isaacs et al., 1994) or CIC targeting TALENs (chapter 5). The transcriptional changes of both sample sets will be normalised against wild-type water injected embryo samples.

This study will provide insight in to how FGF regulates gene transcription in the *X. tropicalis* and given the conserved nature of *CIC*, in other vertebrate animal models. If CIC is established to function downstream of the FGF signalling pathway, it would add an additional RTK to the list of known RTKs that regulate *CIC* through MAPK

transduction. This could suggest CIC functions promiscuously with other RTKs that utilise MAPK transduction in different tissue or cell contexts.

This chapter will address the following aims which are to:

- Establish if CIC operates downstream of the FGF pathway using RNA-seq and next-generation sequencing.
- Establish which downstream targets genes of FGF signalling CIC targets by repression.

6.2 Results

Sibling embryos were micro-injected at 1-2 cell stages with pCSK-eFGF, mRNA of *CIC* targeting TALENS or water. Plasmid based overexpression of *FGF4* rather than mRNA overexpression was used to allow zygotic expression of *FGF4* offering a valid comparison with the TALEN knockdown which only affects zygotic expression. Embryo samples were collected at stage 14 in batches of 10. TALENs *CIC* targeting was confirmed by sequencing. In addition, photographs of phenotypes were recorded for all sample batches prior to mRNA/DNA extraction. Total mRNA was checked for quality control using the 2100 Bioanalyzer. 3 biological replicates batches were analysed for each sample type, CIC knockdown, eFGF overexpression and water injected embryos creating 9 Illumina next-generation sequencing libraries. Samples were done in triplicate to allow appropriate statistical analysis. 9 libraries on a single Illumina sequencing lane allowed ≥ 40 million reads per sample. Results were produced by aligning the raw reads for each sample to the *X. tropicalis* reference transcriptome (v9.1 of the genome) with Salmon (http://salmon.readthedocs.io), producing an estimated read count for transcripts for each sample. Q values and effect size were calculated using the estimated read counts using Sleuth (http://pachterlab.github.io/sleuth/).

6.2.1 Changes in gene expression resulting from *CIC* inhibition and *FGF4* overexpression

43,558 transcripts were analysed for 23,635 annotated genes, measuring transcripts per million (TPM). TPM values are a method of normalising the raw read counts which can be affected by factors such as sequencing biases, read lengths and total number of reads (Conesa et al., 2016). The TPM value is a measure of abundance of transcripts in each sample. The TPM value represents the number of transcripts observed if one million transcripts from the whole transcriptome were counted

(Conesa et al., 2016). 618 transcripts were found to have significant changes in expression using the statistical Q value of ≤ 0.06. The Q value is an adjusted P value that allows for false positives and is used as a measure of statistical significance of differential expression (Storey and Tibshirani, 2003). The Q value of ≤ 0.06 would allow for 6% false positives meaning that of the 618 genes analysed 37 would be false positives. 416 (67.3%) of the 618 transcripts were found to be upregulated using the statistical effect size of >1.2 (fig 60 & table 4) which would indicate a significant increase in expression and the remaining 202 (32.7%) transcripts were found to be downregulated using the effect size of < 0.8 (fig 61 & table 5) indicating a significant decrease in expression. Any transcripts which had an effect size value between 0.8 and 1.2 would be considered to not be significantly upregulated or downregulated.
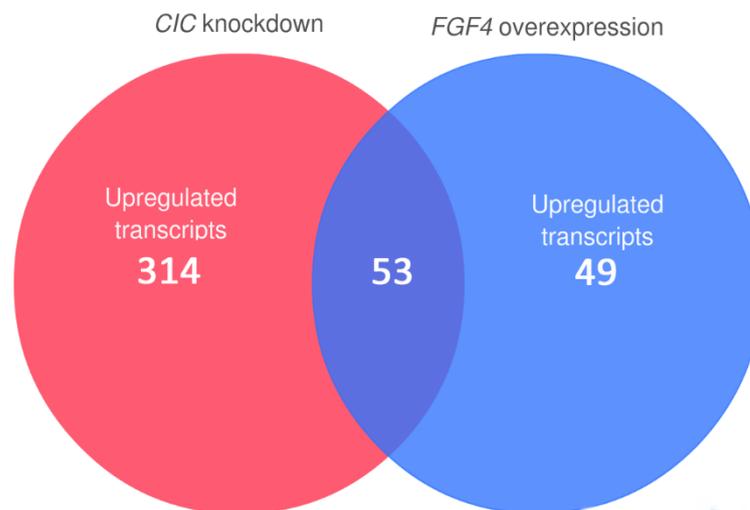
Figure 60, the Venn diagram of upregulated genes for *CIC* knockdown and *FGF4* overexpression created from RNA-seq data using the statistical Q value of ≤ 0.06 and effect size of ≥ 1.2. 314 transcripts were upregulated in CIC knockdown and 48 transcripts were upregulated in the FGF4 expression. 53 transcripts were found to be upregulated both knockdown of *CIC* and overexpression of *FGF4*.

Of the total 416 transcripts which were upregulated, 314 (75.5%) of those were found to be upregulated specifically when *CIC* was knocked down (fig 60). 49 (11.8%) transcripts were found to be specifically upregulated when FGF4 was overexpressed. The remaining 53 (12.7%) transcripts were found to have overlapping upregulated expression in both *CIC* knockdown and *FGF4* overexpression. Of the 53 overlapping transcripts which were upregulated in both *CIC* knockdown and *FGF4* overexpression, 9 were unannotated transcripts from the *X. tropicalis* reference transcriptome (v9.1 of the genome) (table 4).

Gene/Locus id	Effect size in CIC knockdown	Effect size in FGF4 overexpression	Q value in CIC knockdown	Q-value in FGF4 overexpression	Xenbase Gene ID
LOC101731310	28.5305756	11.5266165	7.19E-05	0.02520134	
bmp7.2	25.2288141	15.0711704	2.16E-05	0.00144765	XB-GENE-855954
Sept2	25.205563	29.2370423	0.04664137	0.05606717	XB-GENE-480201
htr1b	24.1132058	7.74610398	5.14E-06	0.02716225	XB-GENE-988006
fgd3	20.3978228	20.212536	0.02390131	0.04470355	XB-GENE-995891
tmcc1	14.5872784	11.3618342	0.00531905	0.03185826	XB-GENE-5881367
LOC100485153	11.7873895	24.588171	0.04122085	0.00369167	XB-GENE-5965587
mmp1	8.19095565	4.65359449	2.65E-04	0.04253908	XB-GENE-485009
exoc3l1	5.81320415	9.96754289	0.00537694	8.94E-05	XB-GENE-6258780
apold1	4.69028423	5.44816231	0.00447528	0.00202074	XB-GENE-5814316
LOC101732940	4.61418571	3.24325078	5.75E-07	7.21E-04	
Frzb	4.49221318	3.20793272	8.35E-04	0.0446586	XB-GENE-481353
Fos	4.43048696	5.38117736	1.42E-12	7.15E-17	XB-GENE-866811
usp2	3.47775889	2.22497282	1.74E-05	0.0446586	XB-GENE-962796
fosl1	3.0906611	2.99609256	0.00531905	0.01541677	XB-GENE-6257957
ier3	2.62726252	2.33077077	4.83E-05	0.00129991	XB-GENE-5884223
rasl11b	2.59528505	1.98702446	3.45E-07	0.00144765	XB-GENE-491056
arrdc2	2.55508345	2.47186556	5.75E-07	3.06E-06	XB-GENE-946088
gpcpd1	2.49121285	2.45243676	0.00748687	0.01815647	XB-GENE-967385
sgk1	2.45400097	2.65673011	2.26E-04	7.13E-05	XB-GENE-1003518

Junb	2.43654277	2.038884	0.00207761	0.0553722	XB-GENE-945864
riok3	2.30934499	1.55128754	1.86E-07	0.05777269	XB-GENE-971902
LOC100486038	2.23504154	2.65740808	0.01944072	0.00308028	
cbx4	2.21646702	2.87099344	0.02778942	0.00129991	XB-GENE-986716
LOC101733948	2.21612572	2.67505682	0.00849379	7.16E-04	
b4galt1.1	2.21408599	2.4369297	0.02982526	0.01800023	XB-GENE-5871724
c4bpa	2.17436868	2.62680095	0.04092973	0.0075817	XB-GENE-1014581
nfkbiz	2.12618753	2.01723886	0.00609923	0.02733025	XB-GENE-981412
LOC105947461	2.11043456	2.44356055	0.00115911	7.13E-05	
arrdc2	2.05151718	2.4395286	2.17E-04	2.50E-06	XB-GENE-946088
sgk1	2.03864276	2.50510045	0.00158382	1.86E-05	XB-GENE-1003518
fat1	1.9642374	1.63165863	1.41E-04	0.03552574	XB-GENE-920279
LOC105945708	1.96219871	2.74501717	0.04793613	5.16E-04	
atf3	1.96068697	1.75737319	3.91E-04	0.01364873	XB-GENE-6085251
fam83c	1.89317727	2.11820891	0.00655736	0.00106863	XB-GENE-5953445
cldn6.1	1.84627752	1.78367668	0.0034821	0.01541677	XB-GENE-959181
wnt8a	1.83394387	2.20096083	0.04092973	0.00357804	XB-GENE-493706
chic1	1.82839031	1.67206119	0.00303852	0.04111942	XB-GENE-6039630
fam83c	1.79889328	1.93048057	0.03986846	0.02567513	XB-GENE-5953445
LOC101730746	1.79691185	1.98596143	0.04149332	0.01711502	
sat1	1.72794386	1.85810879	0.03240249	0.01742614	XB-GENE-1002781
mst1	1.68170416	1.89125821	0.04025172	0.00890713	XB-GENE-487985

sox17b.2	1.65879165	1.53188698	6.00E-04	0.0160725	XB-GENE-495335
tsc22d3	1.64757661	1.53669973	0.00727572	0.05993999	XB-GENE-1004589
sat1	1.64409085	1.59429617	1.44E-04	0.00100716	XB-GENE-1002781
pnpla3	1.64330681	1.6004365	0.00159287	0.00786306	XB-GENE-5855123
tmcc1	1.5922001	1.62244506	0.01423816	0.01815647	XB-GENE-5881367
gadd45a	1.57948454	1.50609716	0.00345639	0.02705165	XB-GENE-482257
prkd1	1.57663995	1.40215039	3.18E-04	0.04253908	XB-GENE-487626
tsc22d3	1.5509114	1.42398264	6.31E-04	0.02716225	XB-GENE-1004589
cdc14b	1.4682714	1.46709574	0.02720604	0.05146224	XB-GENE-977314
nuak2	1.45738152	1.6730974	0.03641397	0.00124724	XB-GENE-5857794
LOC100127682	1.38157643	1.41023856	0.03859384	0.04125888	

Table 4, shows the RNA-seq data of overlapping upregulated genes of *CIC* knockdown and *FGF4* overexpression. Genes are ranked in order of effect size of *CIC* knockdown.
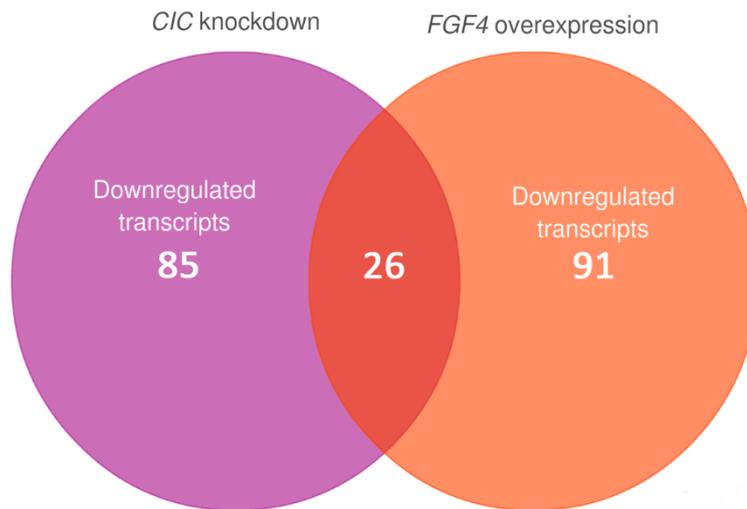
Figure 61, the Venn diagram of downregulated genes for *CIC* knockdown and *FGF4* overexpression created from RNA-seq data using the statistical Q value of ≤ 0.06 and effect size of ≤ 0.8. 85 transcripts were upregulated in *CIC* knockdown and 91 transcripts were upregulated in the *FGF4* expression. 26 transcripts were found to be downregulated in both knockdown of *CIC* and overexpression of *FGF4*.

Of the 202 transcripts which were found to be downregulated in the RNA-seq analysis, 85 (42.1%) of the transcripts were found to be downregulated specifically when CIC was knocked down (fig 61). 91 (45.1%) of the transcripts were found to be downregulated specifically when FGF4 was overexpressed and the remaining 26 (12.9%) transcripts were found to be have overlapping downregulated expression in both CIC knockdown and FGF4 overexpression (fig 61). 2 of the 26 transcripts which were found to be downregulated in both CIC knockdown and FGF4 overexpression were found to be unannotated in the reference transcriptome (table 5).

Gene/Locus id	Effect size in CIC knockdown	Effect size in FGF4 overexpression	Q-value in CIC knockdown	Q-value in FGF4 overexpression	Xenbase Gene ID
atp2a2	0.74134727	0.68014408	0.04727284	0.0055786	XB-GENE-1012933
sox11	0.72311495	0.73109298	0.02093286	0.05317234	XB-GENE-483418
s1pr5	0.70207191	0.6913191	0.01613349	0.01989754	XB-GENE-5902381
efnb3	0.68241848	0.67841232	0.01060003	0.01742614	XB-GENE-5902269
LOC100158459	0.67975584	0.56527671	0.05323558	7.84E-04	
serpina1	0.66968602	0.66836844	0.01637609	0.029958	XB-GENE-5804366
celsr2	0.65255064	0.67688529	0.00400162	0.02508506	XB-GENE-919957
ripply2.2	0.63869511	0.6707801	0.00392118	0.03028268	XB-GENE-6539633
pkdcc.2	0.61940826	0.57379925	0.05798883	0.02733025	XB-GENE-941501
notch3	0.61658192	0.70122949	6.86E-05	0.02304013	XB-GENE-6450008
LOC100485697	0.60328469	0.6695121	9.88E-04	0.04182617	
irx3	0.59869973	0.71728262	1.71E-05	0.04189998	XB-GENE-480486
dpysl3	0.58830679	0.60933431	0.01090688	0.04253908	XB-GENE-944729
slc23a2	0.58601389	0.45933212	0.0107265	2.41E-05	XB-GENE-966414
axl	0.5669826	0.57109684	0.01389473	0.02993852	XB-GENE-922445
nkain1	0.55385555	0.5452458	0.02720604	0.04111942	XB-GENE-946558

cygb	0.54503804	0.48108769	0.02435031	0.0048875	XB-GENE-987724
msi1	0.50282806	0.58732033	1.82E-04	0.01989754	XB-GENE-490596
znf219	0.49320622	0.58336698	6.31E-04	0.0446586	XB-GENE-1001695
foxi4.1	0.47408734	0.50427205	0.01423816	0.05845581	XB-GENE-5996107
cebpa	0.36285755	0.27682834	0.01681196	0.00124122	XB-GENE-853397
pax6	0.31575659	0.27341949	0.0176532	0.00927766	XB-GENE-484088
pax6	0.29963222	0.27350506	0.03339538	0.03028268	XB-GENE-484088
tmem119	0.29299271	0.30980779	0.01830742	0.05317234	XB-GENE-5796050
unc13d	0.20529008	0.2693114	2.98E-04	0.01286832	XB-GENE-1008072
spib	0.18295904	0.13705067	0.0052548	8.37E-04	XB-GENE-479615

Table 5, shows the RNA-seq data of overlapping downregulated genes of *CIC* knockdown and *FGF* overexpression. Genes are ranked in order of effect size of *CIC* knockdown.

This data suggest that CIC operates as a transcriptional repressor downstream of FGF transduction. This data also highlights that CIC sits upstream of a number of important FGF gene targets.

6.2.2 Gene ontology analysis and enrichment

The Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification tool (Mi et al., 2013) was utilised to establish the gene ontology (GO) of the overlapping upregulated genes resulting from *CIC* knockdown and *FGF4* overexpression (table 4). This information would provide a functional profile of the upregulated transcripts. The information provided by GO analysis would be used give understanding as to what function and role CIC plays downstream of FGF signalling. By establishing a functional profile from the overlapping upregulated gene set (table 4), it would provide a greater understanding of the undelaying biological process of *CIC* repression downstream of FGF signalling.

6.2.3 Pathway analysis of upregulated overlapping genes

Genes that were found to have overlapping upregulated expression in both *CIC* knockdown and *FGF4* overexpression were entered into the PANTHER classification system to establish which biological process, molecular function and pathways they were associated with (fig 62-64, & table 6). Both *X. tropicalis* and *H. sapiens* species established pathways were analysed.

Upregulated gene	Pathway	Gene Identifier
fos *atf3*	Apoptosis signalling pathway	*X. tropicalis* - XB-GENE-866811 *X. tropicalis* - XB-GENE-6085251
fos *junb* *bmp7.2* *atf3*	Gonadotropin releasing hormone receptor pathway	*X. tropicalis* - XB-GENE-866811 *X. tropicalis* - XB-GENE-945864 *X. tropicalis* - XB-GENE-855954 *X. tropicalis* - XB-GENE-6085251
mmp1	Plasminogen activating cascade	*X. tropicalis* - XB-GENE-485009
lims1	Integrin signalling pathway	*H. sapiens* - HGNC=6616
fos	Huntington disease	*X. tropicalis* - XB-GENE-866811
lims1 *rasl11b*	Integrin signalling pathway FGF pathway	*X. tropicalis* - XB-GENE-967923
junb *bmp7.2* *fosl1* *rasl11b*	TGF-beta signalling pathway	*X. tropicalis* - XB-GENE-945864 *X. tropicalis* - XB-GENE-855954 *X. tropicalis* - XB-GENE-6257957
sept2	Parkinson disease	*H. sapiens* - HGNC=7729
wnt8a *mmp1*	Alzheimer disease-presenilin pathway	*X. tropicalis* - XB-GENE-493706 *X. tropicalis* - XB-GENE-485009
fos	T cell activation	*X. tropicalis* - XB-GENE-866811
gadd45a	PI3 kinase pathway	*X. tropicalis* - XB-GENE-482257

htr1b	5HT1 type receptor mediated signalling pathway	*X. tropicalis* - XB-GENE-988006
fos	PDGF signalling pathway	*X. tropicalis* - XB-GENE-866811
prkd1 *fos* *frzb*	Angiogenesis	*X. tropicalis* - XB-GENE-487626 *X. tropicalis* - XB-GENE-866811 *X. tropicalis* - XB-GENE-481353
gadd45a	p53 pathway	*X. tropicalis* - XB-GENE-482257
prkd1	EGF receptor signalling pathway	*X. tropicalis* - XB-GENE-487626
junb	Inflammation mediated by chemokine and cytokine signalling pathway	*X. tropicalis* - XB-GENE-945864
wnt8a *frzb* *fat1*	Wnt signalling pathway	*X. tropicalis* - XB-GENE-493706 *X. tropicalis* - XB-GENE-481353 *X. tropicalis* - XB-GENE-920279
gadd45a	p38 MAPK pathway	*X. tropicalis* - XB-GENE-482257
prkd1	VEGF signalling pathway	*X. tropicalis* - XB-GENE-487626
wnt8a *fat1*	Cadherin signalling pathway	*X. tropicalis* - XB-GENE-493706 *X. tropicalis* - XB-GENE-920279
prkd1 *fos*	CCKR signalling map	*X. tropicalis* - XB-GENE-487626 *X. tropicalis* - XB-GENE-866811
fos	B cell activation	*X. tropicalis* - XB-GENE-866811
htr1b	Heterotrimeric G-protein signalling pathway-Gi alpha and Gs alpha mediated pathway	*X. tropicalis* - XB-GENE-988006
fos	Interleukin signalling pathway	*X. tropicalis* - XB-GENE-866811
fos	Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade	*X. tropicalis* - XB-GENE-866811

Table 6, genes from the upregulated group associated with different pathways discovered by the PATNHER GO pathway tool for *H. sapiens* and *X. tropicalis*. (Red text) Genes added to pathways and not found in the Panther database from unpublished data.

The PANTHER GO tool revealed 15 of the overlapping upregulated genes were associated with 26 pathways (table 6). The gonadotropin releasing hormone (GnRH) receptor pathway (*fos*, *junb*, *bmp7.2* & *atf3*) , TGF-β signalling pathway (*junb, bmp7.2, fosl1* & *rasl11b*), Angiogenesis (*prkd1, fos* & *frzb*), Wnt signalling pathway (*wnt8a, frzb* & *fat1*), Alzheimer disease-presenilin pathway (*wnt8a* & *mmp1*), Cadherin signalling pathway (*wnt8a* & *fat1*) and CCKR signalling map (*prkd1* & *fos*) were found to contain multiple upregulated genes (fig 60 & table 4, 6). *fos* was found to associated with 10 of the pathways, more than any other of the upregulated genes (table 6). Other genes such as *wnt8a*, *gadd45a*, *prkd1*, *frzb, junb* and *mmp1* were found in multiple pathways (table 6).

A Molecular function

B Molecular function - Binding

C Molecular function – catalytic activity

Transporter activity, 1

Binding, 13

Structural molecule activity, 3

Signal transducer activity, 2

Receptor activity, 2

Channel regulator activity, 1

Catalytic activity, 10

Lipid binding, 1

Protein binding, 7

Nucleic acid binding, 4

Enzyme regulator activity , 1

Transferase activity, 5

Hydrolase activity, 4

Figure 62, pie charts of the categorised molecular functions from the overlapping upregulated genes of *CIC* knockdown and *FGF4* overexpression using the PANTHER GO tools. (A) 33 genes where found to fit into 32 functional categories (binding, catalytic, channel regulator activity, receptor activity, signal transducer activity, structural molecule activity and transporter activity). (B) The largest molecular functional group, binding, was sub-categorised to establish the type of binding the overlapping upregulated genes were involved with. 13 genes were placed into 12 functional categories within protein binding, lipid binding and nucleic acid binding. (C) The catalytic activity sub-group contained 10 genes in 10 functional categories within transferase activity, enzyme regulator activity and hydrolase activity sub-groups.

A Biological process

- Response to stimulus , 8
- Biological adhesion , 1
- Biological regulation , 9
- Multicellular organismal process , 3
- Cellular component organization or biogenesis 4
- Metabolic process , 15
- Localisation , 2
- Developmental process , 5
- Cellular process , 20

B Biological process - Cellular process

- Cytokinesis, 1
- Cell proliferation, 1
- Cell cycle, 2
- Cell communication, 7

C Biological process - Metabolic process

- Primary metabolic process, 6
- Biosynthetic process, 6
- Catabolic process, 3
- Phosphate-containing compound metabolic process, 8
- Nitrogen compound metabolic process, 9

D Biological process – Biological regulation

- Regulation of molecular function, 1
- Homeostatic process, 1
- Regulation of biological process, 8

E Biological process – Response to stimulus

- Response to stress , 1
- Response to abiotic stimulus.
- Response to external stimulus , 1
- Response to endogenous stimulus , 1
- Response to biotic stimulus.

F Biological process – Developmental process

- Nervous system…
- Anatomical structure morphogenesis, 1
- Mesoderm development, 1
- Embryo development, 1
- Ectoderm development, 1
- Death, 2
- Cell differentiation, 3

Figure 63, pie charts of the biological processes from the overlapping upregulated genes of *CIC* knockdown and *FGF4* overexpression using the PANTHER GO tools. (A) 33 genes where found to fit into 67 process categories within biological adhesion, biological regulation, cellular component organisation/biogenesis, cellular process, developmental process, localisation, metabolic

process, multicellular organismal process or response to stimulus sub-groups. (B) The largest biological process sub-group, cellular processes, was sub-categorised to establish the type of cellular processes the overlapping upregulated genes were involved in. 20 genes were placed into a further 11 sub-categories (cytokinesis, cell proliferation, cell cycle & cell communication). (C) The metabolic processes sub-group which was found to contain 15 genes in 32 further process sub-categories (primary metabolic process, biosynthetic process, catabolic process, nitrogen compound metabolic process & phosphate containing compound metabolic process). (D) The biological regulation sub-category contained 9 genes in 32 further process sub-categories split into regulation of molecular function, homeostatic process & regulation of biological processes sub-groups. (E) The response to stimulus sub-category contained 9 genes in 32 further process sub-categories within response to stress, response to abiotic stimulus, response to biotic stimulus, response to endogenous stimulus, response to external stimulus sub-groups. (F) The developmental processes sub-category contained 5 genes in 10 further process sub-categories within nervous system development, anatomical structure morphogenesis, cell differentiation, death, ectoderm development, embryo development & mesoderm development sub-groups.

Overlapping upregulated group associated pathways
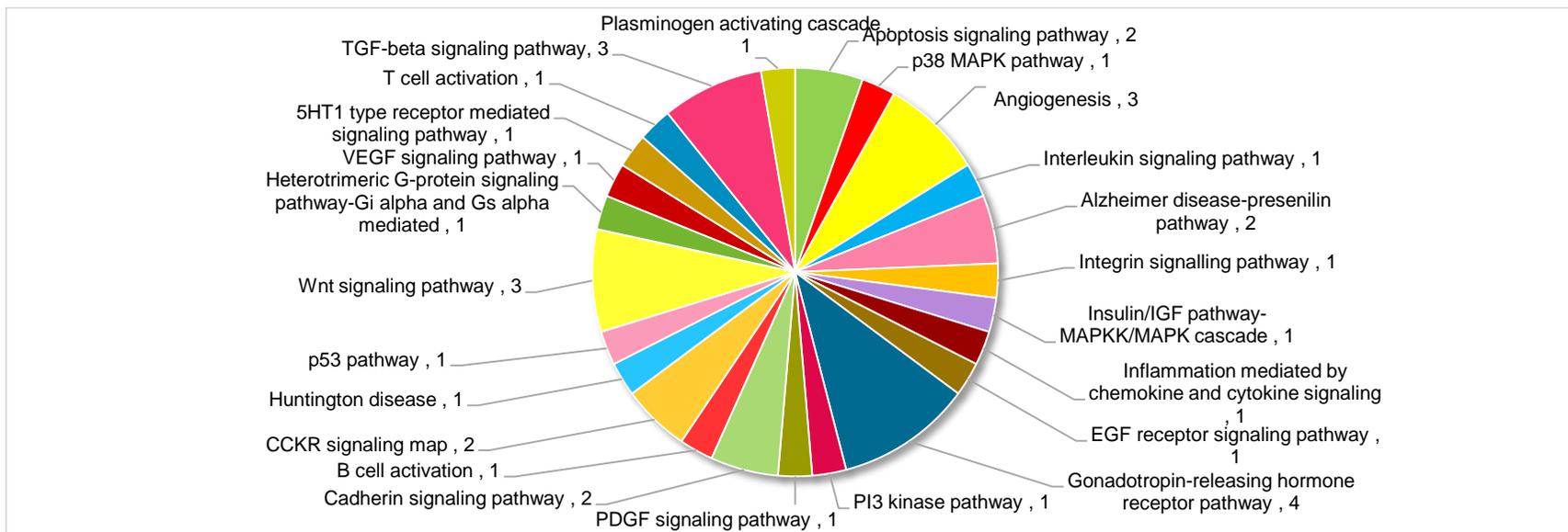
Figure 64, a pie chart showing the 37 pathway hits associated from 33 genes from the upregulated group using the PATNHER GO pathway tool for *X. tropicalis*.

6.2.4 The ontology of upregulated overlapping genes in *CIC* knockdown and *FGF4* overexpression.

Due to the limitations of the annotation of PANTHER *X. tropicalis* database, the Xenbase (v4.2) GO tools was also used to establish GO. The Xenbase GO terms associated with *X. tropicalis* genes were applied to the upregulated overlapping gene set of *CIC* knockdown and *FGF4* overexpression (table 7). The Xenbase GO tools were utilised to establish the molecular function, biological process and area in the cell that the translated protein would function, providing more detailed information regarding the *X. tropicalis* genome. This information would provide information for the undelaying biological process of *CIC* repression downstream of FGF signalling.

Gene	GO Term	
apold1	Molecular function	• Lipid binding
	Biological Process	• Lipid transport
		• Lipoprotein metabolic process
	Cellular component	• Extracellular region
		• Integral component of membrane
arrdc2	Biological Process	• Signal transduction
aft3	Molecular function	• Dna binding
		• Dna-binding transcription factor activity
		• Rna polymerase ii proximal promoter sequence-specific dna binding
		• Sequence-specific dna binding
		• Transcription factor activity, rna polymerase ii proximal promoter sequence-specific dna binding
	Biological Process	• Regulation of transcription by rna polymerase ii response to virus
	Cellular component	• Nucleus
b4galt1.1	Molecular function	• Transferase activity, transferring glycosyl groups
	Biological Process	• Carbohydrate metabolic process
	Cellular component	• Integral component of membrane
bmp7.2	Molecular function	• Bmp receptor binding
		• Cytokine activity

		• Growth factor activity
		• Transforming growth factor beta receptor binding
	Biological Process	• Bmp signalling pathway
		• Smad protein signal transduction
		• Cartilage development
		• Cell development
		• Cell differentiation
		• Growth
		• Ossification
		• Positive regulation of pathway-restricted smad protein phosphorylation
		• Regulation of MAPK cascade
		• Regulation of apoptotic process
	Cellular component	• Extracellular region
		• Extracellular space
cbx4	Cellular component	• Nucleus
cdc14b	Molecular Function	• Protein serine/threonine phosphatase activity
		• Protein tyrosine phosphatase activity
		• Protein tyrosine/serine/threonine phosphatase activity
	Biological Process	• Cell division
		• Cilium assembly
		• Mitotic cell cycle arrest
		• Mitotic spindle midzone assembly
		• Regulation of exit from mitosis
	Cellular component	• Centrosome
		• Condensed chromosome
		• Cytoplasm
		• Integral component of membrane
		• Mitotic spindle
		• Nucleolus
		• Nucleus
		• Spindle pole
cldn6.1	Molecular function	• Structural molecule activity
	Cellular function	• Bicellular tight junction
		• Integral component of membrane

		•	Plasma membrane
exoc3l1	Molecular function	•	Snare binding
	Biological Process	•	Exocyst localisation
		•	Exocytosis
	Cellular component	•	Exocyst
fat1	Molecular function	•	Calcium ion binding
	Biological Process	•	Homophilic cell adhesion via plasma membrane adhesion molecules
	Cellular component	•	Integral component of membrane
		•	Plasma membrane
fgd3	Molecular function	•	Rho guanyl-nucleotide exchange factor activity
		•	Metal ion binding
	Biological process	•	Regulation of rho protein signal transduction
fos	Molecular function	•	Dna-binding
		•	Dna-binding transcription factor activity
		•	Double-stranded dna binding
		•	Sequence-specific dna binding
	Biological process	•	Multicellular organismal response to stress
		•	Regulation of transcription by rna polymerase ii
	Cellular component	•	Transcription factor complex
fosl1	Molecular function	•	Dna-binding
		•	Dna-binding transcription factor activity
		•	Sequence-specific dna binding
	Biological process	•	Regulation of transcription by rna polymerase ii
	Cellular component	•	Nucleus
frzb	Molecular function	•	G-protein coupled receptor activity
		•	Wnt-activated receptor activity
		•	Wnt-protein binding
	Biological process	•	Wnt signalling pathway
		•	Canonical wnt signalling pathway
		•	Multicellular organism development
		•	Non-canonical wnt signalling pathway

		• Positive regulation of wnt signalling pathway by establishment of wnt protein localization to extracellular region
	Cellular component	• Extracellular region • Extracellular space • Integral component of membrane
gadd45a	Biological process	• Activation of MAPKKK activity • Regulation of cell cycle • Response to stress
	Cellular component	• Cytoplasm • Nucleus
gpcpd1	Molecular function	• Glycerophosphocholine phosphodiesterase activity • Phosphoric diester hydrolase activity • Starch binding
	Biological process	• Glycerophospholipid catabolic process • Lipid metabolic process
htr1b	Molecular function	• G-protein coupled serotonin receptor activity • Neurotransmitter receptor activity • Serotonin binding
	Biological process	• Adenylate cyclase-inhibiting g-protein coupled receptor signalling pathway • Bone remodelling • Chemical synaptic transmission • Feeding behaviour • Phospholipase c-activating g-protein coupled receptor signalling pathway • Regulation of behaviour • Release of sequestered calcium ion into cytosol • Response to drug • Vasoconstriction
	Cellular component	• Dendrite • Integral component of plasma membrane
junb	Molecular function	• Dna-binding • Dna-binding transcription factor activity

		• Dna-binding transcription factor activity, rna polymerase ii-specific
		• Rna polymerase ii proximal promoter sequence-specific dna-binding
		• Sequence-specific dna-binding
		• Transcription coactivator activity
		• Transcription factor binding
	Biological process	• Cellular response to hormone stimulus
		• Negative regulation of transcription by rna polymerase ii
		• Positive regulation of cell differentiation
		• Positive regulation of transcription by rna polymerase ii
		• Regulation of cell cycle
		• Regulation of cell death
		• Regulation of cell proliferation
		• Regulation of transcription by rna polymerase ii
		• Response to camp
		• Response to cytokine
		• Response to drug
		• Response to lipopolysaccharide
		• Response to mechanical stimulus
		• Response to radiation
		• Transcription, dna-templated
	Cellular component	• Nuclear chromatin
		• Transcription factor complex
lims1	Molecular function	• Metal ion binding
		• Zinc ion binding
mmp1	Molecular function	• Calcium ion binding
		• Metalloendopeptidase activity
		• Zinc ion binding
	Cellular component	• Collagen trimer
		• Extracellular matrix
mst1	Molecular function	• Serine-type endopeptidase activity
	Cellular component	• Extracellular space
nfkbiz	Molecular function	• Transcription coregulator activity

	Biological process	• Regulation of transcription by rna polymerase ii
	Cellular component	• Nucleus
nuak2	Molecular function	• Atp-binding • Protein kinase activity • Protein serine/threonine kinase activity
	Biological process	• Cellular response to glucose starvation • Intracellular signal transduction • Protein phosphorylation
	Cellular component	• Cytoplasm • Nucleus
pnpla3	Molecular function	• Triglyceride lipase activity
	Biological process	• Lipid catabolic process • Lipid homeostasis • Liver development • Positive regulation of triglyceride catabolic process • Triglyceride catabolic process
	Cellular component	• Cytoplasm • Lipid droplet • Membrane
prkd1	Molecular function	• Atp-binding • Metal ion binding • Protein kinase c activity
	Biological process	• Ntracellular signal transduction • Protein kinase d signalling • Regulation of angiogenesis • Regulation of lymphangiogenesis
	Cellular component	• Cytoplasm • Intracellular
rasl11b	Molecular function	• Gtp binding • Gtpase activity
	Biological process	• Mesendoderm development • Signal transduction • Small gtpase mediated signal transduction
	Cellular component	• Intracellular • Membrane

riok3	Molecular function	• Atp-binding • Metal ion binding • Protein serine/threonine kinase activity
sat1	Molecular function	• Diamine n-acetyltransferase activity • Spermidine binding
	Biological process	• Spermidine acetylation • Spermine catabolic process
	Cellular component	• Cytosol
sept2	Molecular function	• Gtp-binding
	Biological process	• Cell cycle • Cell division • Cilium assembly • Mitotic nuclear division • Smoothened signalling pathway
	Cellular component	• Cell cortex • Ciliary membrane • Cleavage furrow • Cytoplasm • Midbody • Spindle
sgk1	Molecular function	• Atp-binding • Calcium channel regulator activity • Chloride channel regulator activity • Potassium channel regulator activity • Protein serine/threonine kinase activity • Sodium channel regulator activity
	Biological process	• Apoptotic process • Cellular sodium ion homeostasis • Inflammatory response • Intracellular signal transduction • Neuron projection morphogenesis • Peptidyl-serine phosphorylation • Positive regulation of sodium ion transport • Positive regulation of transporter activity • Regulation of apoptotic process • Regulation of cell growth • Regulation of cell proliferation

	Cellular component	• Cytoplasm • Endoplasmic reticulum • Nucleus
sox17b.2	Molecular function	• Dna-binding
	Biological process	• Wnt signalling pathway • Gastrulation • Regulation of transcription, dna-templated • Transcription, dna-templated
	Cellular component	• Nucleus
tmcc1	Cellular component	• Integral component of membrane
tsc22d3	Molecular function	• Dna-binding transcription factor activity
	Biological process	• Negative regulation of activation-induced cell death of t-cells
	Cellular component	• Cytoplasm • Nucleus
usp2	Molecular function	• Thiol-dependent ubiquitinyl hydrolase activity
	Biological process	• Protein deubiquitination • Ubiquitin-dependent protein catabolic process
wnt8a	Molecular function	• Frizzled binding • Protease binding
	Biological process	• Spemann organizer formation • Wnt signalling pathway • Canonical wnt signalling pathway • Canonical wnt signalling pathway involved in neural crest cell differentiation • Cell fate commitment • Embryonic axis specification • Negative regulation of cardiac cell fate specification • Neural crest cell fate commitment • Neuron differentiation • Positive regulation of dna-binding transcription factor activity • Regulation of transcription involved in anterior/posterior axis specification

Cellular component	• Extracellular region	
	• Extracellular space	
	• Proteinaceous extracellular matrix	

Table 7, Gene ontology and GO annotations from overlapping upregulated transcripts from *CIC* knockdown and *FGF4* overexpression taken from the Xenbase database (http://www.xenbase.org).

6.2.5 Pathway analysis of downregulated overlapping genes

Genes that were found to have overlapping downregulated levels of expression in both *CIC* knockdown and *FGF4* overexpression were entered into the PANTHER classification system to establish which biological process, molecular function and pathways they were associated with (fig 65-67 & table 8). Both *X. tropicalis* and *H. sapiens* species established pathways were analysed.

Downregulated gene	Pathway	Gene Identifier
dpysl3	Pyrimidine Metabolism	*X. tropicalis* - XB-GENE-944729
celsr2	Cadherin signalling pathway	*X. tropicalis* - XB-GENE-919957
celsr2	Wnt signalling pathway	*X. tropicalis* - XB-GENE-919957
unc13d	Synaptic vesicle trafficking	*H. sapiens* - HGNC- 23147
spib	Interleukin signalling pathway	*X. tropicalis* - XB-GENE-479615
serpina1	Blood coagulation	*X. tropicalis* - XB-GENE-5804366
dpysl3	Axon guidance mediated by semaphorins	*X. tropicalis* - XB-GENE-944729
notch3	Notch signalling pathway	*H. sapiens* - HGNC=7883
notch3	Alzheimer disease-presenilin pathway	*H. sapiens* - HGNC=7883

Table 8, genes from the overlapping downregulated gene set associated with different pathways discovered by the PATNHER GO pathway tool for *H. sapiens* and *X. tropicalis*.

The PANTHER GO tools revealed 9 pathways (7 in *X. tropicalis*) were found to be involved in the 7 of the overlapping genes which were downregulated in *ClC* knockdown and FGF4 overexpression (fig 61 & table 8). Unlike the upregulated group of transcripts, multiple transcripts were not found in a single pathway. *Celsr2 (*cadherin signalling pathway & wnt signalling pathway)*, dpysl3* (axon guidance mediated by semaphorins & pyrimidine Metabolism) and *notch3* (notch signalling pathway & alzheimer disease-presenilin pathway) were found to be involved in multiple pathways (table 8).

A   Molecular function

Transmembrane transporter activity, 2
Signal transducer activity, 1
Binding, 5
Receptor activity, 2
Catalytic activity - Hydrolase activity, 2

B   Molecular function - Binding

Protein binding, 1
Calcium ion binding, 1
Nucleic acid binding, 3
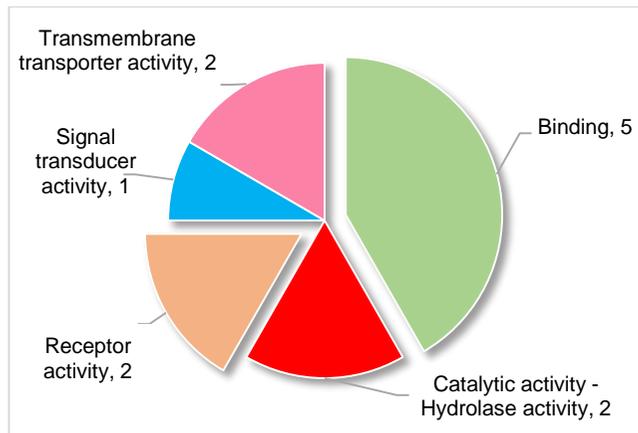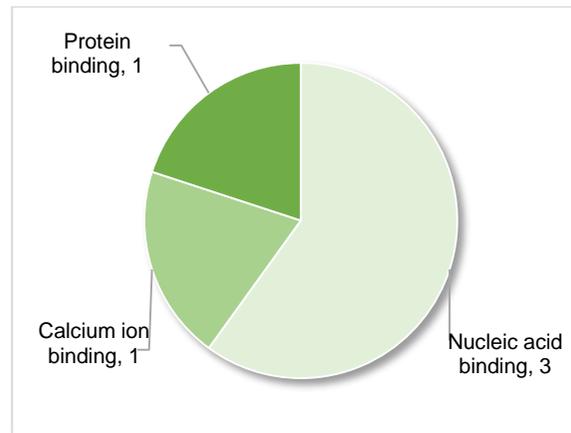
C   Molecular function - Receptor activity

Transmembrane receptor protein tyrosine kinase activity, 1
G-protein coupled receptor activity, 1
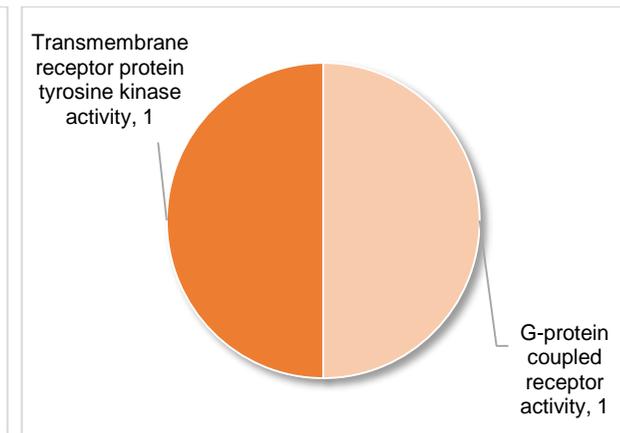
Figure 65, pie charts of the molecular function from the overlapping downregulated genes of *CIC* knockdown and *FGF4* overexpression using PANTHER GO tools. (A) 17 genes where found to fit into 12 functional categories within transmembrane transporter activity, binding, catalytic activity hydrolase activity, receptor activity & signal transducer activity sub-groups. (B) The largest molecular functional sub-group, binding, was further sub-categorised to establish the type of binding the upregulated genes were involved with. 5 genes were placed into 5 functional categories within protein binding, nucleic acid binding & calcium ion binding sub-groups. (C) The receptor activity sub-group contained 2 genes in 2 functional categories within the transmembrane receptor protein tyrosine kinase activity & G-protein coupled receptor activity sub-groups.

A Biological process

B Biological process - Metabolic process

C Biological process - Biological regulation

D Biological process - Developmental process
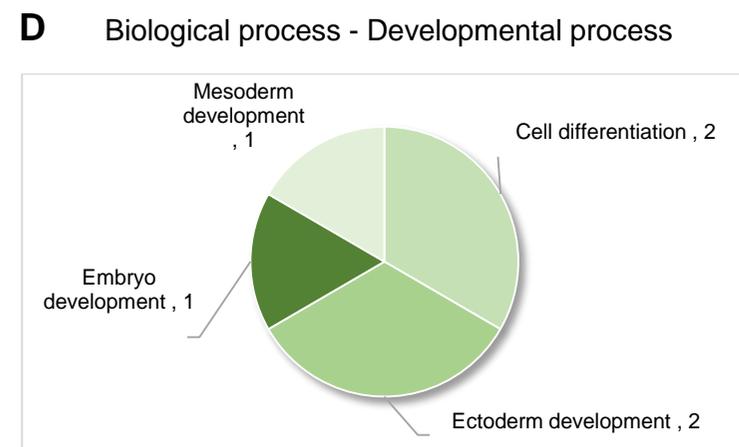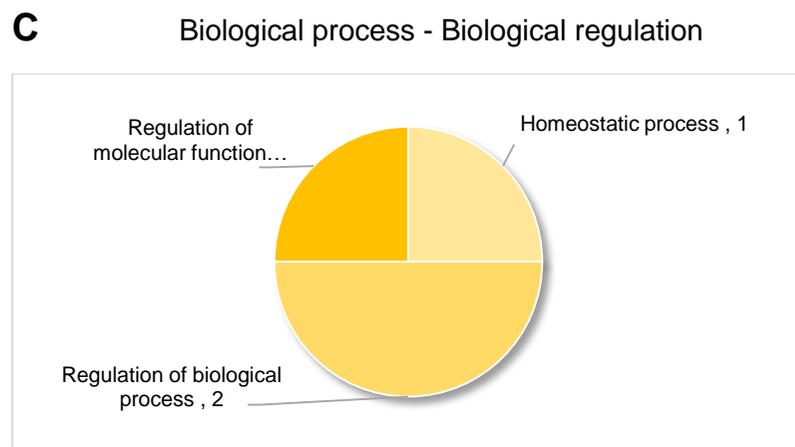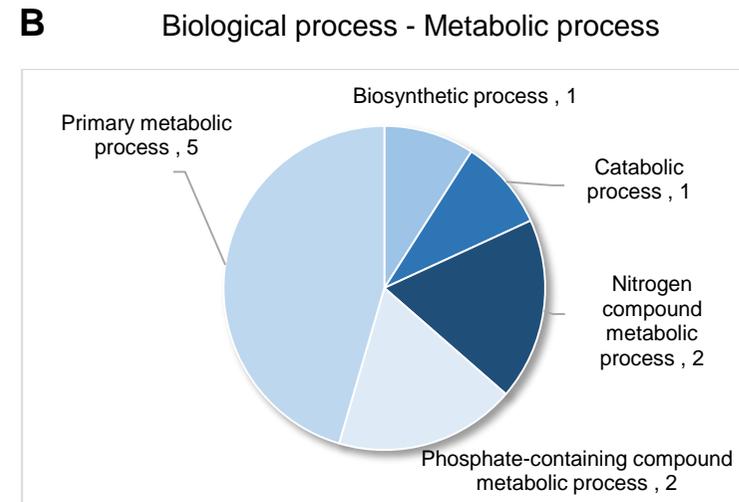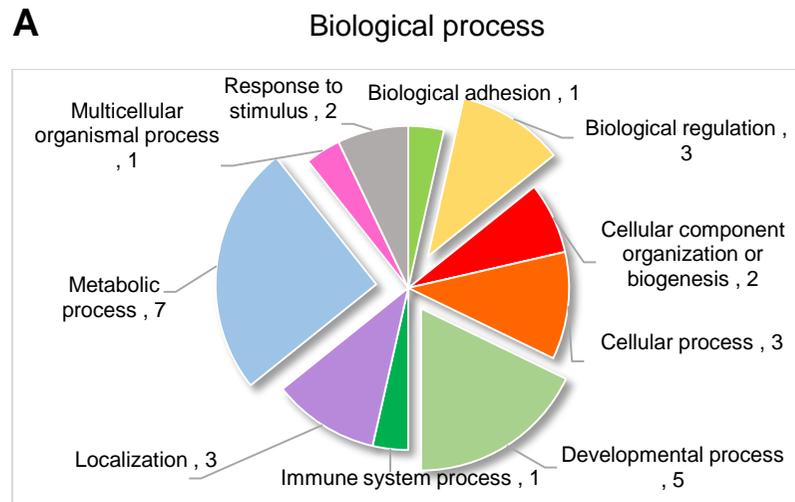
Figure 66, a pie charts of the biological processes from the overlapping downregulated genes of CIC knockdown and FGF4 overexpression. (A) 17 genes where found to fit into 28 process categories within biological adhesion, biological regulation, cellular component organization/biogenesis, cellular process, developmental process, immune system process, localisation,

metabolic process, multicellular organismal process & response to stimulus sub-groups. (B) The largest biological process group, metabolic processes, was sub-categorised to establish the type of cellular processes the upregulated genes were involved in. 17 genes were placed into 28 a further processes primary metabolic process, biosynthetic process, catabolic process, nitrogen compound metabolic process & phosphate-containing compound metabolic process sub-groups. (C) The biological process sub-categories of biological regulation processes which was found to contain 3 genes in 4 further process sub-categories with regulation of molecular function, homeostatic process & regulation of biological process sub-groups. (D) The developmental sub-category contained 5 genes in 6 further process sub-categories within mesoderm development, cell differentiation, ectoderm development & embryo development sub-groups.

Overlapping downregulated group associated pathways
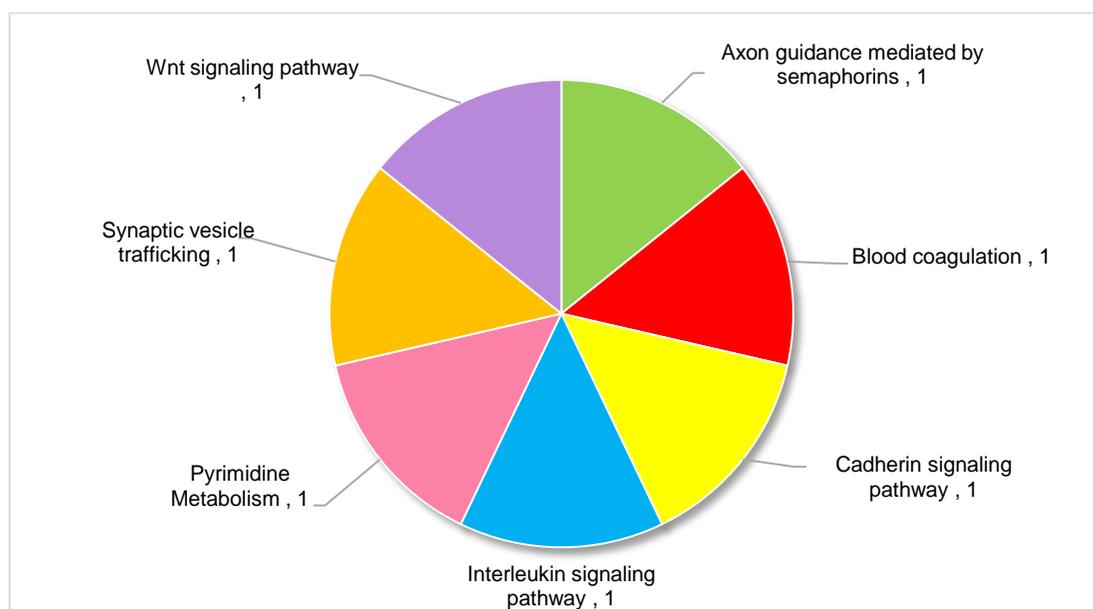
Figure 67, a pie chart showing the 7 pathway hits associated from 17 genes from the upregulated group using the PATNHER GO pathway tool for *X. tropicalis.*

6.2.6 The ontology of overlapping downregulated genes in *CIC* knockdown and *FGF4* overexpression.

The Xenbase (v4.2) GO term tools associated with *X. tropicalis* genes were applied to the overlapping downregulated genes in *CIC* knockdown and *FGF4* overexpression (table 9). The Xenbase GO tools were utilised to create a list of terms which applied to the molecular function, the biological process and area where the protein component would be found in the cell to each of the genes from the downregulated group (table 9).

Gene	GO Term	
atp2a2	Molecular function	• ATP binding • Calcium-transporting atpase activity • Metal ion binding
	Cellular component	• Integral component of membrane • Sarcoplasmic reticulum membrane
axl	Molecular function	• ATP binding • Protein tyrosine kinase activity
	Cellular component	• Integral component of membrane
cebpa	Molecular function	• DNA binding • DNA-binding transcription factor activity • Sequence-specific DNA binding
	Biological Process	• Definitive hemopoiesis • Hematopoietic stem cell migration • Liver development • Neutrophil differentiation • Primitive hemopoiesis • Transcription, dna-templated
celsr2	Molecular function	• G-protein coupled receptor activity • Calcium ion binding
	Biological Process	• Cell surface receptor signalling pathway • Homophilic cell adhesion via plasma membrane adhesion molecules
	Cellular component	• Integral component of membrane • Plasma membrane
cygb	Molecular function	• Heme binding • Iron ion binding • Oxygen binding

		• Oxygen carrier activity
dpysl3	Molecular function	• Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
	Biological Process	• Actin filament bundle assembly • Peripheral nervous system neuron axonogenesis • Positive regulation of filopodium assembly • Regulation of cell migration • Regulation of neuron projection development • Response to axon injury
	Cellular component	• Cytoplasm • Growth cone
efnb3	Molecular function	• Integral component of membrane
foxi4.1	Molecular function	• DNA-binding transcription factor activity • DNA-binding transcription factor activity, RNA polymerase II-specific • Sequence-specific DNA binding
	Biological Process	• Anatomical structure morphogenesis • Camera-type eye development • Cell differentiation • Positive regulation of transcription, dna-templated • Transcription, dna-templated
	Cellular component	• Nucleus
irx3	Molecular function	• Sequence-specific DNA binding
	Biological Process	• Brain development • Dorsal/ventral pattern formation • Endocrine pancreas development • Gastrulation • Kidney development • Maintenance of kidney identity • Mesoderm development • Metanephros development • Negative regulation of neuron differentiation • Negative regulation of transcription by rna polymerase ii • Neural plate development • Positive regulation of neuron differentiation

		• Positive regulation of transcription by rna polymerase ii
		• Positive regulation of transcription, dna-templated
		• Pronephros development
		• Proximal/distal pattern formation
		• Proximal/distal pattern formation involved in nephron development
		• Proximal/distal pattern formation involved in pronephric nephron development
		• Regulation of transcription, dna-templated
		• Specification of loop of henle identity
		• Specification of pronephric tubule identity
		• Transcription, dna-templated
	Cellular component	• Axon • Cytoplasm • Nucleus
msi1	Molecular function	• RNA binding • Nucleic acid binding • Nucleotide binding
	Biological Process	• Central nervous system development
nkain1	Biological Process	• Regulation of sodium ion transport
	Cellular component	• Integral component of membrane • Plasma membrane
notch3	Molecular function	• Calcium ion binding • Receptor activity • Signalling receptor activity
	Biological Process	• Notch signalling pathway • Cell differentiation • Multicellular organism development • Regulation of developmental process • Regulation of transcription, DNA-templated
	Cellular component	• Integral component of membrane
pax6	Molecular function	• Sequence-specific DNA binding
	Biological Process	• Multicellular organism development • Regulation of transcription, dna-templated • Transcription, dna-templated
	Cellular component	• Nucleus

pkdcc.2	Molecular function	• ATP binding • Protein kinase activity
ripply2.2	Biological Process	• Negative regulation of DNA-binding transcription factor activity • Negative regulation of transcription by RNA polymerase II • Somitogenesis • Transcription, DNA-templated
	Cellular component	• Nucleus • Transcriptional repressor complex
s1pr5	Molecular function	• Sphingosine-1-phosphate receptor activity
	Cellular component	• Integral component of membrane
serpina1	Molecular function	• Serine-type endopeptidase inhibitor activity
	Cellular component	• Extracellular space
slc23a2	Molecular function	• Sodium-dependent L-ascorbate transmembrane transporter activity
	Cellular component	• Integral component of membrane
sox11	Molecular function	• DNA binding • DNA-binding transcription factor activity • Sequence-specific DNA binding
	Biological Process	• Cell differentiation • Nervous system development • Transcription, dna-templated
	Cellular component	• Nucleus
spib	Molecular function	• DNA-binding transcription factor activity • DNA-binding transcription factor activity, RNA polymerase II-specific • Sequence-specific DNA binding
	Biological Process	• Cell differentiation • Regulation of myeloid leukocyte differentiation • Regulation of transcription by rna polymerase ii
	Cellular component	• Nucleus
tmem119	Biological Process	• Ossification
	Cellular component	• Integral component of membrane
znf219	Molecular function	• Nucleic acid binding

Table 9, Gene ontology and GO annotations from overlapping downregulated transcripts from CIC knockdown and FGF4 overexpression taken from the Xenbase database. All transcripts had GO entries except unc13d.

6.3 Discussion

6.3.1 CIC acts as a transcriptional repressor downstream of the FGF pathway.

The data from this RNA-seq experiment is consistent with the hypothesis that CIC operates as a transcriptional repressor downstream of FGF signalling, operating to repress a subset of genes of the FGF pathway at the late gastrulation, early neurulation stages of development. This finding corresponds to the phenotypes of knockdown *CIC* and FGF4/FGF8 overexpression, producing a similar posturised phenotype (Fletcher et al., 2006; Isaacs et al., 1995a). 17 putative genes of FGF signalling were found to be upregulated in the RNA-seq data for *CIC* knockdown and *FGF4* overexpression (*Apold1, atf3, bmp7.2, cbx4, cdc14b, fat1, fgd3, fos, fosl1, frzb, ier3, junb, mmp1, nuak2, rasl11b, sgk1, sox17b, tmcc1* and *wnt8a*) (fig 59 & table 4). Of these genes *fos, junb, Bmp7* and *rasl11b* are known targets of FGF signalling involved in the induction of mesoderm formation (table 10) (Branney et al., 2009; Dale et al., 1997; Pezeron et al., 2008). 19 genes were found to be upregulated which have not been previously identified as targets of positive regulation by FGF signalling (*Arrdc2, b4galt1.1, c4bpa, chic1, cldn6.1, exoc3l1, fam83c, gadd45a, gpcpd1, htr1b, lims1, mst1, nfkbiz, pnpla3, prkd1, riok3, sat1, sept2, tsc22d3* and *usp2*). These genes are known to be involved in multiple pathways (fig 60 & table 4).

One gene which is involved in the same process of development as FGF but has not been shown to be regulated by FGF signalling until now is *macrophage-stimulating 1 (Mst1). Mst1* and *Mst2* are tumour suppressors which restrict cell proliferation and cell survival. *Mst1* is a component of the highly evolutionary conserved hippo pathway which CIC is known to be a part of in the *D. melanogaster* animal model (Song et al., 2010; Thompson et al., 2006). Both FGF and *Mst1* are involved in limp development in the developing embryo. It is well known that limb size is mediated by epithelial-mesenchymal interactions with factors such as FGFs (Su et al., 2014; Sun et al., 2002). Song *et al.* (2010) suggested that limb size and organ growth is modulated by cell proliferation or cell death by reduction of *Mst1/Mst2* expression, likely through regulation of epithelial-mesenchymal factors such as FGFs. *CIC* expression is also known to be a key regulator of organ growth and regulator of proliferation in the *D. melanogaster* development controlled by the hippo pathway at the mRNA level (Yang et al., 2016).

These genes are not only putative genes of the FGF pathway but have known roles across multiple pathways, which is the likely reason FGF is able to cross talk in other

pathways. Multiple genes of the Wnt signalling and TGFβ signalling pathways had upregulated genes suggesting cross talk of signalling pathways through regulation of CIC repressional activity. These findings provide a new pathway of enquiry to the crosslink signalling of FGF and other pathways.

Crosstalk between signalling pathways is an important process that allows the cell signals to co-operate in the development of the embryo. Previous research on the signalling crosstalk has shown that FGF and Wnt pathways interact with one another regulating a diverse range of biological processes although the exact interactive mechanism was unknown (Burks et al., 2009; Dailey et al., 2005; Keenan et al., 2006a). Keenan et al. demonstrated FGF and Wnt co-operate together to induce the expression of *Cdx* genes within the mesoderm. In addition, FGF and Wnt signalling is known to function together to pattern anteroposterior neural ectoderm (McGrew et al., 1997) and are involved in the regulation of somitogenesis (Dequeant et al., 2006; Dequeant and Pourquie, 2008).

Activation of the canonical Wnt pathway leads to the stabilisation of β-catenin an intracellular signal transducer (Peifer et al., 1994). *β-catenin* initiates the transcription of the target genes of the canonical Wnt pathway by binding to T-cell factor/lymphoid enhancer-binding factor (TCF/LEF) transcription factor (MacDonald et al., 2009). TCF/LEF family of transcription factors contain a high-mobility group domain allowing them to bind to DNA imitating target gene transcription (Cadigan and Waterman, 2012). Groucho a transcriptional co-repressor of CIC, is known to associate with enhancers of target genes of the TCF/LEF transcription factor, which results in the downregulation of transcription of target genes (Brantjes et al., 2001). In *D. melanogaster*, Groucho is known to be inhibited by MAPK transduction (Hasson et al., 2005). In *X. laevis,* FGF4 and Wnt8 combined activity was shown to increase the expression of *XmyoD* through relief of Groucho (Burks et al., 2009). The RNA-seq data suggests a new mechanism for how FGF interreacts with the Wnt pathway, through relief of CIC repression.

Gene	Previous evidence of positive regulation by FGF signalling
apold1	Microarray analysis to identify differentially expressed transcripts in HAECs and HUVECs. Apold1 was found to have increased expression in cell cultures which had increased expression of FGF2 and FGF5 (Seo et al., 2016).

atf3	ATF-3 and c-Jun form a complex and levels of both proteins are increased in response to FGF (Tan et al., 1994).
bmp7.2	FGF interacts with Bmp7 regulate development of metanephric mesenchyme (Dudley et al., 1999).
cbx4	Cbx4 is found to be positively regulated by FGF signalling in early development (Branney et al., 2009).
cdc14b	The Cdc14B and FGF8 are in a parallel pathway in regulation of ciliogenesis in *D. rerio* (Clément et al., 2011).
fat1	Western blot analysis revealed that when RASMCs were treated with FGF2 Fat1 protein levels increased (Hou et al., 2006).
fgd3	FGF2 regulates the expression of *fgd3* in chondrocytes cells (Buchtova et al., 2015).
fos	During mesoderm Induction FGF2 treatment leads to increased AP-1 activity in animal cap explants (Kim et al., 1998).
fosl1	In microarray analysis of Bovine granulosa cells treated with FGF8 *fosl1* expression was increased (Jiang et al., 2013; Price, 2016)
frzb	Frzb is found to be positively regulated by FGF signalling in early development (Branney et al., 2009).
ier3	*Ier3* is upregulated when treated with FGF1 in *M. musculus* osteoblastic OB-TOP#1 cells (Ambrosetti et al., 2008).
junb	FGF2 induces expression of *junb* in *M. musculus* Y1 cell line (Vitorino et al., 2018).
mmp1	Treatment of primary *H. sapiens* OA chondrocytes with FGF2 increased expression of mmp1 (Nummenmaa et al., 2015).
nuak2	nuak2 found to have increased expression in iFGFR4 treated *X. laevis* embryos (H.B., unpublished).

rasl11b	rasl11b was found to be upregulated in iFGFR4 treated X. laevis embryos (H.B., unpublished).
sgk1	FGF23 upregulates SGK1 in proximal tubular epithelial cells (Andrukhova et al., 2012).
sox17b.2	sox17 expression is dependent on a FGF signal in *M. musculus* embryonic stem cells (Hansson et al., 2009).
tmcc1	tmcc1 found to have increased expression in iFGFR4 treated X. laevis embryos (H.B., unpublished).
wnt8a	In *G. domesticus* ectodermal explants, FGF4 induces Wnt8a expression (Urness et al., 2010).

Table 10, the list of genes which have published evidence of positive regulation by FGF signalling.

6.3.2 Pathways associated with overlapping transcripts upregulated in *CIC* knockdown and *FGF4* overexpression in *X. tropicalis*

6.3.2.1 Expression of genes involved in the formation of the central nervous system

Several pathways were found to be associated with the upregulated overlapping genes of *CIC* knockdown and *FGF4* overexpression. PANTHER analysis revealed both the GnRH receptor (*fos*, *junb*, *bmp7.2* & *atf3*) and TGFβ signalling (*junb, bmp7.2, fosl1* & *rasl11b*) pathways contained the most upregulated genes associated with any one pathway, each containing 4 of the upregulated genes (fig 64, table 4 & 6). GnRH receptor and TGFβ signalling pathways are known to be involved with the formation of the nervous system in early development (Wierman et al., 2011; Wu and Hill, 2009).

The GnRH receptor pathway is the primary regulator responsible the release of gonadotrophs, luteinizing hormone (LH) and follicle-stimulating hormone (FSH). LH and FSH are regulated by the secretion of GnRH from the hypothalamus which acts upon G-protein coupled receptors at anterior pituitary to regulate their synthesis. LH and FSH stimulate the formation of the gonads and although not critical for life, the GnRH receptor pathway is essential for reproduction. LH secretion due to stimulation by GnRH leads to the stimulation of the gonads to secrete other hormones testosterone, oestrogen and progesterone.

These sex hormones act in a negative feedback loop to inhibit secretion of GnRH and in turn downregulating the secretion of LH and FSH. Studies have revealed inhibition of BMP initiates neural induction via FGF signalling (Marchal et al., 2009). The PANTHER GO analysis revealed that *bmp7.2* is involved in both the GnRH receptor and TGFβ signalling. FGF8 and BMP/TGFβ antagonists signalling is known to regulate differentiation and determination of GnRH and olfactory sensory neurons cell fate (Chiba et al., 2008; Chmielnicki et al., 2004; Marchal et al., 2009; Streit et al., 2000; Wilson and Hemmati-Brivanlou, 1995; Zimmerman et al., 1996a).

The involvement of upregulated genes in the both the GnRH receptor and TGFβ signalling pathways suggests that CIC has a role in initiation of neural induction by FGF signalling in development of the *X. tropicalis* embryo. The 'default' model of neural induction suggests that BMP inhibition alone is required and appropriate for neural induction (Muñoz-Sanjuán and Brivanlou, 2002), whilst an alternative theory due to findings in *G. domesticus,* suggest that supplementary instructive early signals from FGF are required neural induction (Streit et al., 2000; Wilson et al., 2000). The finding in the RNA-seq data from *CIC* knockdown and *FGF4* overexpression suggests that inhibition of *bmp7.2* signalling itself is regulated by FGF4 signalling by control of repression of CIC. These findings reinforce are understanding that a supplementary early signal from FGF is required neural induction. Interfering with normal development of the olfactory sensory neurons cell fates could explain the unresponsive phenotype observed in CIC-S morpholino targeting seen previously in this thesis (chapter 5).

Other genes of note which are upregulated in CIC knockdown and FGF4 overexpression which are known key components of nervous system are *5-Hydroxytryptamine receptor 1B* (*htr1b*), *Serum/glucocorticoid regulated Kinase 1 (sgk1)* and *Wingless-Type MMTV Integration Site Family, Member 8A (wnt8a)*. *htr1b* gene encodes for G-protein coupled receptor for serotonin, regulating mood and social behaviour, appetite and digestion, sleep, memory (Villafuerte et al., 2009).

The *sgk1* gene encodes for a serine/threonine protein kinase involved sodium ion homeostasis by the regulation of a wide variety of ion channels and neuron excitability by activating a number of ion channels (Lang et al., 2010). *Sgk1* is also known to play a role in membrane transporters, cellular enzymes, transcription factors, neuronal excitability, cell growth, proliferation, survival, migration and apoptosis (Lang et al., 2010). *Wnt8a* expression is required for normal development of brain structures. Although Wnt8a is considered to be a posturising agent. Wnt8 is required for the initial

subdivision of the neuroectoderm in the developing *X. laevis* embryo (Rhinn et al., 2005). As previous studies in *H. sapiens* (Rousseaux et al., 2018), *M. musculus* (Yang et al., 2017) and *D. melanogaster (Yang et al., 2016)* have shown CIC retains a role in neural development in amphibian.

6.3.2.2 Expression of genes involved in mesoderm induction

We know that activation of FGF signalling leads to increases of expression of the Activator protein-1 (AP-1) transcription factor which plays an important role in mesoderm induction (Kim et al., 1998). The mechanism for how FGF signalling leads to increases of *jun* and *fos* has been unknown until now. The RNA-seq data reveals that CIC sits upstream of both *jun* and *fos* and upon activation of FGF transduction lead to relief of CIC repression and increase of their expression.

The structure of AP-1 consists of dimer proteins belonging to the c-Fos, c-Jun, ATF and JDP families which form the heterodimer structure (O'Shea et al., 1989). Overexpression of two components of the AP-1 transcription factor, *jun* and *fos* lead to the posteriorised phenotype in *X. laevis* embryos and leads to induction of mesoderm formation in animal caps (Kim et al., 1998). In a similar fashion, knockdown of *CIC* and overexpression FGF4 leads to a similar posteriorized phenotype in *X. tropicalis* (chapter 5). The mechanism for the posteriorized phenotype in *CIC* knockdown is likely due to the increases of expression of *jun* and *fos.* In this context CIC appears to play an important role in mesoderm induction and the development of the anterior-posterior axis due to control of the expression of *jun* and *fos.*

Interestingly, *bmp7.2* has increased expression in *CIC* knockdown and *FGF4* overexpression and in *M. musculus, bmp7* is a known regulator of AP-1 function in the context of renal organogenesis (Muthukrishnan et al., 2015). Knockout of *bmp7* has highlighted its importance. Knockout of *bmp7* leads to ectopic cell death in the developing nephrogenic zone of the foetal kidney and developing eye (Dudley et al., 1995; Dudley and Robertson, 1997; Luo et al., 1995). Expression of *bmp7* leads to the phosphorylation and activation of MAPKs TAK1 and JNK. The activation of the MAPKKK TAK1 and JNK lead to the targeting of jun for phosphorylation. The phosphorylated/activated jun targets genes by binding to AP-1 elements leading to increases in gene transcription.

One such gene that contains an AP-1 element is *Myc* which is required for renewal of mesenchymal nephron progenitor cells (NPCs) *in vivo.* Jun is also known to bind

to itself increasing its own expression (Muthukrishnan et al., 2015). Jun and Myc are known to be key regulators of the expression of genes which are involved in the G1-phase of the cell cycle. In addition, the same study suggested that bmp7 and FGF9 synergistically control AP-1 transcription and in doing so regulate the G1-phase of the cell cycle (Muthukrishnan et al., 2015). NPCs treated with bmp7 revealed that significant increase in S and M-phase cells promoting proliferation due to accelerated G1 phase of the cell cycle. When FGF9 and bmp7 treatments were combined in the NPCs it was found to give additive effect with increases of S and M-phase cells and cell proliferation.

Other studies have also highlighted synergistic interaction between FGF and Bmp signalling in regulating development of metanephric mesenchyme (Dudley et al., 1999).The findings in RNA-seq data suggests, rather than having synergistic additive effect, the treatment of FGF in the NPCs is likely to further increases the expression of *bmp7,* increasing the concentration within the NPCs leading to increases cell proliferation.

Jun does not only form complexes by dimerization with fos (AP-1), but there is cross-talk with other families transcription familes (ATF/CREB) (Hai and Curran, 1991). *Activating transcription factor 3* (*atf3*) is a member of the cAMP responsive element-binding (CREB) protein family of transcription factors. FGF expression co-induces atf3 and *jun* gene expression which leads to increase of a heterodimeric complex formed between atf3 and jun (Tan et al., 1994). The formation of the atf3/jun complex leads to increased expression of *proenkephalin* gene expression by Ras dependent activation.

Previous data from the Isaacs lab has shown that the gene coding for a small GTPase, Rasl11b*,* a member of the Ras subfamily of putative tumour suppressor genes has increased expression when embryos are treated with FGF (H.B., unpublished). Evidence has been found that Rasl11b is known to influence mesendoderm formation in *D. rerio* (Pezeron et al., 2008) although no evidence exists in the *X. tropicalis* animal model to suggest Rasl11b acts in the same way. *Rasl11b* knockdown in *D. rerio* suppresses the one-eyed-pinhead mutant phenotype caused by mutations in the *one-eyed-pinhead* (*oep*) gene (Kiecker et al., 2000; Pezeron et al., 2008).

The *oep* gene is a member of the conserved EGF-CFC gene family required promoting Nodal signalling in axis and dorsal midline formation in the developing *D. rerio* embryo. EGF-CFC genes encode for membrane-bound extracellular factors

which are required for mesoderm, endoderm induction and left-right axis formation *D. rerio* (Gritsman et al., 1999). In *M. musculus*, EGF-CFC genes include *Cripto1* (*Cr-1*) in *M. musculus*, *Fgf receptor ligand 1* (*Frl1*) in *Xenopus* and *criptic (CR-1)* in *H. sapiens*. Mutations of the *Cr-1* gene in *M. musculus* lead to failure of the formation of the primitive streak with embryos lacking embryonic mesoderm and head structures (Ding et al., 1998). There is no known receptor for the EGF-CFC proteins depends upon an activin-type RIIB receptor system acting upon Smad-2 (Saloman et al., 2000). Whilst knockdown of *Rasl11b* alleviates the anteriorly truncated one-eyed-pinhead mutant phenotype caused by mutation of the *oep* gene, CIC knockdown leads to over expression of *Rasl11b* which is likely one of the mechanisms to explain the anterior truncation phenotype observed in TALENs *CIC* knockdown. Further analysis will be required to validate this hypothesis.

6.3.2.3 CIC repression targets the expression anterior-posterior regulating genes

Although misregulation of Rasl11b could give an indication for the mechanism for truncation of the anterior of the anterior-posterior axis, Wnt8a is also known to be involved with the proper formation of the anterior-posterior axis. Studies in mesoderm induction using the *Xenopus* model have led to the conclusion that growth factors such as activin, Vg1, wnt8 and noggin play an important role in the formation of dorsal mesoderm/Spemanns organizer (Watabe et al., 1995). *Wnts* are known to have ventralising and posteriorising properties, antagonising the formation of head structures in the developing *Xenopus* embryo when overexpressed (Christian and Moon, 1993b; Fredieu et al., 1997).

The phenotype produced by *wnt8* overexpression appears to much like that of *CIC* knockdown with total loss of head structure (Leyns et al., 1997). *Wnt8* expression is also known to lead to the formation of a secondary axis when injected in the vegetal blastomere at the 32-cell stage in *X. laevis (Leyns et al., 1997)*. Wnt8 is suggested to primarily function during pattern formation at gastrulation stage of development (Christian and Moon, 1993a). *Wnt8a* expression is known to be induced by FGF signalling and when *sprouty* (*spry1/ spry2)* genes are overexpressed *wnt8a* expression is reduced (Mahoney Rogers et al., 2011; Wright et al., 2015). The *sprouty* genes are known to encode for proteins which are antagonists of RTK signalling. When *spry1/spry2* are knocked out*,* FGF regulated *wnt8a* expression is expanded (Mahoney Rogers et al., 2011). When FGF expression is absent *wnt8a* expression is also absent during early hindbrain development.

The *Sox17* gene is a member of the (SRY-related HMG-box) family of transcription factors, known to be involved in cell fate determination and regulation of embryonic development (Irie et al., 2015; Sinner et al., 2004; Vallier et al., 2009). *Sox17b* is involved in the Wnt/β-catenin signalling (Hudson et al., 1997). In animal cap experiments expression Sox17b was found to be expressed by treatment of activin and not FGF.

Frizzled related protein (*Frzb*) is gene which encodes for a secreted protein which interacts with wnt8a by binding (Wang et al., 1997). During gastrulation *Frzb* is found to be highly expressed in the region of the Spemann organizer. *Frzb* is secreted from Spemann organizer/ dorsal blastopore lip regulating dorsoventral and anteroposterior patterning to neighbouring cells (Leyns et al., 1997; Spemann and Mangold, 2001). Overexpression of *Frzb* by injection mRNA into single ventral blastomeres at the four-cell stage produced *X. laevis* embryos with phenotypes that had two partial posterior dorsal axes (Wang et al., 1997), unlike wnt8a overexpression which can produce a phenotype with two anterior dorsal axes (Leyns et al., 1997). Whilst overexpression of *Frzb* by injections mRNA from the one cell stage in *X. laevis* lead to a embryos having a posterised phenotype, with enlarged head, eyes, and cement glands (Leyns et al., 1997) much like that of FGF knockdown by dominant negative receptor expression (Isaacs et al., 1995a). *Frzb* overexpression does not change levels of Wnt8, but instead exerts dorsalising effects by impeding the action of *wnt8* by binding to the wnt8 protein (Wang et al., 1997). Further co-injections experiments with the CSKA-Xwnt-8 plasmid revealed that when *frzb* mRNA was coinjected head defects observed in the CSKA-Xwnt-8 plasmid injected embryos were not found (Wang et al., 1997).

Wnt8 is known to be required for the formation of the paraxial mesoderm or somatic mesoderm in the developing embryo (Hoppler et al., 1996). *Wnt8* expression is known to induce the expression of *myogenic regulatory factors D (MyoD)* which is a master regulatory gene of muscle cell formation (Weintraub et al., 1991). When embryos were treated with a dominant negative dnXwnt-8 block leading to knockdown of Wnt8 expression, *MyoD* expression was blocked (Hoppler et al., 1996). Given that *Frzb* is known to be a Wnt inhibitor, impeding the dorsalising action of *wnt8,* Wang et al. (1997) wanted to establish if *Frzb* overexpression also impacted somite formation. Interestingly, *Ripply Transcriptional Repressor 2 (RIPPLY2) a* gene which encodes for a nuclear protein belonging to a novel family of proteins found to be essential vertebrate somitogenesis (Chan et al., 2007) was found to be downregulated in the

RNA-seq transcriptome of CIC knockdown and FGF4 overexpression. *Frzb* overexpression was found to block *MyoD* expression.

Recently *Frzb* has been found regulate, chondrocyte maturation in long bone during skeletogenesis development in *G. domesticus* embryo and in chondrogenesis, hypertrophy in mesenchymal stem cells (Zhong et al., 2016). In addition, studies in mice have shown that Frzb regulates cartilage integrity, cortical bone thickness and density through the regulation of Wnt/β-catenin target genes such as *matrix metalloproteinase* (*Mmp*) (Lories et al., 2007), *Mmp1* is one of which that is highly expressed in the upregulated genes of FGF4 overexpression and CIC knockdown. Mmp1 is a downstream target of Jun N-terminal kinase (JNK) signalling in *D. melanogaster* (Uhlirova and Bohmann, 2006). Mmps are central to the process of collagen degradation, matrix breakdown and bone remodelling (Everts et al., 1992).

Achondroplasia, is a condition caused by a severe impairment of cartilage growth due to mutations of the FGFR3 which leads to the misregulation of FGF signalling (Teven et al., 2014). The condition could indeed be caused by the misregulation of FGF signalling leading to misregulation of genes like *frzb and mmp1* that sit-down stream of CIC. *FYVE, RhoGEF and PH Domain Containing 3* (*Fgd3*) is a gene which is involved in the signalling pathway of the Rho family of GTPases. The Rho family of GTPases are G proteins which belong to a subfamily of the Ras family. The Rho family of GTPases act as "molecular switches" which play a role in organelle development, cytoskeletal dynamics, cell movement, and other common cellular functions in response to extracellular stimuli (Boureux et al., 2007; Bustelo et al., 2007; Heasman and Ridley, 2008; Ridley, 2015). Rho GTPases function by regulating changes in actin of the cytoskeleton. *Fgd3* acts as a RhoA pathway activator (Buchtova et al., 2015; Maeda et al., 2011). FGF2 and Wnt3a modulates the expression of fgd3 in RCS chondrocytes leading to changes in actin fibres (Buchtova et al., 2015).

Similar to Fgf8, Wnt8a possesses an upstream RA response element that binds RA receptors. Retinoic acid (RA) signalling is known to create boundary of wnt8a and FGF8 expression (Cunningham et al., 2015). When embryos were deficient in retinoic acid (RA) synthesis due to *Raldh2* knockout embryos appeared to produce somites which were smaller than normal, and embryos had a shortened drunk. *Raldh2* is a gene which codes for a protein which catalyses the synthesis of RA (Duester, 2008). *Raldh2$^{-/-}$* embryos were found to have broader boundary expression of *FGF8* and *Wnt8a* had increased expression and extension of Wnt8a at the 7 somite stage,

demonstrating that the caudal and hindbrain expression domains were merged (Cunningham et al., 2015).

Although this study and other previous studies in the Isaacs lab suggest that *Frzb* expression is increased with FGF overexpression (Branney et al., 2009), this could lead to the question, is the increases of *Frzb* transcription a result of the cell trying to compensate for increased gene transcription of *wnt8a* or are both *Frzb* and *wnt8a* transcription finely regulated by FGF signalling? Further analysis will be required to unpick and explain the mechanism through the validation of these results. Whichever hypothesis remains to be true, CIC repression appears to be involved and important component of anterior-posterior axis formation by the regulation of FGF signalling.

6.3.2.4 Expression of genes involved in the MAPK signalling pathway.

FGF transduction has been well established to act through 3 signalling pathways; Ras/MAPK, P13/AKT and PLC-Y (Thisse and Thisse, 2005). The MAPK pathway utilises a chain of proteins which communicate an external signal via the binding of a signalling molecule to a cell surface receptor of the cell. The conformational changes of the receptors lead to the signal becoming internalised and leading to changes in the nucleus. This signal can lead to transcription factors being turned "on" or "off". An example of the activation of an immediate-early gene activation is *fos* (Gille et al., 1995). MAPK stabilises fos by phosphorylation, allowing it to form the AP-1 complex (Murphy et al., 2002; Whitmarsh and Davis, 1996). *Bmp7.2, gadd45a, ier3, and Rasl11b* are all genes which are known to be involved in MAPK signalling in some way which were upregulated by *CIC* knockdown and *FGF4* overexpression in the transcriptome data. Studies in *D. melanogaster* show that fos, a known target of MAPK targeting, is an effector MAPK/JNK signalling during the formation of wing vein and photoreceptor of the developing retina, but no evidence exists to say this is the same in mammalian animal models (Ciapponi et al., 2001).

Bmp7 is known to be involved in epithelial mesenchymal transition (EMT), not only in the transition of EMT (Lim et al., 2011), but in the reversal of EMT process (Zeisberg et al., 2003). During Bmp7 stimulation corneal epithelial cells showed a marked increase in TGF-β, cell cycle, JAK-STAT and MAPK signalling (Blank et al., 2009; Carreira et al., 2014; Kowtharapu et al., 2018; Lim et al., 2011). These increases of signalling allow adaption of corneal epithelial cells, to transition from a sedentary to the migratory state and is critical for movement (Barriere et al., 2015). When treated with Bmp7, corneal epithelial cells showed continued activation of

MAPK cascade proteins p44/42 MAPK and p38. JNK in comparison was only activated briefly (Kowtharapu et al., 2018; Zeisberg et al., 2003). What this means in context to CIC and regulation of EMT is still unknown but suggests that CIC may have a role in EMT regulation. It may also suggest that if increases of Bmp7.2 lead to activation of MAPK, CIC may regulate its own repressional activity in some fashion. Further analysis will be required to address these questions.

The *growth arrest and DNA damage-inducible 45* (*Gadd45*) family of genes are a group of genes which have increased expression following cellular stress (Liebermann and Hoffman, 2008) and introduction of DNA-damaging agents (mutagens) to the cell (Hughes et al., 2012). Gadd45a is known to play a role in the MAPK signal transduction by the activation of MAPKKK (Takekawa and Saito, 1998). Interestingly, *Immediate early response 3* (*Ier3*) another gene which is induced in response to cellular stress (Arlt and Schafer, 2011). *Ier3* is also associated MAPK1 and MAPK14 signalling (Garcia et al., 2002; Letourneux et al., 2006). Could increases of *Ier3* and *Gadd45a* expression rather than being a product of direct control of CIC repression, have increased expression due a response to cell death caused by the knockdown of CIC or overexpression of FGF. This question remains to be unanswered.

Lastly, *Ras-Like Family 11 Member B* (*Rasl11b)* is a gene which encodes for a small GTPase protein belonging to a family of proteins with a high degree of similarity to Ras proteins (Louro et al., 2004; Pezeron et al., 2008; Stolle et al., 2007). *Rasl11b* expression has previously been identified as a target of FGF signalling. *Rasl11b* a member of the Ras superfamily involved in MAPK signalling (H.B., unpublished). *Rasl11b* has increased expression in both *CIC* knockdown and *FGF4* overexpression.

6.3.3 Transcripts downregulated by CIC knockdown and FGF 4 overexpression in the *X. tropicalis* transcriptome.

Standout genes of overlapping downregulation in *CIC* knockdown and *FGF4* overexpression are *neurogenic locus notch homolog protein 3 (notch 3) and Paired box protein-6* (*pax6). Notch3* is the *H. sapiens* homologue of the *D. melanogaster* type I membrane protein notch (Bray, 1998). *Notch3* is known to promote neuronal differentiation in *M. musculus* cell culture (Rusanescu and Mao, 2014). Notch3 is known to restrict FGF ability to enable adult hippocampus-derived multipotent

progenitor (AHP) cells to be multipotent and instead induce the cells to into an astroglial fate (Tanigaki et al., 2001).

Pax6 is a "master control" transcription factor expressed during embryonic development regulating organs and tissue development. Like *Hox* genes, the *Pax6* gene is known to play a role specifying the body plan of the organism (Gehring, 1996). Pax6 primary role is in the regulation of eye, central nervous system and epidermal tissue (Davis et al., 2008). *Pax6* is known to be downregulated by overexpression of *FGF8* (Bertrand et al., 2000), whilst inhibition of FGF2 lead to increases of *Pax6* expression (Greber et al., 2011). Interestingly, Frzb is known to downregulate other *pax* genes, no evidence exists to suggest it downregulates *pax6* expression. This data is consistent with the published data suggest an increase in FGF expression leads to the down regulation of *Pax6*.

6.3.4 Knockdown of *CIC* and *FGF4* overexpression does not increase transcription of known targets involved in both pathways

Surprisingly, the RNA-seq data presented an absence of known transcripts involved in both downstream pathways of CIC and FGF, such as the most well-known of CIC targets, the ETS family of transcription factors Pea3 subfamily, Etv1, Etv4 and Etv5 (Brent and Tabin, 2004; Dissanayake et al., 2011; Herriges et al., 2015; Kawamura-Saito et al., 2006; Lee et al., 2011; Wang et al., 2017; Weissmann et al., 2018; Zhang et al., 2009b). *Etv1, Etv4* and *Etv5* are known to be induced by FGF MAPK transduction and repressed by CIC activity in *H. sapiens and M. musculus* (Chen et al., 1999; de Launoit et al., 2006; Fores et al., 2017; Graves and Petersen, 1998; Lee et al., 2011; Remy and Baltzinger, 2000). Despite CIC and FGF having a role in *Etv* gene expression, RNA-seq data suggests that there is no significant increase in expression of the *Etv* genes in *CIC* knockdown, *FGF4* overexpression compared to the water injected wild type sibling embryos. Currently no temporal *Etv5* expression profile exists in *X. tropicalis,* although previous expression data shows *Etv1* and *Etv4* are expressed in early development, with *Etv1* being expressed maternally and *Etv4* becoming expressed at last gastrulation (stage 12) in *X. tropicalis* (Owens et al., 2016). The likely reason for the lack of increase of *Etv* gene expression could be attributed to the requirement of co-repressor transcription factors Atxn1l or Atxn1, known to be important for the regulation of Pea3 family of genes (Wang et al., 2017).

Both Atxn1l and Atxn1 are expressed maternally but have very low expression during gastrulation and early neurulation stages of development (Owens et al., 2016).

Previous studies have shown that Atxn1I is able to stabilise the CIC protein by forming a co-repressor complex (Bowman et al., 2007; Crespo-Barreto et al., 2010; Lee et al., 2011), synergistically enhancing CIC ability to act as a transcriptional repressor (Lee et al., 2011). When expression of Atxn1I was decreased it lead to increases in *Etv* gene expression and destabilisation of the CIC protein (Wang et al., 2017). Therefore, if CIC requires the formation of a co-repressor complex with Atxn1I or Atxn1 to target *Etv* genes, lack of Atxn1I or Atxn1 could lead to increased expression despite the presence of the CIC protein in the cell or even with increased expression of FGF4. Importantly, this highlights despite the spatial and temporal expression profiles of *CIC* expression and presence of the CIC protein within the embryo, the ability of CIC to act as a transcriptional repressor in different contexts is modulated by other interacting transcription factors which have their own spatial and temporal expression profiles such as Atxn1I and Atxn1 (Owens et al., 2016).

Atxn1, Atxn1I and CIC are known regulators of the extracellular matrix remodelling in the formation of the lung in early development of the *M. musculus* embryo (Lee et al., 2011). Atxn1, Atxn1I and CIC regulate the normal formation of the lung alveolarization by controlling the expression of the *Mmp* genes (Lee et al., 2011). One member of the Mmp family of genes, Mmp1 was found to have overlapping upregulated expression of CIC knockdown and FGF4 overexpression in the RNA-seq data (table 61). In the context of lung development, no evidence exists to suggest that the Atxn1/Atxn1I-CIC repressional complex is involved with the regulation of *Mmp1,* although studies in hepatocellular carcinoma (HCC) in H. sapiens have shown that Mmp1 expression is controlled by CIC repressional activity (Kim et al., 2018). Increases in cell proliferation and malignancy in HCC were shown to be increased when CIC was post transcriptionally reduced. Survival rates of patients were shown to correlate with CIC levels in in HCC, with decreasing levels of the CIC protein leading to increased mortality rates. Analyse of the carcinoma cells revealed that *Etv4* a known target of CIC repression and *Mmp1* were significantly up-regulated in HCC cells lacking CIC (Kim et al., 2018). Etv4 is known to upregulate the expression of *Mmp1* which is important factor for HCC progression.

Although the RNA-seq data shows an increase of *Mmp1* overlapping expression in both *CIC* knockdown and *FGF* overexpression there is no evidence that *Etv4* has increased in expression in either *CIC* knockdown and *FGF* overexpression in comparison to the wild-type *X. tropicalis* expression. This suggests that neither FGF4 nor CIC regulate the expression of *Etv4* at the late gastrulation early or neurula stages of development. That could be due to the requirement of the formation of the

Atxn1/Atxn1l-CIC complex as previously mentioned, adding to the complexity of signalling. Interestingly, *Etv4* expression at the early neurula stages is known to become increased when the RNA-seq samples were collected (Owens et al., 2016), but Mmp1 expression is increased in *CIC* knockdown and *FGF* overexpression in comparison to the wild-type embryos suggesting the mechanism for *Mmp1* expression at the early neurula stage is still unknown.

Cdx and Hox transcription factors are known to have their expression regulated by FGF signalling (Keenan et al., 2006b; Reece-Hoyes et al., 2002). They play an important role in the development of the vertebrate body axis (Guo et al., 2004). Given their importance in vertebrate body axis elongation it is therefore surprising that there are no significant changes in *Cdx/Hox* expression. Although there were slight increases of *Cdx* expression in FGF4 treated embryos they were not significant compared to the wildtype siblings (See Supplementary Tables: Spreadsheet of RNA-seq results). In contrast CIC knockout embryos showed no change in *Cdx* or *Hox* gene expression suggesting that CIC is not responsible for *Cdx* or *Hox* gene expression.

In summary the RNA-seq data suggests that many of the genes which are involved with downstream of processes FGF signalling across MAPK transduction, neurogenesis, induction of mesoderm and anterior-posterior axis patterning throughout development are regulated by CIC repressional activity. This data taken in conjunction with other phenotypical evidence, overexpression and knockdown strongly suggest that the hypothesis of CIC acting downstream of FGF signalling is correct.

Chapter 7: General discussion

7.1 Summary

The regulation of gene expression through RTKs signalling plays an important role in early development. Despite the wealth of information, the mechanism for how changes of gene transcription through FGF signalling by MAPK transduction is poorly understood. The findings within this thesis could provide important insight into the mechanism of gene changes in a subset of FGF target genes. The research in this thesis has established strong evidence that CIC has a role in mediating FGF transcriptional regulation in *X. tropicalis.* This project has for the first time provided in-depth analysis of *CIC* in an early developmental biology context. These findings could explain numerous signalling mechanisms in early development and how FGF is able to cross-talk with other signalling pathways, such as the canonical Wnt signalling pathway (chapter 6) (Leyns et al., 1997; Marchal et al., 2009; Munoz-Sanjuan and Brivanlou, 2002; Streit et al., 2000; Wilson et al., 2000).

Identifying the gene locus and exon structure of *CIC* by cloning in *X. tropicalis* enabled the discovery of the CIC-L and CIC-S spatial and temporal expression profiles by *in situ* hybridisation and RT-PCR/qPCR analysis (chapter 3). Analysis revealed that the prominent isoforms of *CIC* had overlapping expression with several FGF's (chapter 4) (Lea et al., 2009). Interestingly, CIC-L was found to be expressed maternally, whilst the shorter CIC isoform, CIC-S, becomes expressed post MBT at the zygotic phase of embryonic development. Whilst analysing the expression of *CIC*, uncharacterised alternative isoforms were found that if expressed could also have alternative functions to the prominent CIC-L and CIC-S isoforms of *CIC.*

Knockdown analysis of *CIC* by TALENs lead to embryos becoming posturised at the anterior-posterior axis, a phenotype which has previously seen in FGF4 overexpression, hinting at the likelihood that FGF and CIC were involved in the same pathway for regulation of anterior-posterior patterning. Morpholinos were used as an alternative approach to TALENs knockdown, allowing targeting of the CIC-L and CIC-S isoforms. Morpholinos also enabled targeting of the maternally expressed CIC-L. The knockdown of the CIC-L isoform by morpholinos produced a kinked back phenotype which suggested irregular somite formation in the embryo, whilst targeting of CIC-S isoform produced embryos which had a wild-type phenotype but were un-responsive to touch suggesting disruption of neuro-muscular development in the

embryo. These findings provide further evidence that the isoforms of *CIC* have alternative functions in embryonic development.

Overexpression analysis of *CIC* revealed that overexpression of *H. sapiens* or *M. musculus* homologs of CIC-S in *X. tropicalis* did not produce a mutant phenotype. The reason for the lack of mutant phenotype is unknown at this time, but previous evidence has shown that the CIC protein requires additional factors to bind optimally to gene target enhancers/promoters to repress transcription. Factors such as Groucho in *D. melanogaster* (Dubnicoff et al., 1997; Papagianni et al., 2018; Valentine et al., 1998) and Ataxin-1 in mammalian animal models (Kim et al., 2013; Lam et al., 2006) are required for CIC to act optimally as a transcriptional repressor in some contexts.

When analysing the effects of FGF on the CIC protein, co-injecting GFP-CIC-S mRNA with FGF4 mRNA lead to the post-translational modification/degradation of CIC, whilst when treating CIC with FGF8 caused total loss of CIC-S protein. This leads to the question, do different FGF's affect CIC protein in different ways and how does this effect the regulation of transcriptional repression of CIC? Further research will be required to address these questions and establish the effects on CIC when treated with different ligands of FGF.

Transcriptome analysis of *CIC* knockdown and *FGF4* overexpression and the comparison to wild-type siblings at the late gastrulation/early neurula stages of development has provided an abundance of FGF gene network information. Overlapping genes have been discovered in the transcriptome data which provides substantial evidence that FGF signalling leads to regulation of CIC, causing changes in gene transcription within the developing embryo.

Genes of note which were not shown to be changed in the transcriptome analysis of upregulated overlapping *CIC* knockdown and *FGF4* overexpression were Etv1, Etv4 and Etv5 (Chen et al., 1999; Dissanayake et al., 2011; Herriges et al., 2015; Lee et al., 2011; Weissmann et al., 2018; Zhang et al., 2009). These genes are known targets of FGF and CIC regulation but were not found to have a significant change in expression compared to the wild type transcriptome. The likely reason is that neither FGF4 nor CIC regulate gene expression at post-gastrulation, early neurula stages of development. The interaction of FGF and CIC may change dependant on the context, activating different genes at different time points increasing the complexity of expression of the FGF-CIC network of genes. Further analyse will be required.

7.2 Regulation of CIC activity through RTK signalling

CIC controls cell proliferation downstream of the RTK Ras-MAPK signalling pathway and when knocked out enables cell growth without the need for MAPK transduction (Tseng et al., 2007). Because of CICs role as a regular of proliferation it comes as no surprise that numerous types of cancer have been shown to have mutations within the *CIC* gene (Bettegowda et al., 2011; Jiao et al., 2012; Kim et al., 2018; Sjoblom et al., 2006; Tanaka et al., 2017; Yip et al., 2012). CIC is well established to act downstream of the Torso and EGFR RTK signalling pathways in *D. melanogaster* and the findings from this project in *X. tropicalis* provides strong evidence that CIC functions downstream of the FGFR RTK in vertebrate development. Jiménez (2012) hypothesised that CIC was a general regulator of RTK–Ras–MAPK signalling response, which appears to correspond the findings within this thesis regarding FGF signalling. We know that Torso and EGFR have a role in anterior-posterior axis (Jimenez et al., 2000) and neurogenesis (Ajuria et al., 2011; Stathopoulos and Levine, 2005; von Ohlen and Doe, 2000) by modifying CIC repression. Evidence provided in this thesis now suggest that FGF is now known to be another RTK pathway which regulates CIC repressional activity in anterior-posterior axis and neurogenesis.

We know the importance of RTKs MAPK transduction in the regulation of CIC, it is likely that other RTKs which utilise the Ras-MAPK transduction also regulate CIC transcriptional repression. Further analysis of CIC will be required to see how many other RTKs play a role in the modulation of CIC repression and in what context throughout embryonic development.

Establishing which growth factors/RTKs that utilise MAPK transduction to regulate CIC, such as FGF and EGFR could be of critical importance in cancer therapeutics (Simon-Carrasco et al., 2018). If more RTKs are shown to regulate genes transcription through CICs regulation of repression it would prove interesting to see how multiple RTKs acting on CIC at any given time would modulate CIC transcriptional repression during embryonic development.

7.3 Alternative functions of CIC isoforms.

There is very limited data on the alternative functions of the prominent isoforms CIC-L and CIC-S in *H. sapiens*, *M. musculus* and *D. melanogaster* (Chittaranjan et al., 2014; Fores et al., 2015) . In addition, very little information exists of the temporal or spatial expression of *CIC* in development outside of the limited data in *D. melanogaster* and *D. rerio* model systems (Chen et al., 2014).

In *X. tropicalis* embryonic development, the CIC-L isoform is maternally expressed and when targeted by morpholinos produces a kinked back phenotype indicative of irregular somite formation. *Wnt8a* is a component the canonical Wnt pathway which is upregulated in the group of genes which has overlapping upregulation in *CIC* knockdown and *FGF4* overexpression data. Wnt8a is a known activator of MyoD expression (Fletcher and Harland, 2008; Hoppler et al., 1996) which is a 'master regulator' of myogenesis and somitogenesis (Maguire et al., 2012; Rudnicki et al., 1993). Interfering with the normal expression of Wnt8a by knockout of CIC is likely to cause upregulation of *wnt8a*, leading to abnormal increased activation of *MyoD*. This could explain the kinked back phenotype seen in CIC-L knockdown and the requirement of cross-talk between FGF and canonical Wnt pathway.

Unlike CIC-L, CIC-S becomes activated during zygotic expression. The alternative zygotic expression of CIC-S and its targeting by morpholinos reveals a very different phenotype from the CIC-L isoform, with embryos appearing to be wild-type but suffering from varying degrees of paralysis and twitching, which is suggestive of issues with neuromuscular formation in the developing embryo. This could be the strongest evidence of alternative functions of the prominent isoforms of *CIC*. Analysis of these phenotypes is still required to understand the molecular mechanism and differences between CIC-L and CIC-S isoform knockdown and function, but investigation of genes in the neuromuscular/myogenic pathways could provide the answer.

7.4 Modulation of repression of CIC during development of the embryo

Several studies have shown that CIC repression is modulated by additional transcription factors in different contexts. CSN1b a subunit of the COP9 signalosome is known to protect CIC from ubiquitylation by Cullin 1/SKP1-related A/Archipelago E3 ligase (Suisse et al., 2017). CSN1b acts to protect CIC from EGFR MAPK-dependent/-independent modes of degradation, maintaining basal levels in *D. melanogaster* (Suisse et al., 2017). In addition, we know that in *D. melanogaster,* Dorsal is a co-repressor that binds to conserved A/T-rich sites at its gene target enhancers (Papagianni et al., 2018). The recruitment of CIC to Dorsal enables it to bind to low-affinity DNA sites which would have previously not been targeted for transcriptional repression.

In *X. laevis* the wnt pathway has been shown to regulate the Groucho-related transcriptional repressors (Roose et al., 1998). Although Groucho is a known co-

repressor of CICs in *D. melanogaster*, no evidence has been shown to indicate that is the case in vertebrate animal models (Ajuria et al., 2011; Cinnamon et al., 2004; Cinnamon and Paroush, 2008; Paroush et al., 1997). Instead Ataxin-1 is the only known co-repressor so far discovered in vertebrate animal models (Lam et al., 2006; Lu et al., 2017). As this is the first project analysing *CIC* in amphibians, no evidence presently exists to confirm CIC forms co-repressor complex seen in vertebrate or invertebrate animal models. This is an area of research that will need to be addressed in the future.

Although co-repressors are known to modulate CIC binding ability, evidence of treating CIC with different FGF leads to different effects on CIC. FGF8 appeared to be more potent at degrading CIC, whilst FGF4 leads to the post-translational modification of the CIC-tagged protein (chapter 5). During the formation of mesoderm, in addition to FGF4 and FGF8, FGF3 and FGF20 are expressed (Branney et al., 2009; Christian and Moon, 1993). FGF3 and FGF20 are required for the expression of other mesoderm specific genes (Fletcher and Harland, 2008). Analysis of FGFs and how they affect the CIC protein is an important question in how CIC is modulated. The modulation of CIC and its interactions with various transcription factors suggests that the network of regulation is highly complex and will need further analysis to understand.

7.5 A mechanism for cross-talk between FGF and other signalling pathways

One key area of research is the cross-talk between FGF signalling and other signalling pathways, such as the canonical Wnt signalling pathway. Wnt8a was found to be one the upregulated transcripts in the overlapping genes of CIC knockdown and FGF4 overexpression. This indicates that FGF signalling and the relief of CIC repression leads to the activation of Wnt8a expression.

7.6 Future work

Previous studies have identified MAPK transduction as the mode of post-transcriptional regulation of CIC. In *D. melanogaster* Torso MAPK transductions leads to the degradation of CIC, whilst EGFR MAPK transduction leads to the relocation of nuclear CIC to the cytoplasm. In both examples of RTK signalling, once MAPK becomes internalised, as a result of the Ras-MAPK signalling cascade, MAPK binds to CIC directly at the C2 domain leading to its phosphorylation. At present, no

phosphorylation profiles exist to explain the difference in how either of Torso or EGFR can act upon CIC in these different ways.

Although western blot analysis of CIC treatment reveals that there are increase in dpERK upon treatment within the *X. tropicalis* embryo with FGF4/FGF8, no direct evidence has been found to confirm that CIC is phosphorylated. Pulldown analysis could provide the solution to the unanswered question of how CIC is post-translationally modified/degraded when treated with FGF. GFP-pulldown down of the *H. sapiens* GFP-tagged homolog of CIC-S and western blot using an antibody to detect phosphorylated motifs within the protein could provide important information in regard to CIC phosphorylation when treated with FGF.

Although RNA-seq data suggests that CIC regulates a subset of FGF targets genes (chapter 6), further validation will be required to determine which of the subset of genes are direct binding targets of CIC and which genes are regulated by the targets of CIC repression. One such method to determine if CIC acts to directly target gene expression would be to use bioinformatics approach to look for the octameric binding sequences of the HMG-box of CIC within the overlapping upregulated target genes. This approach was previously used within the lab to confirm that the *X. tropicalis* homologs of ETS Etv genes contained binding sequences and could be used to give an indication of which genes are directly targeted by CIC transcriptional repression. An additional approach, would be the use of ChIP-seq analysis. This approach would provide identification of the global binding sites of target genes of CIC and provide a wealth of information in regard to the targets of CIC repression.

7.7 Conclusion and implications

This thesis provides a significant body of evidence that the highly conserved transcriptional repressor, CIC, acts upon a subset of target genes of the FGF signalling pathway (chapter 6).

This subset of FGF target genes act upon a broad range of important pathways in early development, having roles in anterior-posterior axis formation, induction of neurogenesis to maintenance of mesoderm tissue and cell proliferation in the X. tropicalis embryo. Insight into the mechanism of FGF regulation target gene expression will provide important understanding for developmental disorders in *H. sapiens* diseases and therapeutic treatment of cancer. Like many genes which are critical regulators of cell proliferation in early development, *CIC* is a known tumour suppressor gene. Understanding of the molecular events that trigger cancer may

provide a more effective approach to treatment of cancer. The mechanism of FGF-CIC signalling may prove to be important for therapeutic treatment of developmental disease or cancer. Understanding the FGF-CIC signalling mechanism will also prove to be significant for stem cell research and tissue regeneration. This work will lay the ground work for deciphering the gene network of FGF-CIC signalling and play an important role in future research of the FGF pathway.

Appendices

Appendix 1
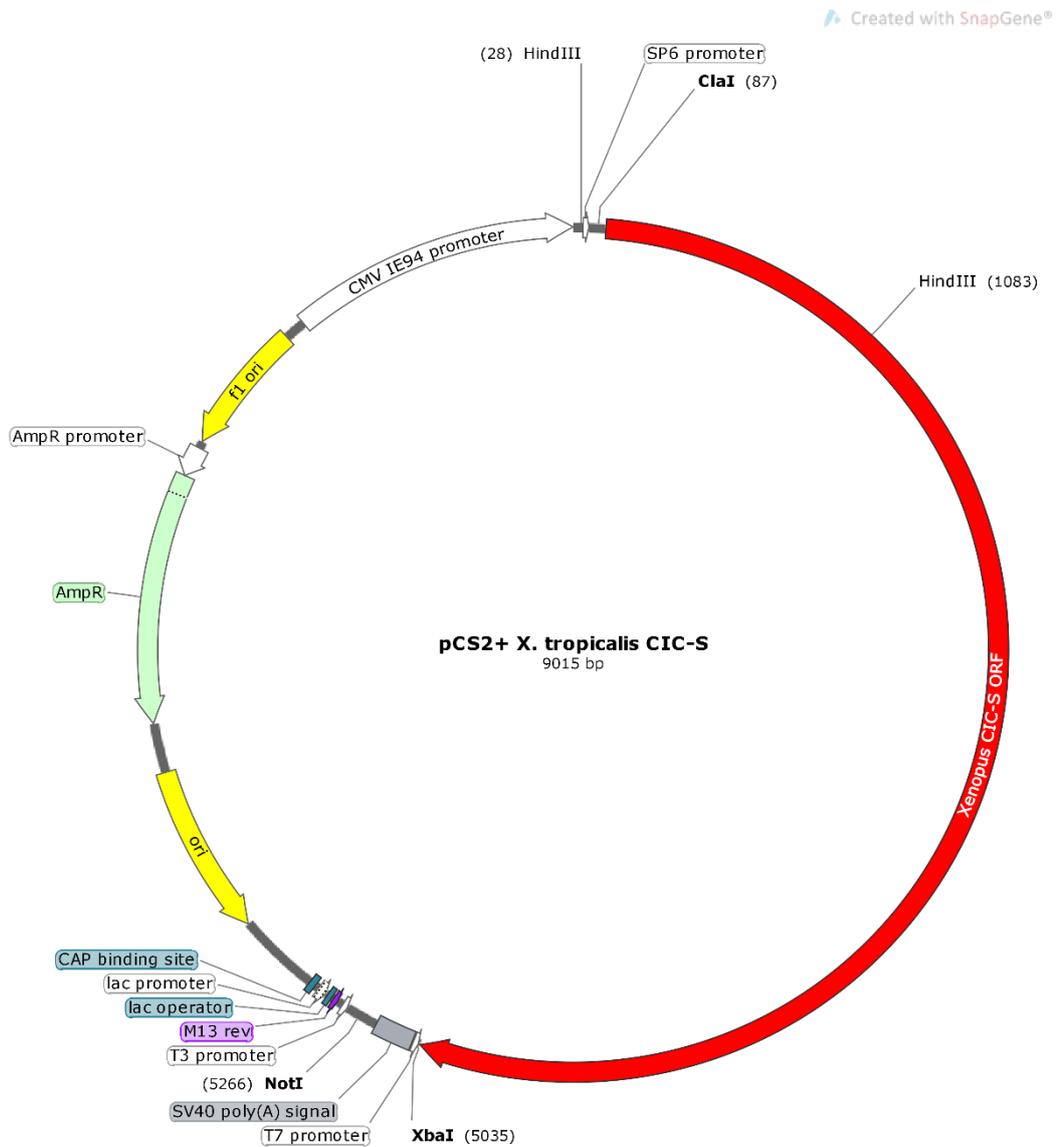
Figure 68, a plasmid map of pCS2+ CIC-S plasmid containing the ORF of *X. tropicalis* CIC-S homolog.

Abbreviations

2HG	2-Hydroxyglutarate
ACLY	ATP-Citrate Lyase
AKT	Protein Kinase B
AP-1	Activator Protein-1
ATXN(L)…	ATAXIN (Like)
Bmp…	Bone morphogenetic protein
Cdx…	caudal type homeobox
CIC	Capicua
CRLs	Cullin-RING E3 ubiquitin ligase complexes
CSN…	COP9 signalosome
Cux1	Cut Like Homeobox 1
DAG	Diacylglycerol
DIAP…	Death-associated inhibitor of apoptosis
dpERK	Diphosphorylated Extracellular Signal-Regulated Kinase
DUX4	Double homeodomain 4
DYRK1A	Dual specificity tyrosine-phosphorylation-regulated kinase 1A
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EFTs	Ewing's family of tumours
EGFR	Epidermal growth factor receptor
Egr1	Early growth response 1
Elk-1	ETS-Like Gene 1
ERK	mitogen-activated protein kinase 1
EST	Expressed sequence tag
Ets	E-twenty-six transformation-specific
FGF	Fibroblast growth factors

GO	Gene ontology
Gsx	Genomic Screened Homeobox
hCG	Human chorionic gonadotrophin
hkb	Huckebein
HMG	High-Mobility Group
HOG	Human Oligodendroglioma
Hox…	Homeobox
HS-GAG	Heparin/Heparan sulfate
HSST	Heparan sulfate 2-O-sulfotransferase
IDH…	Isocitrate dehydrogenase
Ind	Neuroblasts defective
KPNA3	Importin α4
LADD	Lacrimo-auriculo-dento-digital syndrome
MAPK	Mitogen-activated protein kinase
MBT	Mid-blastrula transition
mirr	Mirror
Mmp…	Matrix metalloproteinase
Mnb	Minibrain
MRF…	Myogenic regulatory factors
MSCs	Mesenchymal stem cells
MyoD	Myogenin D
NAM	Normal Amphibian Medium
NLS	Nuclear localization sequence
ODG	Oligodendroglioma
ORF	Open reading frame
Otp	Orthopedia homeobox
P13K	Phosphoinositide 3-kinase
p90^{RSK}	p90 ribosomal S6 kinase
PEA3	Polyomavirus Enhancer Activator 3
PKB	Protein kinase B
PKC	Protein kinase C
PLC…	Phospholipase C

PSM	Paraxial presegmental (or presomitic) mesoderm	
RPE	retinal pigment epithelium	
RTK	Receptor tyrosine kinase	
SCA1	Spinocerebellar ataxia type 1	
SMAD…	Sma- And Mothers Against DPP	
SOS	Son of seven	
Sox…	Sex Determining Region Y-Box	
SRF	Serum Response Factor	
TALEN	Transcription activator-like effector nucleases	
TGF-β	Transforming growth factor beta	
Tll	Tailless	
Tor	Torso	
TPM	Transcripts per million	
UTR	Untranslated region	
Vg1	Growth differentiation factor 1	
wg	Protein wingless	
Wnt…	Wingless-Type MMTV Integration Site Family	
Xbra	Brachyury	
Xcad4	Caudal type homeobox 4	
Xnr…	Xenopus nodal related	

References

Ahmad, S.T., Rogers, A.D., Chen, M.J., Dixit, R., Adnani, L., Frankiw, L., Lawn, S.O., Blough, M.D., Alshehri, M., Wu, W.*, et al.* (2018). Regulation of neural stem cell fate by the transcriptional repressor Capicua.

Ajuria, L., Nieva, C., Winkler, C., Kuo, D., Samper, N., Andreu, M.J., Helman, A., Gonzalez-Crespo, S., Paroush, Z., Courey, A.J.*, et al.* (2011). Capicua DNA-binding sites are general response elements for RTK signaling in Drosophila. Development *138*, 915-924.

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*, 257-270.

Amaya, E., Stein, P.A., Musci, T.J., and Kirschner, M.W. (1993). FGF signalling in the early specification of mesoderm in Xenopus. Development *118*, 477-487.

Ambrosetti, D., Holmes, G., Mansukhani, A., and Basilico, C. (2008). Fibroblast Growth Factor Signaling Uses Multiple Mechanisms To Inhibit Wnt-Induced Transcription in Osteoblasts ▽. Mol Cell Biol *28*, 4759-4771.

Andreu, M.J., Ajuria, L., Samper, N., Gonzalez-Perez, E., Campuzano, S., Gonzalez-Crespo, S., and Jimenez, G. (2012a). EGFR-dependent downregulation of Capicua and the establishment of Drosophila dorsoventral polarity. Fly *6*, 234-239.

Andreu, M.J., Gonzalez-Perez, E., Ajuria, L., Samper, N., Gonzalez-Crespo, S., Campuzano, S., and Jimenez, G. (2012b). Mirror represses pipe expression in follicle cells to initiate dorsoventral axis formation in Drosophila. Development *139*, 1110-1114.

Andrukhova, O., Zeitz, U., Goetz, R., Mohammadi, M., Lanske, B., and Erben, R.G. (2012). FGF23 acts directly on renal proximal tubules to induce phosphaturia through activation of the ERK1/2–SGK1 signaling pathway. Bone *51*, 621-628.

Arlt, A., and Schafer, H. (2011). Role of the immediate early response 3 (IER3) gene in cellular stress response, inflammation and tumorigenesis. Eur J Cell Biol *90*, 545-552.

Armelin, H.A. (1973). Pituitary Extracts and Steroid Hormones in the Control of 3T3 Cell Growth. Proceedings of the National Academy of Sciences of the United States of America *70*, 2702-2706.

Astigarraga, S., Grossman, R., Diaz-Delfin, J., Caelles, C., Paroush, Z., and Jimenez, G. (2007a). A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. The EMBO journal *26*, 668-677.

Astigarraga, S., Grossman, R., Díaz-Delfín, J., Caelles, C., Paroush, Z., and Jiménez, G. (2007b). A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. In The EMBO journal, pp. 668-677.

Atcha, F.A., Syed, A., Wu, B., Hoverter, N.P., Yokoyama, N.N., Ting, J.H., Munguia, J.E., Mangalam, H.J., Marsh, J.L., and Waterman, M.L. (2007). A unique DNA binding domain converts T-cell factors into strong Wnt effectors. Molecular and cellular biology *27*, 8352-8363.

Atkey, M.R., Lachance, J.F., Walczak, M., Rebello, T., and Nilson, L.A. (2006). Capicua regulates follicle cell fate in the Drosophila ovary through repression of mirror. Development *133*, 2115-2123.

Babu, G.J., Lalli, M.J., Sussman, M.A., Sadoshima, J., and Periasamy, M. (2000). Phosphorylation of elk-1 by MEK/ERK pathway is necessary for c-fos gene activation during cardiac myocyte hypertrophy. Journal of molecular and cellular cardiology *32*, 1447-1457.

Baeuerle, P.A. (1991). The inducible transcription activator NF-kappa B: regulation by distinct protein subunits. Biochimica et biophysica acta *1072*, 63-80.

Barriere, G., Fici, P., Gallerani, G., Fabbri, F., and Rigaud, M. (2015). Epithelial Mesenchymal Transition: a double-edged sword. Clin Transl Med *4*, 14.

Basu-Roy, U., Ambrosetti, D., Favaro, R., Nicolis, S.K., Mansukhani, A., and Basilico, C. (2010). The transcription factor Sox2 is required for osteoblast self-renewal. Cell death and differentiation *17*, 1345-1353.

Beenken, A., and Mohammadi, M. (2009). The FGF family: biology, pathophysiology and therapy. Nature reviews Drug discovery *8*, 235-253.

Beira, J.V., and Paro, R. (2016). The legacy of Drosophila imaginal discs. Chromosoma *125*, 573-592.

Belvin, M.P., Jin, Y., and Anderson, K.V. (1995). Cactus protein degradation mediates Drosophila dorsal-ventral signaling. Genes & development *9*, 783-793.

Bertrand, N., Medevielle, F., and Pituello, F. (2000). FGF signalling controls the timing of Pax6 activation in the neural tube. Development *127*, 4837-4843.

Besnard, A., Galan-Rodriguez, B., Vanhoutte, P., and Caboche, J. (2011). Elk-1 a Transcription Factor with Multiple Facets in the Brain. Front Neurosci *5*.

Bettegowda, C., Agrawal, N., Jiao, Y., Sausen, M., Wood, L.D., Hruban, R.H., Rodriguez, F.J., Cahill, D.P., McLendon, R., Riggins, G.*, et al.* (2011a). Mutations in CIC and FUBP1 contribute to human oligodendroglioma. Science *333*, 1453-1455.

Bettegowda, C., Agrawal, N., Jiao, Y., Sausen, M., Wood, L.D., Hruban, R.H., Rodriguez, F.J., Cahill, D.P., McLendon, R., Riggins, G.*, et al.* (2011b). Mutations in CIC and FUBP1 Contribute to Human Oligodendroglioma. Science *333*, 1453-1455.

Blank, U., Brown, A., Adams, D.C., Karolak, M.J., and Oxburgh, L. (2009). BMP7 promotes proliferation of nephron progenitor cells via a JNK-dependent mechanism. Development *136*, 3557-3566.

Boch, J. (2011). TALEs of genome targeting. Nature Biotechnology *29*, 135.

Bottcher, R.T., and Niehrs, C. (2005). Fibroblast growth factor signaling during early vertebrate development. Endocrine reviews *26*, 63-77.

Boureux, A., Vignal, E., Faure, S., and Fort, P. (2007). Evolution of the Rho family of ras-like GTPases in eukaryotes. Mol Biol Evol *24*, 203-216.

Bowman, A.B., Lam, Y.C., Jafar-Nejad, P., Chen, H.K., Richman, R., Samaco, R.C., Fryer, J.D., Kahle, J.J., Orr, H.T., and Zoghbi, H.Y. (2007). Duplication of Atxn1l suppresses SCA1 neuropathology by decreasing incorporation of polyglutamine-expanded ataxin-1 into native complexes. Nat Genet *39*, 373-379.

Branney, P.A., Faas, L., Steane, S.E., Pownall, M.E., and Isaacs, H.V. (2009). Characterisation of the fibroblast growth factor dependent transcriptome in early development. PLoS One *4*, e4951.

Brantjes, H., Roose, J., van De Wetering, M., and Clevers, H. (2001). All Tcf HMG box transcription factors interact with Groucho-related co-repressors. Nucleic Acids Res *29*, 1410-1419.

Bray, S. (1998). Notch signalling in Drosophila: three ways to use a pathway. Semin Cell Dev Biol *9*, 591-597.

Brent, A.E., and Tabin, C.J. (2004). FGF acts directly on the somitic tendon progenitors through the Ets transcription factors Pea3 and Erm to regulate scleraxis expression. Development *131*, 3885-3896.

Brown, T.A., and McKnight, S.L. (1992). Specificities of protein-protein and protein-DNA interaction of GABP alpha and two newly defined ets-related proteins. Genes & development *6*, 2502-2512.

Brunner, D., Oellers, N., Szabad, J., Biggs, W.H., 3rd, Zipursky, S.L., and Hafen, E. (1994). A gain-of-function mutation in Drosophila MAP kinase activates multiple receptor tyrosine kinase signaling pathways. Cell *76*, 875-888.

Brusse, E., de Koning, I., Maat-Kievit, A., Oostra, B.A., Heutink, P., and van Swieten, J.C. (2006). Spinocerebellar ataxia associated with a mutation in the fibroblast growth factor 14 gene (SCA27): A new phenotype. Movement disorders : official journal of the Movement Disorder Society *21*, 396-401.

Buchtova, M., Oralova, V., Aklian, A., Masek, J., Vesela, I., Ouyang, Z., Obadalova, T., Konecna, Z., Spoustova, T., Pospisilova, T.*, et al.* (2015). Fibroblast growth factor and canonical WNT/beta-catenin signaling cooperate in suppression of chondrocyte

differentiation in experimental models of FGFR signaling in cartilage. Biochim Biophys Acta *1852*, 839-850.

Burks, P.J., Isaacs, H.V., and Pownall, M.E. (2009). FGF signalling modulates transcriptional repression by Xenopus groucho-related-4. Biol Cell *101*, 301-308.

Burnside, M.B., and Jacobson, A.G. (1968). Analysis of morphogenetic movements in the neural plate of the newt Taricha torosa. Developmental biology *18*, 537-552.

Bustelo, X.R., Sauzeau, V., and Berenjeno, I.M. (2007). GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo. Bioessays *29*, 356-370.

Cadigan, K.M., and Waterman, M.L. (2012). TCF/LEFs and Wnt Signaling in the Nucleus. Cold Spring Harb Perspect Biol *4*.

Cai, Z., Feng, G.S., and Zhang, X. (2010). Temporal requirement of the protein tyrosine phosphatase Shp2 in establishing the neuronal fate in early retinal development. The Journal of neuroscience : the official journal of the Society for Neuroscience *30*, 4110-4119.

Carreira, A.C., Alves, G.G., Zambuzzi, W.F., Sogayar, M.C., and Granjeiro, J.M. (2014). Bone Morphogenetic Proteins: structure, biological function and therapeutic applications. Arch Biochem Biophys *561*, 64-73.

Chan, T., Kondow, A., Hosoya, A., Hitachi, K., Yukita, A., Okabayashi, K., Nakamura, H., Ozawa, H., Kiyonari, H., Michiue, T*., et al.* (2007). Ripply2 is essential for precise somite formation during mouse early development. FEBS Lett *581*, 2691-2696.

Chen, T., Zhou, L., Yuan, Y., Fang, Y., Guo, Y., Huang, H., Zhou, Q., and Lv, X. (2014). Characterization of Bbx, a member of a novel subfamily of the HMG-box superfamily together with Cic. Development genes and evolution *224*, 261-268.

Chen, Y., Hollemann, T., Grunz, H., and Pieler, T. (1999). Characterization of the Ets-type protein ER81 in Xenopus embryos. Mechanisms of development *80*, 67-76.

Chiba, S., Lee, Y.M., Zhou, W., and Freed, C.R. (2008). Noggin enhances dopamine neuron production from human embryonic stem cells and improves behavioral outcome after transplantation into Parkinsonian rats. Stem Cells *26*, 2810-2820.

Chittaranjan, S., Chan, S., Yang, C., Yang, K.C., Chen, V., Moradian, A., Firme, M., Song, J., Go, N.E., Blough, M.D*., et al.* (2014). Mutations in CIC and IDH1 cooperatively regulate 2-hydroxyglutarate levels and cell clonogenicity. Oncotarget *5*, 7960-7979.

Chmielnicki, E., Benraiss, A., Economides, A.N., and Goldman, S.A. (2004). Adenovirally expressed noggin and brain-derived neurotrophic factor cooperate to induce new medium spiny neurons from resident progenitor cells in the adult striatal ventricular zone. J Neurosci *24*, 2133-2142.

Cho, E., Feng, Y., Rauskolb, C., Maitra, S., Fehon, R., and Irvine, K.D. (2006). Delineation of a Fat tumor suppressor pathway. Nature genetics *38*, 1142.

Cho, K.W., and De Robertis, E.M. (1990). Differential activation of Xenopus homeo box genes by mesoderm-inducing growth factors and retinoic acid. Genes Dev *4*, 1910-1916.

Chotteau-Lelièvre, A., Desbiens, X., Pelczar, H., Defossez, P.-A., and Launoit, Y.d. (1997). Differential expression patterns of the PEA3 group transcription factors through murine embryonic development. Oncogene *15*, 937.

Chow, S.Y., Yu, C.Y., and Guy, G.R. (2009). Sprouty2 interacts with protein kinase C delta and disrupts phosphorylation of protein kinase D1. The Journal of biological chemistry *284*, 19623-19636.

Christen, B., and Slack, J.M. (1997a). FGF-8 is associated with anteroposterior patterning and limb regeneration in Xenopus. Developmental biology *192*, 455-466.

Christen, B., and Slack, J.M. (1999). Spatial response to fibroblast growth factor signalling in Xenopus embryos. Development *126*, 119-125.

Christen, B., and Slack, J.M.W. (1997b). FGF-8 is associated with anteroposterior patterning and limb regeneration in Xenopus. Developmental biology *192*, 455-466.

Christian, J.L., and Moon, R.T. (1993a). Interactions between Xwnt-8 and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of Xenopus. Genes Dev *7*, 13-28.

Christian, J.L., and Moon, R.T. (1993b). Interactions between Xwnt-8 and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of Xenopus.

Ciapponi, L., Jackson, D.B., Mlodzik, M., and Bohmann, D. (2001). Drosophila Fos mediates ERK and JNK signals via distinct phosphorylation sites. Genes Dev *15*, 1540-1553.

Cinnamon, E., and Paroush, Z. (2008). Context-dependent regulation of Groucho/TLE-mediated repression. Current opinion in genetics & development *18*, 435-440.

Clément, A., Solnica-Krezel, L., and Gould, K.L. (2011). The Cdc14B phosphatase contributes to ciliogenesis in zebrafish. Development *138*, 291-302.

Coebergh, J.A., Fransen van de Putte, D.E., Snoeck, I.N., Ruivenkamp, C., van Haeringen, A., and Smit, L.M. (2014). A new variable phenotype in spinocerebellar ataxia 27 (SCA 27) caused by a deletion in the FGF14 gene. European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society *18*, 413-415.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X.*, et al.* (2016). A survey of best practices for RNA-seq data analysis. Genome Biol *17*.

Consortium, A. (2000). Autosomal dominant hypophosphataemic rickets is associated with mutations in FGF23. Nature genetics *26*, 345-348.

Cooke, J. (1989). Xenopus mesoderm induction: evidence for early size control and partial autonomy for pattern development by onset of gastrulation. Development *106*, 519-529.

Cornell, R.A., Musci, T.J., and Kimelman, D. (1995). FGF is a prospective competence factor for early activin-type signals in Xenopus mesoderm induction. Development.

Coutu, D.L., and Galipeau, J. (2011). Roles of FGF signaling in stem cell self-renewal, senescence and aging. Aging *3*, 920-933.

Crespo-Barreto, J., Fryer, J.D., Shaw, C.A., Orr, H.T., and Zoghbi, H.Y. (2010). Partial loss of ataxin-1 function contributes to transcriptional dysregulation in spinocerebellar ataxia type 1 pathogenesis. PLoS Genet *6*, e1001021.

Cunningham, T.J., Kumar, S., Yamaguchi, T.P., and Duester, G. (2015). Wnt8a and Wnt3a Cooperate in the Axial Stem Cell Niche to Promote Mammalian Body Axis Extension. Dev Dyn *244*, 797-807.

Dailey, L., Ambrosetti, D., Mansukhani, A., and Basilico, C. (2005). Mechanisms underlying differential responses to FGF signaling. Cytokine Growth Factor Rev *16*, 233-247.

Dale, J.K., Vesque, C., Lints, T.J., Sampath, T.K., Furley, A., Dodd, J., and Placzek, M. (1997). Cooperation of BMP7 and SHH in the induction of forebrain ventral midline cells by prechordal mesoderm. Cell *90*, 257-269.

Davidson, B., Goldberg, I., Gotlieb, W.H., Kopolovic, J., Ben-Baruch, G., and Reich, R. (2003). PEA3 is the second Ets family transcription factor involved in tumor progression in ovarian carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research *9*, 1412-1419.

Davis, L.K., Meyer, K.J., Rudd, D.S., Librant, A.L., Epping, E.A., Sheffield, V.C., and Wassink, T.H. (2008). Pax6 3′ deletion results in aniridia, autism and mental retardation. Hum Genet *123*, 371-378.

Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns to Shape the Cancer Genome. Cell *155*, 948-962.

de las Heras, J.M., and Casanova, J. (2006). Spatially distinct downregulation of Capicua repression and tailless activation by the Torso RTK pathway in the Drosophila embryo. Mechanisms of development *123*, 481-486.

de Launoit, Y., Baert, J.-L., Chotteau-Lelievre, A., Monte, D., Coutte, L., Mauen, S., Firlej, V., Degerny, C., and Verreman, K. (2006). The Ets transcription factors of the PEA3 group: Transcriptional regulators in metastasis. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer *1766*, 79-87.

de Launoit, Y., Chotteau-Lelievre, A., Beaudoin, C., Coutte, L., Netzer, S., Brenner, C., Huvent, I., and Baert, J.L. (2000). The PEA3 group of ETS-related transcription factors. Role in breast cancer metastasis. Advances in experimental medicine and biology *480*, 107-116.

Degoutin, J.L., Milton, C.C., Yu, E., Tipping, M., Bosveld, F., Yang, L., Bellaiche, Y., Veraksa, A., and Harvey, K.F. (2013). Riquiqui and minibrain are regulators of the hippo pathway downstream of Dachsous. Nature cell biology *15*, 1176-1185.

Deimling, S.J., and Drysdale, T.A. (2011). Fgf is required to regulate anterior-posterior patterning in the Xenopus lateral plate mesoderm. Mechanisms of development *128*, 327-341.

Delaune, E., Lemaire, P., and Kodjabachian, L. (2005). Neural induction in Xenopus requires early FGF signalling inaddition to BMP inhibition.

Delclaux, C., Delacourt, C., D'Ortho, M.P., Boyer, V., Lafuma, C., and Harf, A. (2012). Role of gelatinase B and elastase in human polymorphonuclear neutrophil migration across basement membrane. http://dxdoiorg/101165/ajrcmb1438845180.

Delfini, M.-C., Dubrulle, J., Malapert, P., Chal, J., and Pourquié, O. (2005). Control of the segmentation process by graded MAPK/ERK activation in the chick embryo. PNAS.

Dequeant, M.L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A., and Pourquie, O. (2006). A complex oscillating network of signaling genes underlies the mouse segmentation clock. Science *314*, 1595-1598.

Dequeant, M.L., and Pourquie, O. (2008). Segmental patterning of the vertebrate embryonic axis. Nat Rev Genet *9*, 370-382.

Diez del Corral, R., Breitkreuz, D.N., and Storey, K.G. (2002). Onset of neuronal differentiation is regulated by paraxial mesoderm and requires attenuation of FGF signalling. Development *129*, 1681-1691.

Ding, J., Yang, L., Yan, Y.T., Chen, A., Desai, N., Wynshaw-Boris, A., and Shen, M.M. (1998). Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. Nature *395*, 702-707.

Dissanayake, K., Toth, R., Blakey, J., Olsson, O., Campbell, D.G., Prescott, A.R., and MacKintosh, C. (2011). ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor capicua. The Biochemical journal *433*, 515-525.

Dode, C., Levilliers, J., Dupont, J.M., De Paepe, A., Le Du, N., Soussi-Yanicostas, N., Coimbra, R.S., Delmaghani, S., Compain-Nouaille, S., Baverel, F.*, et al.* (2003). Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. Nature genetics *33*, 463-465.

Dogan, R.I., Getoor, L., Wilbur, W.J., and Mount, S.M. (2007). SplicePort--an interactive splice-site analysis tool. Nucleic acids research *35*, W285-291.

Dudley, A.T., Godin, R.E., and Robertson, E.J. (1999). Interaction between FGF and BMP signaling pathways regulates development of metanephric mesenchyme. Genes Dev *13*, 1601-1613.

Dudley, A.T., Lyons, K.M., and Robertson, E.J. (1995). A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. Genes Dev *9*, 2795-2807.

Dudley, A.T., and Robertson, E.J. (1997). Overlapping expression domains of bone morphogenetic protein family members potentially account for limited tissue defects in BMP7 deficient embryos. Dev Dyn *208*, 349-362.

Duester, G. (2008). Retinoic Acid Synthesis and Signaling during Early Organogenesis. Cell *134*, 921-931.

Duffy, J.B., and Perrimon, N. (1994). The torso pathway in Drosophila: lessons on receptor tyrosine kinase signaling and pattern formation. Developmental biology *166*, 380-395.

Edgar, B.A., and Schubiger, G. (1986). Parameters controlling transcriptional activation during early Drosophila development. Cell *44*, 871-877.

Esch, F., Baird, A., Ling, N., Ueno, N., Hill, F., Denoroy, L., Klepper, R., Gospodarowicz, D., Böhlen, P., and Guillemin, R. (1985). Primary structure of bovine pituitary basic fibroblast growth factor (FGF) and comparison with the amino-terminal sequence of bovine brain acidic FGF.

Eswarakumar, V.P., Monsonego-Ornan, E., Pines, M., Antonopoulou, I., Morriss-Kay, G.M., and Lonai, P. (2002). The IIIc alternative of Fgfr2 is a positive regulator of bone formation. Development *129*, 3783-3793.

Everts, V., Delaisse, J.M., Korper, W., Niehof, A., Vaes, G., and Beertsen, W. (1992). Degradation of collagen in the bone-resorbing compartment underlying the osteoclast involves both cysteine-proteinases and matrix metalloproteinases. J Cell Physiol *150*, 221-231.

Falardeau, J., Chung, W.C., Beenken, A., Raivio, T., Plummer, L., Sidis, Y., Jacobson-Dickman, E.E., Eliseenkova, A.V., Ma, J., Dwyer, A.*, et al.* (2008). Decreased FGF8 signaling causes deficiency of gonadotropin-releasing hormone in humans and mice. The Journal of clinical investigation *118*, 2822-2831.

Fatland, B., Anderson, M., Nikolau, B.J., and Wurtele, E.S. (2000). Molecular biology of cytosolic acetyl-CoA generation. Biochemical Society transactions *28*, 593-595.

Fisher, M.E., Isaacs, H.V., and Pownall, M.E. (2002a). eFGF is required for activation of XmyoD expression in the myogenic cell lineage of Xenopus laevis.

Fisher, M.E., Isaacs, H.V., and Pownall, M.E. (2002b). eFGF is required for activation of XmyoD expression in the myogenic cell lineage of Xenopus laevis. Development *129*, 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*, 1703-1714.

Fletcher, R.B., and Harland, R.M. (2008). The role of FGF signaling in the establishment and maintenance of mesodermal gene expression in Xenopus. Developmental dynamics : an official publication of the American Association of Anatomists *237*, 1243-1254.

Fores, M., Ajuria, L., Samper, N., Astigarraga, S., Nieva, C., Grossman, R., Gonzalez-Crespo, S., Paroush, Z., and Jimenez, G. (2015). Origins of context-dependent gene repression by capicua. PLoS Genet *11*, e1004902.

Fores, M., Simon-Carrasco, L., Ajuria, L., Samper, N., Gonzalez-Crespo, S., Drosten, M., Barbacid, M., and Jimenez, G. (2017). A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer. PLoS genetics *13*, e1006622.

Fredieu, J.R., Cui, Y., Maier, D., Danilchik, M.V., and Christian, J.L. (1997). Xwnt-8 and lithium can act upon either dorsal mesodermal or neurectodermal cells to cause a loss of forebrain in Xenopus embryos. Dev Biol *186*, 100-114.

Fryer, J.D., Yu, P., Kang, H., Mandel-Brehm, C., Carter, A.N., Crespo-Barreto, J., Gao, Y., Flora, A., Shaw, C., Orr, H.T.*, et al.* (2011). Exercise and genetic rescue of SCA1 via the transcriptional repressor Capicua. Science *334*, 690-693.

Fu, H., Subramanian, R.R., and Masters, S.C. (2000). 14-3-3 proteins: structure, function, and regulation. Annu Rev Pharmacol Toxicol *40*, 617-647.

Fuchs, A., Cheung, L.S., Charbonnier, E., Shvartsman, S.Y., and Pyrowolakis, G. (2012). Transcriptional interpretation of the EGF receptor signaling gradient. Proceedings of the National Academy of Sciences of the United States of America *109*, 1572-1577.

Fuhrmann, S. (2010). Eye morphogenesis and patterning of the optic vesicle. Current topics in developmental biology *93*, 61-84.

Furriols, M., and Casanova, J. (2003). In and out of Torso RTK signalling. The EMBO journal *22*, 1947-1952.

Furthauer, M., Van Celst, J., Thisse, C., and Thisse, B. (2004). Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. Development *131*, 2853-2864.

Futran, A.S., Kyin, S., Shvartsman, S.Y., and Link, A.J. (2015). Mapping the binding interface of ERK and transcriptional repressor Capicua using photocrosslinking. Proceedings of the National Academy of Sciences of the United States of America *112*, 8590-8595.

Gabay, L., Seger, R., and Shilo, B.Z. (1997). MAP kinase in situ activation atlas during Drosophila embryogenesis. Development *124*, 3535-3541.

Garcia, J., Ye, Y., Arranz, V., Letourneux, C., Pezeron, G., and Porteu, F. (2002). IEX-1: a new ERK substrate involved in both ERK survival activity and ERK activation. Embo j *21*, 5151-5163.

Gehring, W.J. (1996). The master control gene for morphogenesis and evolution of the eye. Genes Cells *1*, 11-15.

Gille, H., Kortenjann, M., Thomae, O., Moomaw, C., Slaughter, C., Cobb, M.H., and Shaw, P.E. (1995). ERK phosphorylation potentiates Elk-1-mediated ternary complex formation and transactivation. Embo j *14*, 951-962.

Gish, W., and States, D.J. (1993). Identification of protein coding regions by database similarity search. Nature genetics *3*, 266-272.

Goff, D.J., Nilson, L.A., and Morisato, D. (2001). Establishment of dorsal-ventral polarity of the Drosophila egg requires capicua action in ovarian follicle cells. Development *128*, 4553-4562.

Gomez-Skarmeta, J.L., Diez del Corral, R., de la Calle-Mustienes, E., Ferre-Marco, D., and Modolell, J. (1996). Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. Cell *85*, 95-105.

Gospodarowicz, D. (1975). Purification of a fibroblast growth factor from bovine pituitary. The Journal of biological chemistry *250*, 2515-2520.

Gospodarowicz, D., Cheng, J., Lui, G.M., Baird, A., and Böhlent, P. (1984). Isolation of brain fibroblast growth factor by heparin-Sepharose affinity chromatography: identity with pituitary fibroblast growth factor.

Gospodarowicz, D., and Moran, J.S. (1975). Mitogenic effect of fibroblast growth factor on early passage cultures of human and murine fibroblasts. J Cell Biol *66*, 451-457.

Goulev, Y., Department of Developmental Biology, I.J.M., Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7592, Université Paris 7 Denis-Diderot, Tour 43 2, Place Jussieu, F-75251 Paris Cedex 05, France, Fauny, J.D., Department of Developmental Biology, I.J.M., Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7592, Université Paris 7 Denis-Diderot, Tour 43 2, Place Jussieu, F-75251 Paris Cedex 05, France, Gonzalez-Marti, B., Department of Developmental Biology, I.J.M., Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7592, Université Paris 7 Denis-Diderot, Tour 43 2, Place Jussieu, F-75251 Paris Cedex 05, France, Flagiello, D., Department of Developmental Biology, I.J.M., Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7592, Université Paris 7 Denis-Diderot, Tour 43 2, Place Jussieu, F-75251 Paris Cedex 05, France, Silber, J., Department of Developmental Biology, I.J.M., Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7592, Université Paris 7 Denis-Diderot, Tour 43 2, Place Jussieu, F-75251 Paris Cedex 05, France*, et al.* (2008). SCALLOPED Interacts with YORKIE, the Nuclear Effector of the Hippo Tumor-Suppressor Pathway in Drosophila. Current Biology *18*, 435-441.

Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. The Journal of general virology *36*, 59-74.

Graves, B.J., and Petersen, J.M. (1998). Specificity within the ets family of transcription factors. Adv Cancer Res *75*, 1-55.

Greber, B., Coulon, P., Zhang, M., Moritz, S., Frank, S., Müller-Molina, A.J., Araúzo-Bravo, M.J., Han, D.W., Pape, H.C., and Schöler, H.R. (2011). FGF signalling inhibits neural induction in human embryonic stem cells. EMBO J *30*, 4874-4884.

Green, P.J., Walsh, F.S., and Doherty, P. (1996). Promiscuity of fibroblast growth factor receptors. BioEssays : news and reviews in molecular, cellular and developmental biology *18*, 639-646.

Greenwood, S., and Struhl, G. (1997). Different levels of Ras activity can specify distinct transcriptional and morphological consequences in early Drosophila embryos. Development *124*, 4879-4886.

Grimm, O., Sanchez Zini, V., Kim, Y., Casanova, J., Shvartsman, S.Y., and Wieschaus, E. (2012). Torso RTK controls Capicua degradation by changing its subcellular localization. Development *139*, 3962-3968.

Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W.S., and Schier, A.F. (1999). The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. Cell *97*, 121-132.

Guo, R.J., Suh, E.R., and Lynch, J.P. (2004). The role of Cdx proteins in intestinal development and cancer. Cancer biology & therapy *3*, 593-601.

Hai, T., and Curran, T. (1991). Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc Natl Acad Sci U S A *88*, 3720-3724.

Hansson, M., Olesen, D.R., Peterslund, J.M.L., Engberg, N., Kahn, M., Winzi, M., Klein, T., Maddox-Hyttel, P., and Serup, P. (2009). A late requirement for Wnt and FGF signaling during activin-induced formation of foregut endoderm from mouse embryonic stem cells. Dev Biol *330*, 286-304.

Hasson, P., Egoz, N., Winkler, C., Volohonsky, G., Jia, S., Dinur, T., Volk, T., Courey, A.J., and Paroush, Z. (2005). EGFR signaling attenuates Groucho-dependent repression to antagonize Notch transcriptional output. Nat Genet *37*, 101-105.

Heasman, J., Kofron, M., and Wylie, C. (2000). Beta-catenin signaling activity dissected in the early Xenopus embryo: a novel antisense approach. Developmental biology *222*, 124-134.

Heasman, S.J., and Ridley, A.J. (2008). Mammalian Rho GTPases: new insights into their functions from in vivo studies. Nat Rev Mol Cell Biol *9*, 690-701.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L.*, et al.* (2010). The genome of the Western clawed frog Xenopus tropicalis. Science *328*, 633-636.

Herranz, H., Hong, X., and Cohen, S.M. (2012a). Mutual repression by bantam miRNA and Capicua links the EGFR/MAPK and Hippo pathways in growth control. Current biology : CB *22*, 651-657.

Herranz, H., Institute of Molecular and Cell Biology, B.D., Singapore 138673, Singapore, Hong, X., Institute of Molecular and Cell Biology, B.D., Singapore 138673, Singapore, Department of Biological Sciences, N.U.o.S., Singapore 117543, Singapore, scohen@imcb.a-star.edu.sg, Institute of Molecular and Cell Biology, B.D., Singapore 138673, Singapore, and Department of Biological Sciences, N.U.o.S., Singapore 117543, Singapore (2012b). Mutual Repression by Bantam miRNA and Capicua Links the EGFR/MAPK and Hippo Pathways in Growth Control. Current Biology *22*, 651-657.

Herriges, J.C., Verheyden, J.M., Zhang, Z., Sui, P., Zhang, Y., Anderson, M.J., Swing, D.A., Zhang, Y., Lewandoski, M., and Sun, X. (2015). FGF-Regulated ETV Transcription Factors Control FGF-SHH Feedback Loop in Lung Branching. Dev Cell *35*, 322-332.

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A.T., and Neugebauer, K.M. (2014). The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. Cell Reports *6*, 285-292.

Holland, P.W., and Garcia-Fernandez, J. (1996). Hox genes and chordate evolution. Dev Biol *173*, 382-395.

Hoppler, S., Brown, J.D., and Moon, R.T. (1996). Expression of a dominant-negative Wnt blocks induction of MyoD in Xenopus embryos. Genes Dev *10*, 2805-2817.

Horiuchi, S., Yamamoto, H., Min, Y., Adachi, Y., Itoh, F., and Imai, K. (2003). Association of ets-related transcriptional factor E1AF expression with tumour progression and overexpression of MMP-1 and matrilysin in human colorectal cancer. The Journal of pathology *200*, 568-576.

Hou, R., Liu, L., Anees, S., Hiroyasu, S., and Sibinga, N.E. (2006). The Fat1 cadherin integrates vascular smooth muscle cell growth and migration signals. J Cell Biol *173*, 417-429.

Huang, J., Department of Physiology, U.o.T.S.M.C.a.D., 5323 Harry Hines Boulevard, Dallas, Texas 75390, Wu, S., Department of Physiology, U.o.T.S.M.C.a.D., 5323 Harry Hines Boulevard, Dallas, Texas 75390, Barrera, J., Department of Physiology, U.o.T.S.M.C.a.D., 5323 Harry Hines Boulevard, Dallas, Texas 75390, Matthews, K., Department of Physiology, U.o.T.S.M.C.a.D., 5323 Harry Hines Boulevard, Dallas, Texas 75390, Pan, D., djpan@jhmi.edu*, et al.* (2005). The Hippo Signaling Pathway Coordinately Regulates Cell Proliferation and Apoptosis by Inactivating Yorkie, the Drosophila Homolog of YAP. Cell *122*, 421-434.

Hudson, C., Clements, D., Friday, R.V., Stott, D., and Woodland, H.R. (1997). Xsox17alpha and -beta mediate endoderm formation in Xenopus. Cell *91*, 397-405.

Hughes, C., Rabinowitz, A., Tate, M., Birrell, L., Allsup, J., Billinton, N., and Walmsley, R.M. (2012). Development of a high-throughput Gaussia luciferase reporter assay for the activation of the GADD45a gene by mutagens, promutagens, clastogens, and aneugens. J Biomol Screen *17*, 1302-1315.

Irie, N., Weinberger, L., Tang, W., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J., and Surani, M. (2015). SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. Cell *160*, 253-268.

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*, 165-173.

Isaacs, H.V., Pownall, M.E., and Slack, J.M. (1994). eFGF regulates Xbra expression during Xenopus gastrulation. The EMBO journal *13*, 4469-4481.

Isaacs, H.V., Pownall, M.E., and Slack, J.M. (1995a). eFGF is expressed in the dorsal midline of Xenopus laevis. Int J Dev Biol *39*, 575-579.

Isaacs, H.V., Pownall, M.E., and Slack, J.M. (1998). Regulation of Hox gene expression and posterior development by the Xenopus caudal homologue Xcad3. Embo j *17*, 3413-3427.

Isaacs, H.V., Pownall, M.E., and Slack, J.M.W. (1995b). Efgf Is Expressed in the Dorsal Midline of Xenopus-Laevis. Int J Dev Biol *39*, 575-579.

Isaacs, H.V., Tannahill, D., and Slack, J.M. (1992). Expression of a novel FGF in the Xenopus embryo. A new candidate inducing factor for mesoderm formation and anteroposterior specification. Development *114*, 711-720.

Itoh, N., Ohta, H., and Konishi, M. (2015). Endocrine FGFs: Evolution, Physiology, Pathophysiology, and Pharmacotherapy. Front Endocrinol (Lausanne) *6*, 154.

Itoh, N., and Ornitz, D.M. (2004). Evolution of the Fgf and Fgfr gene families. Trends in genetics : TIG *20*, 563-569.

James, S., Fox, J., Afsari, F., Lee, J., Clough, S., Knight, C., Ashmore, J., Ashton, P., Preham, O., Hoogduijn, M.*, et al.* (2015). Multiparameter Analysis of Human Bone Marrow Stromal Cells Identifies Distinct Immunomodulatory and Differentiation-Competent Subtypes. Stem Cell Reports *4*, 1004-1015.

Jennings, B.H., and Ish-Horowicz, D. (2008). The Groucho/TLE/Grg family of transcriptional co-repressors. Genome biology 9, 205.

Jiang, H., Grenley, M.O., Bravo, M.J., Blumhagen, R.Z., and Edgar, B.A. (2011). EGFR/Ras/MAPK signaling mediates adult midgut epithelial homeostasis and regeneration in Drosophila. Cell stem cell 8, 84-95.

Jiang, Z., Guerrero-Netro, H.M., Juengel, J.L., and Price, C.A. (2013). Divergence of intracellular signaling pathways and early response genes of two closely related fibroblast growth factors, FGF8 and FGF18, in bovine ovarian granulosa cells. Mol Cell Endocrinol 375, 97-105.

Jimenez, G., Gonzalez-Reyes, A., and Casanova, J. (2002). Cell surface proteins Nasrat and Polehole stabilize the Torso-like extracellular determinant in Drosophila oogenesis. Genes & development 16, 913-918.

Jimenez, G., Guichet, A., Ephrussi, A., and Casanova, J. (2000). Relief of gene repression by torso RTK signaling: role of capicua in Drosophila terminal and dorsoventral patterning. Genes & development 14, 224-231.

Jimenez, G., Shvartsman, S.Y., and Paroush, Z. (2012). The Capicua repressor--a general sensor of RTK signaling in development and disease. Journal of cell science 125, 1383-1391.

Joung, J.K., and Sander, J.D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nature reviews Molecular cell biology 14, 49-55.

Karim, F.D., and Rubin, G.M. (1998). Ectopic expression of activated Ras1 induces hyperplastic growth and increased cell death in Drosophila imaginal tissues. Development 125, 1-9.

Karimi, K., Fortriede, J.D., Lotay, V.S., Burns, K.A., Wang, D.Z., Fisher, M.E., Pells, T.J., James-Zorn, C., Wang, Y., Ponferrada, V.G., et al. (2018). Xenbase: a genomic, epigenomic and transcriptomic model organism database. Nucleic acids research 46, D861-D868.

Kawamura-Saito, M., Department of Carcinogenesis, Department of Molecular Pathophysiology, T.M.a.D.U., Tokyo, Japan, Yamazaki, Y., Department of Carcinogenesis, Kaneko, K., Department of Carcinogenesis, Department of Molecular Pathophysiology, T.M.a.D.U., Tokyo, Japan, Kawaguchi, N., Department of Orthopaedic Oncology, et al. (2006). Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Human Molecular Genetics 15, 2125-2137.

Keenan, I.D., Sharrard, R.M., and Isaacs, H.V. (2006a). FGF signal transduction and the regulation of Cdx gene expression. Dev Biol 299, 478-488.

Keenan, I.D., Sharrard, R.M., and Isaacs, H.V. (2006b). FGF signal transduction and the regulation of Cdx gene expression. Dev Biol 299, 478-488.

Kiecker, C., Muller, F., Wu, W., Glinka, A., Strahle, U., and Niehrs, C. (2000). Phenotypic effects in Xenopus and zebrafish suggest that one-eyed pinhead functions as antagonist of BMP signalling. Mech Dev 94, 37-46.

Kim, E., Kim, D., Lee, J.S., Yoe, J., Park, J., Kim, C.J., Jeong, D., Kim, S., and Lee, Y. (2018). Capicua suppresses hepatocellular carcinoma progression by controlling the ETV4-MMP1 axis. Hepatology 67, 2287-2301.

Kim, E., Lu, H.C., Zoghbi, H.Y., and Song, J.J. (2013). Structural basis of protein complex formation and reconfiguration by polyglutamine disease protein Ataxin-1 and Capicua. Genes & development 27, 590-595.

Kim, J., Lin, J.-J., Xu, R.-H., and Kung, H.-f. (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.

Klingler, M., Erdelyi, M., Szabad, J., and Nusslein-Volhard, C. (1988). Function of torso in determining the terminal anlagen of the Drosophila embryo. Nature 335, 275-277.

Knöchel, S., Schuler-Metz, A., and Knöchel, W. (2000). c-Jun (AP-1) activates BMP-4 transcription in Xenopus embryos. Mechanisms of development *98*, 29-36.

Kondrashov, F.A., and Koonin, E.V. (2003). Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. Trends Genet *19*, 115-119.

Kouhara, H., Hadari, Y.R., SpivakKroizman, T., Schilling, J., BarSagi, D., Lax, I., and Schlessinger, J. (1997). A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. Cell *89*, 693-702.

Kowtharapu, B.S., Prakasam, R.K., Murín, R., Koczan, D., Stahnke, T., Wree, A., Jünemann, A.G.M., and Stachs, O. (2018). Role of Bone Morphogenetic Protein 7 (BMP7) in the Modulation of Corneal Stromal and Epithelial Cell Functions. Int J Mol Sci *19*.

Kretzschmar, M., Doody, J., and Massague, J. (1997). Opposing BMP and EGF signalling pathways converge on the TGF-beta family mediator Smad1. Nature *389*, 618-622.

Kumano, G., and Smith, W.C. (2002). Revisions to the Xenopus gastrula fate map: implications for mesoderm induction and patterning. Developmental dynamics : an official publication of the American Association of Anatomists *225*, 409-421.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895-2902.

Kwan, K.M., and Kirschner, M.W. (2003). Xbra functions as a switch between cell migration and convergentextension in the Xenopus gastrula.

LaBonne, C., and Whitman, M. (1997). Localization of MAP kinase activity in early Xenopus embryos: implications for endogenous FGF signaling. Dev Biol *183*, 9-20.

Lam, Y.C., Bowman, A.B., Jafar-Nejad, P., Lim, J., Richman, R., Fryer, J.D., Hyun, E.D., Duvick, L.A., Orr, H.T., Botas, J*., et al.* (2006). ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. Cell *127*, 1335-1347.

Lang, F., Strutz-Seebohm, N., Seebohm, G., and Lang, U.E. (2010). Significance of SGK1 in the regulation of neuronal function. J Physiol *588*, 3349-3354.

Lea, R., Papalopulu, N., Amaya, E., and Dorey, K. (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*, 1467-1479.

Lee, C.J., Chan, W.I., Cheung, M., Cheng, Y.C., Appleby, V.J., Orme, A.T., and Scotting, P.J. (2002). CIC, a member of a novel subfamily of the HMG-box superfamily, is transiently expressed in developing granule neurons. Brain research Molecular brain research *106*, 151-156.

Lee, Y., Fryer, J.D., Kang, H., Crespo-Barreto, J., Bowman, A.B., Gao, Y., Kahle, J.J., Hong, J.S., Kheradmand, F., Orr, H.T.*, et al.* (2011). ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung alveolarization. Dev Cell *21*, 746-757.

Lei, Y., Guo, X., Liu, Y., Cao, Y., Deng, Y., Chen, X., Cheng, C.H., Dawid, I.B., Chen, Y., and Zhao, H. (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*, 17484-17489.

Letourneux, C., Rocher, G., and Porteu, F. (2006). B56-containing PP2A dephosphorylate ERK and their activity is controlled by the early gene IEX-1 and ERK. Embo j *25*, 727-738.

Leyns, L., Bouwmeester, T., Kim, S.H., Piccolo, S., and De Robertis, E.M. (1997). Frzb-1 Is a Secreted Antagonist of Wnt Signaling Expressed in the Spemann Organizer. Cell *88*, 747-756.

Li, S., Mattar, P., Dixit, R., Lawn, S.O., Wilkinson, G., Kinch, C., Eisenstat, D., Kurrasch, D.M., Chan, J.A., and Schuurmans, C. (2014). RAS/ERK signaling controls proneural genetic programs in cortical development and gliomagenesis. The Journal of neuroscience : the official journal of the Society for Neuroscience *34*, 2169-2190.

Li, X., Newbern, J., Wu, Y., Morgan-Smith, M., Zhong, J., Charron, J., and Snider, W. (2012). MEK is a key regulator of gliogenesis in the developing brain. Neuron *75*, 1035-1050.

Liebermann, D.A., and Hoffman, B. (2008). Gadd45 in stress signaling. J Mol Signal *3*, 15.

Lim, M., Chuong, C.M., and Roy-Burman, P. (2011). PI3K, Erk signaling in BMP7-induced epithelial-mesenchymal transition (EMT) of PC-3 prostate cancer cells in 2- and 3-dimensional cultures. Horm Cancer *2*, 298-309.

Lindlof, A. (2003). Gene identification through large-scale EST sequence processing. Applied bioinformatics *2*, 123-129.

Lories, R.J., Peeters, J., Bakker, A., Tylzanowski, P., Derese, I., Schrooten, J., Thomas, J.T., and Luyten, F.P. (2007). Articular cartilage and biomechanical properties of the long bones in Frzb-knockout mice. Arthritis Rheum *56*, 4095-4103.

Louro, R., Nakaya, H.I., Paquola, A.C., Martins, E.A., da Silva, A.M., Verjovski-Almeida, S., and Reis, E.M. (2004). RASL11A, member of a novel small monomeric GTPase gene family, is down-regulated in prostate tumors. Biochem Biophys Res Commun *316*, 618-627.

Lu, H.C., Tan, Q., Rousseaux, M.W., Wang, W., Kim, J.Y., Richman, R., Wan, Y.W., Yeh, S.Y., Patel, J.M., Liu, X.*, et al.* (2017a). Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes in mice and humans. Nature genetics *49*, 527-536.

Lu, H.C., Tan, Q., Rousseaux, M.W., Wang, W., Kim, J.Y., Richman, R., Wan, Y.W., Yeh, S.Y., Patel, J.M., Liu, X.*, et al.* (2017b). Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes in mice and humans. Nature genetics.

Luo, G., Hofmann, C., Bronckers, A.L., Sohocki, M., Bradley, A., and Karsenty, G. (1995). BMP-7 is an inducer of nephrogenesis, and is also required for eye development and skeletal patterning. Genes Dev *9*, 2808-2820.

Lyapina, S., Cope, G., Shevchenko, A., Serino, G., Tsuge, T., Zhou, C., Wolf, D.A., Wei, N., Shevchenko, A., and Deshaies, R.J. (2001). Promotion of NEDD8-CUL1 Conjugate Cleavage by COP9 Signalosome.

MacDonald, B.T., Tamai, K., and He, X. (2009). Wnt/beta-catenin signaling: components, mechanisms, and diseases. Dev Cell *17*, 9-26.

MacNicol, A.M., Muslin, A.J., Howard, E.L., Kikuchi, A., MacNicol, M.C., and Williams, L.T. (1995). Regulation of Raf-1-dependent signaling during early Xenopus development. Mol Cell Biol *15*, 6686-6693.

Maeda, M., Hasegawa, H., Hyodo, T., Ito, S., Asano, E., Yuang, H., Funasaka, K., Shimokata, K., Hasegawa, Y., Hamaguchi, M.*, et al.* (2011). ARHGAP18, a GTPase-activating protein for RhoA, controls cell shape, spreading, and motility. Mol Biol Cell *22*, 3840-3852.

Mahoney Rogers, A.A., Zhang, J., and Shim, K. (2011). Sprouty1 and Sprouty2 limit both the size of the otic placode and hindbrain Wnt8a by antagonizing FGF signaling. Dev Biol *353*, 94-104.

Marchal, L., Luxardi, G., Thome, V., and Kodjabachian, L. (2009). BMP inhibition initiates neural induction via FGF signaling and Zic genes. Proceedings of the National Academy of Sciences of the United States of America *106*, 17437-17442.

Martinez-Morales, J.R., Del Bene, F., Nica, G., Hammerschmidt, M., Bovolenta, P., and Wittbrodt, J. (2005). Differentiation of the vertebrate retina is coordinated by an FGF signaling center. Developmental cell *8*, 565-574.

Matsuda, T., and Cepko, C.L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc Natl Acad Sci U S A *101*, 16-22.

McFarlane, S., Zuber, M.E., and Holt, C.E. (1998). A role for the fibroblast growth factor receptor in cell fate decisions in the developing vertebrate retina. Development *125*, 3967-3975.

McGrew, L.L., Hoppler, S., and Moon, R.T. (1997). Wnt and FGF pathways cooperatively pattern anteroposterior neural ectoderm in Xenopus. Mech Dev *69*, 105-114.

Merlet, J., Burger, J., Gomes, J.E., and Pintard, L. (2009). Regulation of cullin-RING E3 ubiquitin-ligases by neddylation and dimerization. Cellular and molecular life sciences : CMLS *66*, 1924-1938.

Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res *41*, D377-386.

Milunsky, J.M., Zhao, G., Maher, T.A., Colby, R., and Everman, D.B. (2006). LADD syndrome is caused by FGF10 mutations. Clinical genetics *69*, 349-354.

Mizutani, A., Wang, L., Rajan, H., Vig, P.J., Alaynick, W.A., Thaler, J.P., and Tsai, C.C. (2005). Boat, an AXH domain protein, suppresses the cytotoxicity of mutant ataxin-1.

Mohammadi, M., Honegger, A.M., Rotin, D., Fischer, R., Bellot, F., Li, W., Dionne, C.A., Jaye, M., Rubinstein, M., and Schlessinger, J. (1991). A tyrosine-phosphorylated carboxy-terminal peptide of the fibroblast growth factor receptor (Flg) is a binding site for the SH2 domain of phospholipase C-gamma 1. Molecular and cellular biology *11*, 5068-5078.

Mohammadi, M., McMahon, G., Sun, L., Tang, C., Hirth, P., Yeh, B.K., Hubbard, S.R., and Schlessinger, J. (1997). Structures of the Tyrosine Kinase Domain of Fibroblast Growth Factor Receptor in Complex with Inhibitors.

Monte, D., Baert, J.L., Defossez, P.A., de Launoit, Y., and Stehelin, D. (1994). Molecular cloning and characterization of human ERM, a new member of the Ets family closely related to mouse PEA3 and ER81 transcription factors. Oncogene *9*, 1397-1406.

Moore, J.K., and Haber, J.E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Molecular and cellular biology *16*, 2164-2173.

Moussian, B., and Roth, S. (2005). Dorsoventral axis formation in the Drosophila embryo--shaping and transducing a morphogen gradient. Current biology : CB *15*, R887-899.

Murphy, L.O., Smith, S., Chen, R.H., Fingar, D.C., and Blenis, J. (2002). Molecular interpretation of ERK signal duration by immediate early gene products. Nat Cell Biol *4*, 556-564.

Muthukrishnan, S.D., Yang, X., Friesel, R., and Oxburgh, L. (2015). Concurrent BMP7 and FGF9 signalling governs AP-1 function to promote self-renewal of nephron progenitor cells. Nat Commun *6*, 10027.

Muñoz-Sanjuán, I., and Brivanlou, A.H. (2002). Neural induction, the default model and embryonic stem cells. Nature Reviews Neuroscience *3*, 271.

Mäkelä, J., Tselykh, T.V., Maiorana, F., Eriksson, O., Do, H.T., Mudò, G., Korhonen, L.T., Belluardo, N., and Lindholm, D. (2014). Fibroblast growth factor-21 enhances mitochondrial functions and increases the activity of PGC-1α in human dopaminergic neurons via Sirtuin-1. In Springerplus.

Nakae, K., Nakajima, K., Inazawa, J., Kitaoka, T., and Hirano, T. (1995). ERM, a PEA3 subfamily of Ets transcription factors, can cooperate with c-Jun. The Journal of biological chemistry *270*, 23795-23800.

Nakayama, T., Blitz, I.L., Fish, M.B., Odeleye, A.O., Manohar, S., Cho, K.W., and Grainger, R.M. (2014). Cas9-based genome editing in Xenopus tropicalis. Methods in enzymology *546*, 355-375.

Nasevicius, A., and Ekker, S.C. (2000). Effective targeted gene 'knockdown' in zebrafish. Nature genetics *26*, 216-220.

Neto-Silva, R.M., Gulbenkian PhD Program in Biomedicine, P.-.-O., Portugal, Department of Genetics and Development, C.U.M.C., 701 West 168th Street, New York, NY 10032, USA, Beco, S.d., Department of Genetics and Development, C.U.M.C., 701 West 168th Street, New York, NY 10032, USA, Johnston, L.A., lj180@columbia.edu, and Department of Genetics and Development, C.U.M.C., 701 West 168th Street, New York, NY 10032, USA (2010). Evidence for a Growth-Stabilizing Regulatory Feedback Mechanism between Myc and Yorkie, the Drosophila Homolog of Yap. Developmental cell *19*, 507-520.

Newport, J., and Kirschner, M. (1982). A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell *30*, 675-686.

Nicholson, K.M., and Anderson, N.G. (2002). The protein kinase B/Akt signalling pathway in human malignancy. Cellular signalling *14*, 381-395.

Niehrs, C., and Pollet, N. (1999). Synexpression groups in eukaryotes. Nature *402*, 483-487.

Nieto, M., Monuki, E.S., Tang, H., Imitola, J., Haubst, N., Khoury, S.J., Cunningham, J., Gotz, M., and Walsh, C.A. (2004). Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex. The Journal of comparative neurology *479*, 168-180.

Nieuwkoop, P.D., and Faber, J. (1967). Normal table of Xenopus laevis (Daudin). A systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis, 2. edn (Amsterdam,: North-Holland Pub. Co.).

Nolo, R., Department of Biochemistry and Molecular Biology, U.o.T.M.D.A.C.C., Houston, Texas 77030, Morrison, C.M., Department of Biochemistry and Molecular Biology, U.o.T.M.D.A.C.C., Houston, Texas 77030, Program in Developmental Biology, B.C.o.M., Houston, Texas 77030, Tao, C., Department of Biochemistry and Molecular Biology, U.o.T.M.D.A.C.C., Houston, Texas 77030, Zhang, X., Department of Biochemistry and Molecular Biology, U.o.T.M.D.A.C.C., Houston, Texas 77030, Halder, G.*, et al.* (2006). The bantam MicroRNA Is a Target of the Hippo Tumor-Suppressor Pathway. Current Biology *16*, 1895-1904.

Nummenmaa, E., Hamalainen, M., Moilanen, T., Vuolteenaho, K., and Moilanen, E. (2015). Effects of FGF-2 and FGF receptor antagonists on MMP enzymes, aggrecan, and type II collagen in primary human OA chondrocytes. Scand J Rheumatol *44*, 321-330.

O'Shea, E.K., Rutkowski, R., Stafford, W.F., 3rd, and Kim, P.S. (1989). Preferential heterodimer formation by isolated leucine zippers from fos and jun. Science *245*, 646-648.

Oliver, L.J., Rifkin, D.B., Gabrilove, J., Hannocks, M.J., and Wilson, E.L. (1990). Long-term culture of human bone marrow stromal cells in the presence of basic fibroblast growth factor. Growth factors *3*, 231-236.

Ong, S.H., Guy, G.R., Hadari, Y.R., Laks, S., Gotoh, N., Schlessinger, J., and Lax, I. (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*, 979-989.

Ornitz, D.M., and Itoh, N. (2001). Fibroblast growth factors. Genome biology *2*, REVIEWS3005.

Owens, N.D.L., Blitz, I.L., Lane, M.A., Patrushev, I., Overton, J.D., Gilchrist, M.J., Cho, K.W.Y., and Khokha, M.K. (2016). Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Rep *14*, 632-647.

Padul, V., Epari, S., Moiyadi, A., Shetty, P., and Shirsat, N.V. (2015). ETV/Pea3 family transcription factor-encoding genes are overexpressed in CIC-mutant oligodendrogliomas. Genes, chromosomes & cancer *54*, 725-733.

Panitz, F., Krain, B., Hollemann, T., Nordheim, A., and Pieler, T. (1998). The Spemann organizer-expressed zinc finger gene Xegr-1 responds to the MAP kinase/Ets-SRF signal transduction pathway. The EMBO journal *17*, 4414-4425.

Papagianni, A., Forés, M., Shao, W., He, S., Koenecke, N., Andreu, M.J., Samper, N., Paroush, Z., González-Crespo, S., Zeitlinger, J*., et al.* (2018). Capicua controls Toll/IL-1 signaling targets independently of RTK regulation. In Proceedings of the National Academy of Sciences of the United States of America, pp. 1807-1812.

Parkinson, J., and Blaxter, M. (2009). Expressed sequence tags: an overview. Methods in molecular biology *533*, 1-12.

Paroush, Z., Wainwright, S.M., and Ish-Horowicz, D. (1997). Torso signalling regulates terminal patterning in Drosophila by antagonising Groucho-mediated repression. Development *124*, 3827-3834.

Peifer, M., Berg, S., and Reynolds, A.B. (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. Cell *76*, 789-791.

Peirano, R.I., and Wegner, M. (2000). The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. Nucleic acids research *28*, 3047-3055.

Pera, E.M., Ikeda, A., Eivers, E., and De Robertis, E.M. (2003). Integration of IGF, FGF, and anti-BMP signals via Smad1 phosphorylation in neural induction. Genes & development *17*, 3023-3028.

Pezeron, G., Lambert, G., Dickmeis, T., Strahle, U., Rosa, F.M., and Mourrain, P. (2008). Rasl11b knock down in zebrafish suppresses one-eyed-pinhead mutant phenotype. PLoS One *3*, e1434.

Piccolo, S., Department of Biological Chemistry, H.H.M.I., University of California, Los Angeles, California 90095-1737, USA, Sasai, Y., Department of Biological Chemistry, H.H.M.I., University of California, Los Angeles, California 90095-1737, USA, Lu, B., Department of Biological Chemistry, H.H.M.I., University of California, Los Angeles, California 90095-1737, USA, Robertis, E.M.D., and Department of Biological Chemistry, H.H.M.I., University of California, Los Angeles, California 90095-1737, USA (1996). Dorsoventral Patterning in Xenopus: Inhibition of Ventral Signals by Direct Binding of Chordin to BMP-4. Cell *86*, 589-598.

Pownall, M.E., Tucker, A.S., Slack, J.M., and Isaacs, H.V. (1996a). eFGF, Xcad3 and Hox genes form a molecular pathway that establishes the anteroposterior axis in Xenopus. Development *122*, 3881-3892.

Pownall, M.E., Tucker, A.S., Slack, J.M.W., and Isaacs, H.V. (1996b). eFGF, Xcad3 and Hox genes form a molecular pathway that establishes the anteroposterior axis in Xenopus. Development *122*, 3881-3892.

Price, C.A. (2016). Mechanisms of fibroblast growth factor signaling in the ovarian follicle. J Endocrinol *228*, R31-43.

Prioleau, M.N., Huet, J., Sentenac, A., and Mechali, M. (1994). Competition between chromatin and transcription complex assembly regulates gene expression during early development. Cell *77*, 439-449.

Qian, Y., and Chen, X. (2013). Senescence regulation by the p53 protein family. Methods in molecular biology *965*, 37-61.

Rana, A.A., Collart, C., Gilchrist, M.J., and Smith, J.C. (2006). Defining Synphenotype Groups in Xenopus tropicalis by Use of Antisense Morpholino Oligonucleotides. PLoS Genet *2*.

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*, 134-140.

Reeves, G.T., and Stathopoulos, A. (2009). Graded dorsal and differential gene regulation in the Drosophila embryo. Cold Spring Harbor perspectives in biology *1*, a000836.

Remy, P., and Baltzinger, M. (2000). The Ets-transcription factor family in embryonic development: lessons from the amphibian and bird. Oncogene *19*, 6417-6431.

Rhinn, M., Lun, K., Luz, M., Werner, M., and Brand, M. (2005). Positioning of the midbrain-hindbrain boundary organizer through global posteriorization of the neuroectoderm mediated by Wnt8 signaling. Development *132*, 1261-1272.

Ridley, A.J. (2015). Rho GTPase signalling in cell migration. Curr Opin Cell Biol *36*, 103-112.

Rittenhouse, K.R., and Berg, C.A. (1995). Mutations in the Drosophila gene bullwinkle cause the formation of abnormal eggshell structures and bicaudal embryos. Development *121*, 3023-3033.

Roberts, P., Browne, C.F., Lewis, I.J., Bailey, C.C., Spicer, R.D., Williams, J., and Batcup, G. (1992). 12q13 abnormality in rhabdomyosarcoma. A nonrandom occurrence? Cancer genetics and cytogenetics *60*, 135-140.

Roch, F., Jimenez, G., and Casanova, J. (2002). EGFR signalling inhibits Capicua-dependent repression during specification of Drosophila wing veins. Development *129*, 993-1002.

Rogers, C., Moody, S.A., and Casey, E. (2009). Neural induction and factors that stabilize a neural fate. Birth Defects Res C Embryo Today *87*, 249-262.

Rohmann, E., Brunner, H.G., Kayserili, H., Uyguner, O., Nurnberg, G., Lew, E.D., Dobbie, A., Eswarakumar, V.P., Uzumcu, A., Ulubil-Emeroglu, M.*, et al.* (2006). Mutations in different components of FGF signaling in LADD syndrome. Nature genetics *38*, 414-417.

Rousseaux, M.W.C., Tschumperlin, T., Lu, H.C., Lackey, E.P., Bondar, V.V., Wan, Y.W., Tan, Q., Adamski, C.J., Friedrich, J., Twaroski, K.*, et al.* (2018). ATXN1-CIC Complex Is the Primary Driver of Cerebellar Pathology in Spinocerebellar Ataxia Type 1 through a Gain-of-Function Mechanism. Neuron *97*, 1235-1243.e1235.

Rusanescu, G., and Mao, J. (2014). Notch3 is necessary for neuronal differentiation and maturation in the adult spinal cord. J Cell Mol Med *18*, 2103-2116.

Sakaguchi, K., Lorenzi, M.V., Bottaro, D.P., and Miki, T. (1999). The acidic domain and first immunoglobulin-like loop of fibroblast growth factor receptor 2 modulate downstream signaling through glycosaminoglycan modification. Molecular and cellular biology *19*, 6754-6764.

Saloman, D.S., Bianco, C., Ebert, A.D., Khan, N.I., De Santis, M., Normanno, N., Wechselberger, C., Seno, M., Williams, K., Sanicola, M.*, et al.* (2000). The EGF-CFC family: novel epidermal growth factor-related proteins in development and cancer. Endocr Relat Cancer *7*, 199-226.

Sasai, Y., Lu, B., Steinbeisser, H., and Robertis, E.D. (1995). Regulation of neural induction by the Chd and Bmp-4 antagonisticpatterning signals in Xenopus. Nature *376*, 333.

Schier, A.F. (2007). The Maternal-Zygotic Transition: Death and Birth of RNAs.

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

Sen, R., and Baltimore, D. (1986). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell *46*, 705-716.

Seo, H.R., Jeong, H.E., Joo, H.J., Choi, S.C., Park, C.Y., Kim, J.H., Choi, J.H., Cui, L.H., Hong, S.J., Chung, S.*, et al.* (2016). Intrinsic FGF2 and FGF5 promotes angiogenesis of human aortic endothelial cells in 3D microfluidic angiogenesis system. Sci Rep *6*.

Shermoen, A.W., and O'Farrell, P.H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell *67*, 303-310.

Shilo, B.Z. (2003). Signaling by the Drosophila epidermal growth factor receptor pathway during development. Experimental cell research *284*, 140-149.

Shinya, M., Koshida, S., Sawada, A., Kuroiwa, A., and Takeda, H. (2001). Fgf signalling through MAPK cascade is required for development of the subpallial telencephalon in zebrafish embryos.

Simon-Carrasco, L., Grana, O., Salmon, M., Jacob, H.K.C., Gutierrez, A., Jimenez, G., Drosten, M., and Barbacid, M. (2017). Inactivation of Capicua in adult mice causes T-cell lymphoblastic lymphoma. Genes & development *31*, 1456-1468.

Sinner, D., Rankin, S., Lee, M., and Zorn, A.M. (2004). Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. Development *131*, 3069-3080.

Sivak, J.M., The Wellcome Trust/Cancer Research UK Gurdon Institute, U.o.C., Tennis Court Road Cambridge, CB2 1QN, United Kingdom, Petersen, L.F., The Wellcome Trust/Cancer Research UK Gurdon Institute, U.o.C., Tennis Court Road Cambridge, CB2 1QN, United Kingdom, Department of Zoology, U.o.C., Downing Site, Cambridge, United Kingdom, Amaya, E., ea3@mole.bio.cam.ac.uk, The Wellcome Trust/Cancer Research UK Gurdon Institute, U.o.C., Tennis Court Road Cambridge, CB2 1QN, United Kingdom, and Department of Zoology, U.o.C., Downing Site, Cambridge, United Kingdom (2005). FGF Signal Interpretation Is Directed by Sprouty and Spred Proteins during Mesoderm Formation. Developmental cell *8*, 689-701.

Sive, H.L., Grainger, R.M., and Harland, R.M. (2000). Early Development of Xenopus laevis: A Laboratory Manual (Cold Spring Harbor Laboratory Press).

Slack, J.M., Darlington, B.G., Heath, J.K., and Godsave, S.F. (1987). Mesoderm induction in early Xenopus embryos by heparin-binding growth factors. Nature *326*, 197-200.

Slack, J.M., Isaacs, H.V., Johnson, G.E., Lettice, L.A., Tannahill, D., and Thompson, J. (1992). Specification of the body plan during Xenopus gastrulation: dorsoventral and anteroposterior patterning of the mesoderm. Dev Suppl, 143-149.

Slack, J.M.W., and Isaacs, H.V. (1989). Presence of basic fibroblast growth factor in the early embryo.

Smith, J.C., Price, B.M.J., Green, J.B.A., Weigel, D., and Herrmann, B.G. (1991). Expression of a Xenopus Homolog of Brachyury (T) Is an Immediate-Early Response to Mesoderm Induction. Cell *67*, 79-87.

Smith, W.C., and Harland, R.M. (1992). Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in Xenopus embryos. Cell *70*, 829-840.

Song, H., Mak, K.K., Topol, L., Yun, K., Hu, J., Garrett, L., Chen, Y., Park, O., Chang, J., Simpson, R.M.*, et al.* (2010). Mammalian Mst1 and Mst2 kinases play essential roles in organ size control and tumor suppression. Proc Natl Acad Sci U S A *107*, 1431-1436.

Spemann, H., and Mangold, H. (2001). Induction of embryonic primordia by implantation of organizers from a different species. 1923. Int J Dev Biol *45*, 13-38.

Spence, J.R., Madhavan, M., Aycinena, J.C., and Del Rio-Tsonis, K. (2007). Retina regeneration in the chick embryo is not induced by spontaneous Mitf downregulation but requires FGF/FGFR/MEK/Erk dependent upregulation of Pax6. Molecular vision *13*, 57-65.

Srisakuldee, W., Institute of Cardiovascular Sciences, S.B.G.H.R.C., 351 Taché Avenue, Winnipeg, Manitoba, CanadaR2H 2A6, Department of Physiology, U.o.M., Winnipeg, Manitoba, Canada, Makazan, Z., Institute of Cardiovascular Sciences, S.B.G.H.R.C., 351 Taché Avenue, Winnipeg, Manitoba, CanadaR2H 2A6, Nickel, B.E., Institute of Cardiovascular Sciences, S.B.G.H.R.C., 351 Taché Avenue, Winnipeg, Manitoba, CanadaR2H 2A6, Zhang, F., Department of Pharmacology, S.C.T.M.B., Dalhousie University, Halifax, Nova Scotia, Canada, Thliveris, J.A.*, et al.* (2018). The FGF-2-triggered protection of cardiac subsarcolemmal mitochondria from calcium overload is mitochondrial connexin 43-dependent. Cardiovascular Research *103*, 72-80.

Stein, D.S., and Stevens, L.M. (2014). Maternal control of the Drosophila dorsal-ventral body axis. Wiley interdisciplinary reviews Developmental biology *3*, 301-330.

Stolle, K., Schnoor, M., Fuellen, G., Spitzer, M., Cullen, P., and Lorkowski, S. (2007). Cloning, genomic organization, and tissue-specific expression of the RASL11B gene. Biochim Biophys Acta *1769*, 514-524.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc Natl Acad Sci U S A *100*, 9440-9445.

Streit, A., Berliner, A.J., Papanayotou, C., Sirulnik, A., and Stern, C.D. (2000). Initiation of neural induction by FGF signalling before gastrulation. Nature *406*, 74-78.

Su, N., Jin, M., and Chen, L. (2014). Role of FGF/FGFR signaling in skeletal development and homeostasis: learning from mouse models. Bone Res *2*, 14003-.

Suisse, A., He, D., Legent, K., and Treisman, J.E. (2017a). COP9 signalosome subunits protect Capicua from MAPK-dependent and -independent mechanisms of degradation. Development *144*, 2673-2682.

Suisse, A., He, D., Legent, K., and Treisman, J.E. (2017b). COP9 signalosome subunits protect Capicua from MAPK-dependent and -independent mechanisms of degradation. In Development, pp. 2673-2682.

Summerton, J. (1999). Morpholino antisense oligomers: the case for an RNase H-independent structural type. Biochimica et biophysica acta *1489*, 141-158.

Summerton, J., and Weller, D. (1997). Morpholino antisense oligomers: design, preparation, and properties. Antisense Nucleic Acid Drug Dev *7*, 187-195.

Sun, X., Mariani, F.V., and Martin, G.R. (2002). Functions of FGF signalling from the apical ectodermal ridge in limb development. Nature *418*, 501-508.

Tada, M., and Smith, J.C. (2000). Xwnt11 is a target of Xenopus Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. Development *127*, 2227-2238.

Takekawa, M., and Saito, H. (1998). A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. Cell *95*, 521-530.

Tan, Y., Low, K.G., Boccia, C., Grossman, J., and Comb, M.J. (1994). Fibroblast growth factor and cyclic AMP (cAMP) synergistically activate gene expression at a cAMP response element. Mol Cell Biol *14*, 7546-7556.

Tanaka, M., Yoshimoto, T., and Nakamura, T. (2017). A double-edged sword: The world according to Capicua in cancer. Cancer Sci *108*, 2319-2325.

Tanigaki, K., Nogaki, F., Takahashi, J., Tashiro, K., Kurooka, H., and Honjo, T. (2001). Notch1 and Notch3 instructively restrict bFGF-responsive multipotent neural progenitor cells to an astroglial fate. Neuron *29*, 45-55.

Taniguchi, Y. (2014). Hox Transcription Factors: Modulators of Cell-Cell and Cell-Extracellular Matrix Adhesion. Biomed Res Int *2014*.

Tapon, N., Harvey, K.F., Bell, D.W., Wahrer, D.C.R., Schiripo, T.A., Haber, D.A., Hariharan, I.K., and hariharan@helix.mgh.harvard.edu (2002). salvador Promotes Both Cell Cycle Exit and Apoptosis in Drosophila and Is Mutated in Human Cancer Cell Lines. Cell *110*, 467-478.

Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003). Genome Duplication, a Trait Shared by 22,000 Species of Ray-Finned Fish. In Genome Res, pp. 382-390.

Tekin, M., Hismi, B.O., Fitoz, S., Ozdag, H., Cengiz, F.B., Sirmaci, A., Aslan, I., Inceoglu, B., Yuksel-Konuk, E.B., Yilmaz, S.T.*, et al.* (2007). Homozygous mutations in fibroblast growth factor 3 are associated with a new form of syndromic deafness characterized by inner ear agenesis, microtia, and microdontia. American journal of human genetics *80*, 338-344.

Teven, C.M., Farina, E.M., Rivas, J., and Reid, R.R. (2014). Fibroblast growth factor (FGF) signaling in development and skeletal diseases. Genes Dis *1*, 199-213.

Thisse, B., and Thisse, C. (2005). Functions and regulations of fibroblast growth factor signaling during embryonic development. Dev Biol *287*, 390-402.

Thompson, B.J., European Molecular Biology Laboratory, M., Heidelberg 69117, Germany, Cohen, S.M., cohen@embl.de, and European Molecular Biology

Laboratory, M., Heidelberg 69117, Germany (2006). The Hippo Pathway Regulates the bantam microRNA to Control Cell Proliferation and Apoptosis in Drosophila. Cell *126*, 767-774.

Tong, X., Gui, H., Jin, F., Heck, B.W., Lin, P., Ma, J., Fondell, J.D., and Tsai, C.C. (2011). Ataxin-1 and Brother of ataxin-1 are components of the Notch signalling pathway.

Tsai, C.-C., Kao, H.-Y., Mitzutani, A., Banayo, E., Rajan, H., McKeown, M., and Evans, R.M. (2004). Ataxin 1, a SCA1 neurodegenerative disorder protein, is functionally linked to the silencing mediator of retinoid and thyroid hormone receptors.

Tseng, A.S., Tapon, N., Kanda, H., Cigizoglu, S., Edelmann, L., Pellock, B., White, K., and Hariharan, I.K. (2007). Capicua regulates cell proliferation downstream of the receptor tyrosine kinase/ras signaling pathway. Current biology : CB *17*, 728-733.

Turner, N., and Grose, R. (2010). Fibroblast growth factor signalling: from development to cancer. Nature reviews Cancer *10*, 116-129.

Ueda, Y., Hirai, S., Osada, S., Suzuki, A., Mizuno, K., and Ohno, S. (1996). Protein kinase C activates the MEK-ERK pathway in a manner independent of Ras and dependent on Raf. The Journal of biological chemistry *271*, 23512-23519.

Uhlirova, M., and Bohmann, D. (2006). JNK- and Fos-regulated Mmp1 expression cooperates with Ras to induce invasive tumors in Drosophila. EMBO J *25*, 5294-5304.

Umbhauer, M., Marshall, C.J., Mason, C.S., Old, R.W., and Smith, J.C. (1995). Mesoderm Induction in Xenopus Caused by Activation of Map Kinase. Nature *376*, 58-62.

Urness, L.D., Paxton, C.N., Wang, X., Schoenwolf, G.C., and Mansour, S.L. (2010). FGF signaling regulates otic placode induction and refinement by controlling both ectodermal target genes and hindbrain Wnt8a. Dev Biol *340*, 595-604.

Vallier, L., Touboul, T., Chng, Z., Brimpari, M., Hannan, N., Millan, E., Smithers, L.E., Trotter, M., Rugg-Gunn, P., Weber, A.*, et al.* (2009). Early Cell Fate Decisions of Human Embryonic Stem Cells and Mouse Epiblast Stem Cells Are Controlled by the Same Signalling Pathways. PLoS One *4*.

Venkatarama, T., Lai, F., Luo, X., Zhou, Y., Newman, K., and King, M.L. (2010). Repression of zygotic gene expression in the Xenopus germline. Development *137*, 651-660.

Verheyden, J.M., Lewandoski, M., Deng, C., Harfe, B.D., and Sun, X. (2005). Conditional inactivation of Fgfr1 in mouse defines its role in limb bud establishment, outgrowth and digit patterning. Development *132*, 4235-4245.

Vetrini, F., Tammaro, R., Bondanza, S., Surace, E.M., Auricchio, A., De Luca, M., Ballabio, A., and Marigo, V. (2006). Aberrant splicing in the ocular albinism type 1 gene (OA1/GPR143) is corrected in vitro by morpholino antisense oligonucleotides. Hum Mutat *27*, 420-426.

Villafuerte, S.M., Vallabhaneni, K., Śliwerska, E., McMahon, F.J., Young, E.A., and Burmeister, M. (2009). SSRI response in depression may be influenced by SNPs in HTR1B and HTR1A. Psychiatr Genet *19*, 281-291.

Vitorino, F.N.L., Montoni, F., Moreno, J.N., de Souza, B.F., Lopes, M.C., Cordeiro, B., Fonseca, C.S., Gilmore, J.M., Sardiu, M.I., Reis, M.S.*, et al.* (2018). FGF2 Antiproliferative Stimulation Induces Proteomic Dynamic Changes and High Expression of FOSB and JUNB in K-Ras-Driven Mouse Tumor Cells. Proteomics *18*, e1800203.

Wallingford, J.B., Fraser, S.E., and Harland, R.M. (2002). Convergent Extension: The Molecular Control of Polarized Cell Movement during Embryonic Development. Developmental cell *2*, 695-706.

Wang, B., Krall, E.B., Aguirre, A.J., Kim, M., Widlund, H.R., Doshi, M.B., Sicinska, E., Sulahian, R., Goodale, A., Cowley, G.S.*, et al.* (2017). ATXN1L, CIC, and ETS Transcription Factors Modulate Sensitivity to MAPK Pathway Inhibition. Cell Rep *18*, 1543-1557.

Wang, S., Krinks, M., Lin, K., Luyten, F.P., and Moos, M., Jr. (1997). Frzb, a secreted protein expressed in the Spemann organizer, binds and inhibits Wnt-8. Cell *88*, 757-766.

Wasylyk, C., Salvi, R., Argentini, M., Dureuil, C., Delumeau, I., Abecassis, J., Debussche, L., and Wasylyk, B. (1999). p53 mediated death of cells overexpressing MDM2 by an inhibitor of MDM2 interaction with p53. Oncogene *18*, 1921-1934.

Watabe, T., Kim, S., Candia, A., Rothbacher, U., Hashimoto, C., Inoue, K., and Cho, K.W. (1995). Molecular mechanisms of Spemann's organizer formation: conserved growth factor synergy between Xenopus and mouse. Genes Dev *9*, 3038-3050.

Weintraub, H. (1993). The MyoD family and myogenesis: redundancy, networks, and thresholds. Cell *75*, 1241-1244.

Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T.K., Turner, D., Rupp, R., Hollenberg, S.*, et al.* (1991). The myoD gene family: nodal point during specification of the muscle cell lineage. Science *251*, 761-766.

Weiss, J.B., Von Ohlen, T., Mellerick, D.M., Dressler, G., Doe, C.Q., and Scott, M.P. (1998). Dorsoventral patterning in the Drosophila central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity. Genes & development *12*, 3591-3602.

Weissmann, S., Cloos, P.A., Sidoli, S., Jensen, O.N., Pollard, S., and Helin, K. (2018). The tumor suppressor CIC directly regulates MAPK pathway genes via histone deacetylation. Cancer Res.

Wessely, O., Kim, J.I., Geissert, D., Tran, U., and De Robertis, E.M. (2004). Analysis of Spemann organizer formation in Xenopus embryos by cDNA macroarrays. Developmental biology *269*, 552-566.

Westall, F.C., Lennon, V.A., and Gospodarowicz, D. (1978). Brain-derived fibroblast growth factor: identity with a fragment of the basic protein of myelin. Proceedings of the National Academy of Sciences of the United States of America *75*, 4675-4678.

Whitmarsh, A.J., and Davis, R.J. (1996). Transcription factor AP-1 regulation by mitogen-activated protein kinase signal transduction pathways. J Mol Med (Berl) *74*, 589-607.

Wierman, M.E., Kiseljak-Vassiliades, K., and Tobet, S. (2011). Gonadotropin Releasing Hormone (GnRH) Neuron Migration: Initiation, Maintenance and Cessation as Critical Steps to Ensure Normal Reproductive Function. Front Neuroendocrinol *32*, 43-52.

Wilson, P.A., and Hemmati-Brivanlou, A. (1995). Induction of epidermis and inhibition of neural fate by Bmp-4. Nature *376*, 331-333.

Wilson, S.I., Graziano, E., Harland, R., Jessell, T.M., and Edlund, T. (2000). An early requirement for FGF signalling in the acquisition of neural cell fate in the chick embryo. Curr Biol *10*, 421-429.

Wright, K.D., Mahoney Rogers, A.A., Zhang, J., and Shim, K. (2015). Cooperative and independent functions of FGF and Wnt signaling during early inner ear development. BMC Dev Biol *15*.

Wu, M.Y., and Hill, C.S. (2009). Tgf-beta superfamily signaling in embryonic development and homeostasis. Dev Cell *16*, 329-343.

Wu, S., Department of Physiology, U.o.T.S.M.C.a.D., 5323 Harry Hines Boulevard, Dallas, TX 75390 USA, Huang, J., Department of Physiology, U.o.T.S.M.C.a.D., 5323 Harry Hines Boulevard, Dallas, TX 75390 USA, Dong, J., Department of Physiology, U.o.T.S.M.C.a.D., 5323 Harry Hines Boulevard, Dallas, TX 75390 USA, Pan, D., dpan@mednet.swmed.edu, and Department of Physiology, U.o.T.S.M.C.a.D., 5323 Harry Hines Boulevard, Dallas, TX 75390 USA (2003). hippo Encodes a Ste-20 Family Protein Kinase that Restricts Cell Proliferation and Promotes Apoptosis in Conjunction with salvador and warts. Cell *114*, 445-456.

Yamamoto, H., Horiuchi, S., Adachi, Y., Taniguchi, H., Nosho, K., Min, Y., and Imai, K. (2004). Expression of ets-related transcriptional factor E1AF is associated with

tumor progression and over-expression of matrilysin in human gastric cancer. Carcinogenesis *25*, 325-332.

Yang, L., Paul, S., Trieu, K.G., Dent, L.G., Froldi, F., Fores, M., Webster, K., Siegfried, K.R., Kondo, S., Harvey, K*., et al.* (2016). Minibrain and Wings apart control organ growth and tissue patterning through down-regulation of Capicua. Proc Natl Acad Sci U S A *113*, 10583-10588.

Yang, R., Chen, L.H., Hansen, L.J., Carpenter, A.B., Moure, C.J., Liu, H., Pirozzi, C.J., Diplas, B.H., Waitkus, M.S., Greer, P.K*., et al.* (2017). Cic Loss Promotes Gliomagenesis via Aberrant Neural Stem Cell Proliferation and Differentiation. Cancer research *77*, 6097-6108.

Yip, S., Butterfield, Y.S., Morozova, O., Chittaranjan, S., Blough, M.D., An, J., Birol, I., Chesnelong, C., Chiu, R., Chuah, E*., et al.* (2012). Concurrent CIC mutations, IDH mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers. The Journal of pathology *226*, 7-16.

Yoshii, C., Ueda, Y., Okamoto, M., and Araki, M. (2007). Neural retinal regeneration in the anuran amphibian Xenopus laevis post-metamorphosis: transdifferentiation of retinal pigmented epithelium regenerates the neural retina. Developmental biology *303*, 45-56.

Zeisberg, M., Hanai, J., Sugimoto, H., Mammoto, T., Charytan, D., Strutz, F., and Kalluri, R. (2003). BMP-7 counteracts TGF-beta1-induced epithelial-to-mesenchymal transition and reverses chronic renal injury. Nat Med *9*, 964-968.

Zhang, J., Ji, J.-Y., Yu, M., Overholtzer, M., Smolen, G.A., Wang, R., Brugge, J.S., Dyson, N.J., and Haber, D.A. (2009a). YAP-dependent induction of amphiregulin identifies a non-cell-autonomous component of the Hippo pathway. Nature Cell Biology *11*, 1444.

Zhang, X., Ibrahimi, O.A., Olsen, S.K., Umemori, H., Mohammadi, M., and Ornitz, D.M. (2006). Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. The Journal of biological chemistry *281*, 15694-15700.

Zhang, Z., Verheyden, J.M., Hassell, J.A., and Sun, X. (2009b). FGF-regulated Etv genes are essential for repressing Shh expression in mouse limb buds. Dev Cell *16*, 607-613.

Zhong, L., Huang, X., Rodrigues, E.D., Leijten, J.C., Verrips, T., El Khattabi, M., Karperien, M., and Post, J.N. (2016). Endogenous DKK1 and FRZB Regulate Chondrogenesis and Hypertrophy in Three-Dimensional Cultures of Human Chondrocytes and Human Mesenchymal Stem Cells. Stem Cells Dev *25*, 1808-1817.

Zimmerman, L.B., De Jesus-Escobar, J.M., and Harland, R.M. (1996a). The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. Cell *86*, 599-606.

Zimmerman, L.B., Department of Molecular and Cell Biology, D.o.B.a.M.B., University of California, Berkeley, California 94720, USA, Jesús-Escobar, J.M.D., Department of Molecular and Cell Biology, D.o.B.a.M.B., University of California, Berkeley, California 94720, USA, Harland, R.M., harland@mendel.berkeley.edu, and Department of Molecular and Cell Biology, D.o.B.a.M.B., University of California, Berkeley, California 94720, USA (1996b). The Spemann Organizer Signal noggin Binds and Inactivates Bone Morphogenetic Protein 4. Cell *86*, 599-606.

Znosko, W.A., Yu, S., Thomas, K., Molina, G.A., Li, C., Tsang, W., Dawid, I.B., Moon, A.M., and Tsang, M. (2010). Overlapping functions of Pea3 ETS transcription factors in FGF signaling during zebrafish development. Developmental biology *342*, 11-25.

Zoghbi, H.Y., and Orr, H.T. (2009). Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. The Journal of biological chemistry *284*, 7425-7429.