Molecular analysis of the ZIC3 and TCF7L1 interaction and its implication in human Heterotaxy

Koula Esta Maria Diamand

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University

August 2019

Except where otherwise acknowledged, the work presented in this thesis is my own.

Koula Diamand

Publications contributed to through the course of this thesis:

Diamand K.E.M., Barratt K.S., Arkell R.M. (2018) Overview of Rodent *Zic* Genes. In: Aruga J. (eds) Zic family. Advances in Experimental Medicine and Biology, vol 1046. Springer, Singapore

Barratt, K.S., **Diamand, K. E. M**. & Arkell, R.M (2018) Identification of reference genes suitable for RT-qPCR studies of murine gastrulation and patterning. Mammalian Genome. 29(9-10):656-662

Publication in preparation:

Jehangir N. Ahmed, **Koula E. M. Diamand** and Ruth M. Arkell. Systematized reporter assays reveal ZIC protein regulatory abilities are subclass-specific and dependent upon transcription factor binding site context

Acknowledgments

If I could, my acknowledgements would be short and simple. Something along the lines of:

‘Thank you – you know who you are.’

Vague, to the point, and no one should feel left out. This PhD, however, has been anything but ‘short and simple’ and so I must express my gratitude to those who have helped me get through this milestone.

First and foremost, I would like to offer my special thanks to my supervisor, Ruth Arkell, for the opportunity to conduct my PhD in your research group. I have learnt a great deal from you and you have helped me develop essential research, presentation and communication skills. Thank you for your supervision, encouragement and thesis writing advice.

To all members of the Arkell laboratory, both past and present (Kristen, Alaa, Jerry, Hannah, Helen, Kathryn, Kelsey, Kyle and the adopted member, Lora), who all have at some point provided me with either support, reassurance, inspiration or motivation, particularly in these last few months. Thank you for all the coffee breaks, the ridiculous conversations, heated discussions trying to answer the *real* philosophical questions of the world (such as, at what point is a fruit salad not a fruit salad - a question with many differing opinions) and life advice -maybe one day I will open a bakery. I also want to give special thanks to Kristen, Kyle and Alaa for proof reading my thesis and Hannah for your infectious optimism.

To Kristen, I give you a poem for you as you did for me, to summarise us as acquaintances, as co-workers and, yes, even friends. May this become a respected and honourable Arkell laboratory tradition that I will in no way encourage or enforce.

I need you to know just in case,

Please add this to your fun fact knowledge base,

Uno a sprinkle is light,

Not drizzle, don’t fight

Now how about a game of fruit-face?

I also wish to acknowledge the help, guidance and lab reagents provided by Nicole Thompson and the Fisher laboratory at JCSMR. Without your assistance I would not have been able to set up or conduct my yeast two-hybrid experiments. Thanks also must be given to the Australian Government for their support through the Australian Postgraduate Award.

Last but not the least, I would like to thank my family. To my parents, words cannot describe how thankful I am for your unconditional love and support. I could not have come this far in life without you both and I am grateful for everything you have done in making me who I am today. To my brothers, who never quite understood why I attempted a PhD, what I even do at university, and how I even managed to finish this thesis. Although you may not be aware, you remind me of all the positive opportunities life will present. Happy reading.

Abstract

Heterotaxy is a congenital disorder that occurs due to the defective establishment of the left-right (L-R) body axis during early embryo development. The disorder affects 1/10,000 live births, results in the abnormal alignment of internal organs with respect to the L-R axis, and is often accompanied by severe cardiac malformations. The most commonly mutated gene in human cases of Heterotaxy is the X-linked *ZIC3*, which was first associated with the condition in 1997. ZIC3 belongs to a family of multifunctional transcriptional regulator proteins (ZIC family) all characterised by a C_2H_2 type zinc finger domain (ZFD). This domain is implicated in the transactivation (DNA binding) and co-factor (protein binding) function of the proteins. However, despite both capabilities, it is still not known when during development the co-factor role of ZIC3 is necessary over the transcription factor role. Consequently, the exact cellular and molecular causes underlying ZIC3-associated Heterotaxy remain unknown.

Mouse models of *Zic3* loss-of-function reveal phenotypes that resemble human Heterotaxy and exhibit a defective node (the structure that confers the left-right axis during development). More recently, studies have shown murine *Zic3* alleles lead to incompletely penetrant, partial (posterior) axis duplications and anterior truncation, phenotypes characteristic of elevated canonical Wnt signalling. The expression pattern of *Zic3* during mammalian gastrulation overlaps with a known repressor of the canonical Wnt pathway, *Tcf7l1* and it has also been shown that a related ZIC protein (ZIC2) can interact with TCF proteins to co-repress Wnt/β-catenin mediated transcription when overexpressed in cell lines. **Overall, this leads to the novel hypothesis that ZIC3 controls the establishment of the L-R axis by interacting with TCF7L1 to repress Wnt signalling during mammalian gastrulation. In this thesis I examine the molecular role of ZIC3 in the Wnt pathway and investigate the functional consequences of ZIC3 Heterotaxy-associated protein variants.**

A physical interaction between ZIC3 and TCF7L1 was demonstrated by several molecular assays. Additionally, 17 Heterotaxy-associated human *ZIC3* variants were functionally assessed in well-established cell-based assays. Four of the human protein variants exhibit non-pathogenic properties as they retain wildtype ZIC3 function in these assays. The remaining variants, which affect key residues within the ZFD or contain a premature termination codon, display impaired transcription factor and co-factor functions: they can no longer activate transcription at a ZIC responsive element and have reduced ability to inhibit β-catenin-mediated Wnt transcription. Loss of Wnt inhibition did not always correlate with loss of TCF7L1 interaction, suggesting ZIC3 and/or TCF7L1 must interact with additional co-repressors or repressive chromatin complexes to mediate transcriptional repression in the Wnt pathway. Additionally, in this thesis I identified several evolutionary conserved domains implicated in

transactivation, including the newly annotated SANC and SACC domains, enhancing our understanding of the structural domains required for ZIC3 molecular function. Overall, this thesis supports a new model of ZIC3 molecular function: ZIC3 interacts with TCF7L1 to inhibit canonical Wnt signalling during mammalian gastrulation and correctly establish the L-R axis. Ultimately, ZIC3-associated Heterotaxy arises when ZIC3 is unable to function as a Wnt co-repressor.

Table of Contents

List of Figures

List of Tables

Abbreviations

aa – Amino Acid

ABI – Applied Biosystems

ANOVA – Analysis Of Variance

ANU –Australian National University

APC – Adenoma Polyposis Coli

Bfc - Batface

β-gl – Beta globin minimal promoter

BDT –Big-Dye Terminating mixture

bGH polyA– Bovine growth hormone polyadenylation

bp – base pair

BRF – Biomolecular Research Facility

BMPs- Bone morphogenetic proteins

BSA – Bovine Serum Albumin

BiFC – Bimolecular Fluorescence Complementation

CERI/II - Cytoplasmic Extraction Reagent I/II

CDS – Coding sequence

CMV– Cytomegalovirus Promoter

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

DH5α – DH5-alpha electro-competent *E.coli* strain

Dkk1 - Dickkof

DMEM – Dulbecco's Modified Eagle Medium

DMSO – Dimethyl Sulphoxide

DNA – Deoxyribonucleic Acid

dNTP– di-Nucleotide Triphosphate

dpc – Days post coitum

DTT – Dithiothreitol

E.coli – Escherichia coli

EDTA – Ethylenediaminetetraacetic Acid

ENU – *N*-ethyl-*N*-nitrosourea

EtOH – Ethanol

FBS – Foetal Bovine Serum

Foxa2 – Forkhead box protein A2

GFP- Green Fluorescent Protein

GST - Glutathione-S-transferase

HCL – Hydrochloric Acid

HEK293T – Human Embryonic Kidney 293T Cell line

HRP – Horseradish Peroxidase

hZIC3– Human ZIC3 Coding sequence

IFI – Integrated Fluorescence Intensity

IHF– Intergrase Host Factor

Ka - Katun

LB – Luria-Bertani

LEF – Lymphoid Enhancer-binding Factor

L-R – Left-Right

LSD – Fisher's Least Significant Difference

MCS – multiple cloning site

MeOH – Methanol

mRNA – messenger RNA

MW – Molecular Weight

NEB – New England Biolabs

NER– Nuclear Extraction Reagent

NIH – National Institutes of Health

NLS – Nuclear Localisation Signal

NMD- Nonsense mediated decay

NPCs – Nuclear Pore Complex Components

NPG – n-propyl gallate

PBM- Protein binding microarray

PBS – Phosphate Buffered Saline

PBT – Phosphate Buffered Saline with Tween

PCR– Polymerase Chain Reaction

PFA – Paraformaldehyde

PLA – Proximity Ligation Assay

PTC – Premature termination codon

PVDF – Polyvinylidene Fluoride

RNA – ribonucleic acid

rpm – revolutions per minute

SANC – Subclass A N-terminal conserved

SACC - Subclass A C-terminal conserved

SD – standard deviation

SDS– Sodium Dodecyl Sulfate

SDS-PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEM – Standard Error of the Mean

SHH – Sonic Hedgehog Homolog

SV40 – Simian Virus 40 Origin of Replication

TBE– Tris-borate/ (EDTA) - thylenediaminetetraacetic Acid

TBP – Mouse Anti-TATA Binding Protein

TCF – T-Cell Factor

TCF7l1 – T-Cell Factor like 1

TEMED – N, N, N', N'-Tetramethylethylenediamine

TK – Thymidine Kinase

Tm – Melting temperature

TLE – Transducin-like Enhancer of Split gene

U – Enzyme units

UV – Ultraviolet

VCS – Venus cloning sequence

WMISH – Whole Mount In Situ Hybridization

wt – Wildtype

Y2H – Yeast Two-Hybrid

YFP– Yellow Fluorescent Protein

ZOC – ZIC opa conserved

ZFNC – zinc finger N-flanking conserved

ZF – Zinc finger

ZFD– Zinc finger domain

Zic –Zinc finger of the cerebellum

Chapter 1: Introduction

This chapter contains text and figures from the paper Diamand, K.E.M., Barratt, K.S. & Arkell, R.M (2017) Chapter 10: Overview of rodent Zic genes, "Zic family - Evolution, Development and Disease", Springer

1.1 Heterotaxy

The exterior of the vertebrate body plan is bilaterally symmetrical, with a mirror image along the sagittal plane of the body that creates an identical left and right. Interiorly, however, some organs are placed in a specific, yet asymmetrical, position along the left-right (L-R) axis of the body. For example, organs such as the heart, spleen and pancreas are situated on the left of the body, while the liver and gallbladder are positioned along the right. Paired organs such as the lungs also demonstrate asymmetries, for example, the right lung has three lobes, whilst the left has two. This normal organisation of the internal organs is called *situs solitus* (Figure 1.1) and arises because of an initial symmetry-breaking event during early embryogenesis, which subsequently allows the development of a distinct left and right. Failure to establish the correct L-R axis results in laterality disorders and severe congenital defects. Complete reversal of the L-R axis is termed *situs inversus*, characterised by a mirror image of the internal asymmetries. Although, as the organs are still in a functional arrangement with respect to each other, *situs inversus* is not normally associated with developmental disorders. Heterotaxy, otherwise known as *situs ambiguous*, falls somewhere between *situs solitus* and *situs inversus* as it defines the abnormal positioning of the thoraco-abdominal organs along the L-R body axis (Figure 1.1). Heterotaxy can be further divided into subgroups based on the type of symmetry (right isomerism vs left isomerism) or on spleen status (polysplenia vs asplenia) (Bellchambers and Ware 2018), all of which can cause a wide range of congenital malformations (Zhu et al. 2006; Sutherland and Ware 2009).

Heterotaxy is associated with severe and complex cardiovascular disorders, leading to considerable morbidity and mortality, which can be amplified by issues with the pulmonary, gastrointestinal and genitourinary systems (Zhu et al. 2006; Catana and Apostu 2017). Approximately 3% of congenital heart diseases are attributable to Heterotaxy, which has an overall prevalence of 1/10,000 live births. This may, however, be subject to underestimation and underdiagnoses due to the embryonic lethality, variable expressivity and incomplete penetrance associated with the disorder (Lin et al. 2000; Belmont et al. 2004). Several genes have been associated with human cases of Heterotaxy including *NODAL, CFC1, CRELD1, FOXH1, SESSNI, LEFTYA, GDF1, ACV2B* and *NKX2.5* (Sutherland and Ware 2009; Bellchambers and Ware 2018), however, the gene most commonly mutated in human patients is
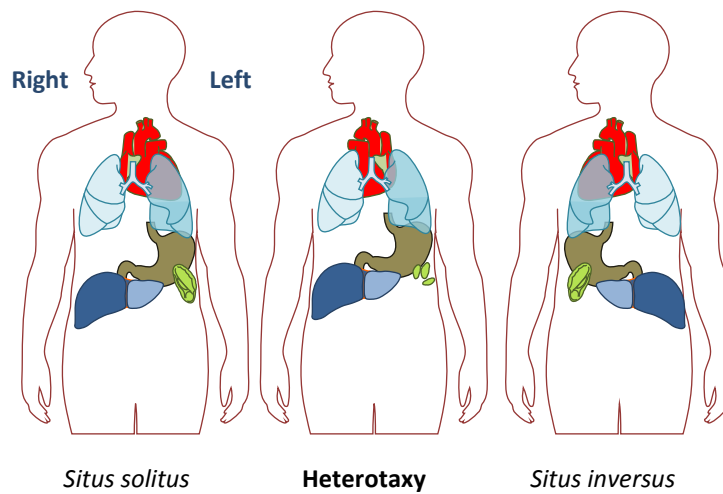
Figure 1.1: Comparison of organ arrangement in Heterotaxy. The normal arrangement of visceral organs is termed *situs solitus,* where organs such as the heart, stomach and spleen are on the left, and the liver on the right. Heterotaxy (also known as *situs ambiguous)* occurs when the heart and visceral organs are oriented independently of one another. *Situs inversus* describes the complete mirror image of internal organs from normal arrangement. Figure adapted from Ruth Arkell.

the X-linked *ZIC3*, with mutations being both familial (>75%) and sporadic (3%) (Gebbia et al. 1997; Sutherland and Ware 2009; Cowan et al. 2014). Mutation of each of these genes also causes aberrant L-R development in the mouse. Due to this, analysis of these mouse strains has identified the steps required for L-R axis formation during embryo development.

1.2 Development of the Left-Right axis

Establishment of the L-R axis occurs during early stages of embryogenesis at a time called gastrulation (6.0 days post coitum [dpc] in the mouse, day 14-17 in humans). A symmetry breaking event at a region of the mouse embryo called the node (or its equivalent structure in other organisms) results in the activation of a signalling cascade on one side of the embryo, leading to a distinct left and right (Figure 1.2). The node is a tear-drop structure located at the distal tip of the embryo, formed by a group of cells derived from the anterior primitive streak at approximately 7.0 dpc. The node is comprised of two distinct cell types. Roughly 200-300 cells, termed pit cells form the tear-drop shaped pit of the node, which are surrounded by 20-30 squamous cells termed crown cells. The pit cells have a single motile cilium that is located on the posterior side of each cell. The cilium tilt to the posterior side (at an angle of approximately 25-40°), protrude into the extracellular space and rotate in a clockwise direction (Okada et al. 2005; Nonaka et al. 2005). The surrounding crown cells also contain a single cilium, though it is not motile. Analysis of mouse models has demonstrated that both motility and cilia length are necessary for the correct development of L-R asymmetry. Embryos with pit cells that lack cilia, or produce cilia that are immotile, randomise L-R axis establishment and present with L-R patterning defects including cardiac abnormalities (Nonaka et al. 1998; Murcia et al. 2000; McGrath et al. 2003). Additionally, mutation of *Rfx3,* a gene that regulates cilia growth, causes significantly shorter cilia than wildtype embryos and results in L-R asymmetry abnormalities (Bonnafe et al. 2004).

Asymmetric gene expression required for correct L-R establishment is created by a process known as nodal flow; the movement of extracellular fluid at the node, generated by cilia rotation, towards the left side of the embryo (Hirokawa et al. 2006) (Figure 1.2). Reversal of nodal flow (by artificially directing fluid to the right side) results in a complete reversal of L-R patterning in cultured mouse embryos (Nonaka et al. 2002). There are, however, two different models of how this signal is processed/interpreted and the mechanisms surrounding this process remain debated amongst developmental biologists.
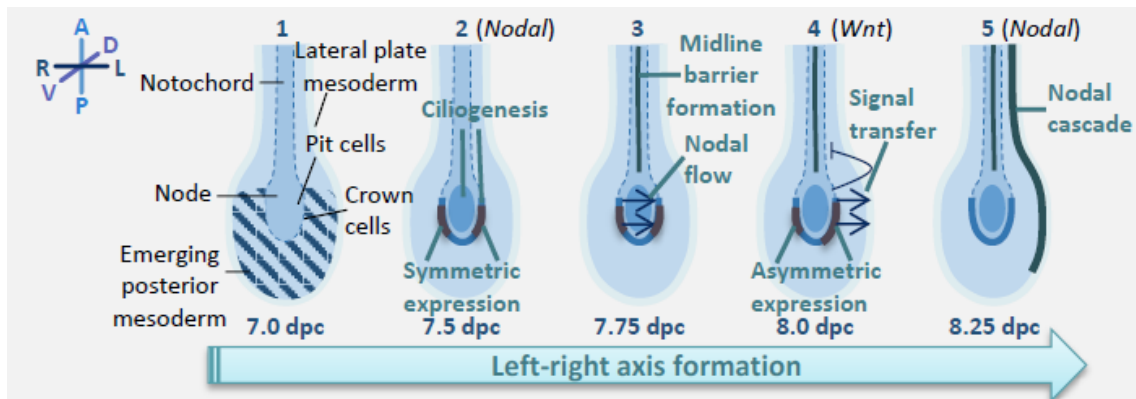
Figure 1.2: L-R symmetry breaking in the murine embryo. The embryonic node is induced to form at mid gastrulation (7.0 dpc) and over 24 hours develops into a shallow, crescent shaped pit consisting of pit and crown cells (Lee and Anderson 2008). **(2)** A monocilium, extending into the extracellular space, forms on the apical surface of each pit and crown cell; these become posteriorly polarised over time. Signalling molecules (such as *Nodal*) are expressed in the crown cells. **(3)** The cilia of the pit cells rotate in a clockwise direction directing leftward flow of extracellular fluid within the node. It is posited that this leftward nodal-flow is sensed by immotile crown cell cilia, prompting a Ca^{2+} flux in the left (but not the right) crown cells which modifies gene expression (Norris 2012). By the end of this 24 hour period (at 8.0 dpc) the first known asymmetries in gene expression are detected within node crown cells (*Dand5*, and soon thereafter *Nodal*, become symmetrically expressed in the node crown). WNT3 ligand is also asymmetrically expressed and canonical signalling then amplifies the initial *Dand5* asymmetry (Nakamura et al. 2012). **(4)** The asymmetric signal(s) are propagated to the left lateral plate mesoderm (LPM) and prevented from spreading to the right LPM by the recently formed midline barrier. **(5)** NODAL signalling in the left LPM controls its own expression, and that of other molecules that ultimately direct organ position and other asymmetries (Norris 2012). A: anterior, P: posterior, D: dorsal, V: ventral, R: right, L: left. Figure taken with permission from Ruth Arkell.

The first model of how the movement of extracellular fluid is translated into asymmetric gene expression is known as the morphogen model. This model suggests that a particular morphogen is secreted in the node cavity and transported to the left side via nodal flow (Nonaka et al. 1998; Okada et al. 2005). This results in a concentration gradient of that morphogen across the node, which is greater on the left side than the right. The high morphogen concentration is detected by cells on the left border of the node and results in the expression of genes such as *Nodal* in the left crown cells, followed by signal transduction triggering asymmetric gene expression in the left lateral plate mesoderm (LPM). The signalling molecule and its corresponding receptors, however, have yet to be identified. An extension of this hypothesis comes from a study by Tanaka et al. 2005, who observed the movement of particles, termed nodal vesicular parcels, in the node cavity. These membrane-bound vesicles are transported to the left of the node and burst upon cilia contact to release their morphogen cargo. The authors propose the transported morphogens include Sonic Hedgehog and Retinoic acid, however, additional studies are needed to corroborate and extend these findings.

The alternative model is known as the two cilia hypothesis. This model proposes that it is the physical stimulation of the flow which is detected by mechano-sensing cilia on the left side crown cells (McGrath et al. 2003). Bending of the sensory cilia by the flow activates an increase of Ca^{2+} signalling in these left-sided cells, which in turn initiates the asymmetric expression of genes such as *Nodal* on the left of the embryo (McGrath et al.2003; Hirokawa et al. 2006). This model is supported by analysis of mice with a mutation in *Pkd2*, a gene which encodes for a Ca^{2+} ion channel that is implicated in polycystic kidney disease. In the kidney, PKD2 interacts with another protein (PKD1) to form a mecaho-transducing complex localised on the immotile cilia of kidney tubule cells. The proteins detect urine flow and generate a Ca^{2+} signal in response (Yoder et al. 2002; Nauli et al. 2003). Although *Pkd1* is not expressed in the embryonic node, a related protein Pkd1-like1 (PKD1L1) is expressed alongside PKD2 in the ciliated cells of the node and is required for response to nodal flow and L-R axis determination in the mouse, implicating Ca^{2+} signalling in L-R patterning (Pennekamp et al. 2002; Field et al. 2011). By analogy, it is proposed that that PKD2 and PKD1L1 form a complex in the left-sided crown cell cilia and upon deflection by nodal flow generate a Ca^{2+} signal. Despite McGrath et al. 2003 observing asymmetric Ca^{2+} signalling in the node of mouse embryos and abnormal Ca^{2+} signalling in mutant embryos, recent studies have challenged this idea and have suggested that Ca^{2+} signalling is not responsible for mechano-sensation of cilia in the node as Ca^{2+} responses were not produced at a physiological level of fluid flow when measured via a transgenic calcium indicator (Delling et al. 2016). Despite their differences, both the morphogen and two cilia models result in a similar outcome – detection of signal on the left side of the embryo and not the right, leading to asymmetric gene expression.

1.3 ZIC3 and human Heterotaxy

To date, 42 mutations in *ZIC3* have been associated with human Heterotaxy in published studies (Gebbia et al. 1997; Mégarbané et al. 2000; Ware et al. 2004; Chhin et al. 2007; Wessels et al. 2010; De Luca et al. 2010; Ma et al. 2012; D'Alessandro et al. 2013a, b; Cowan et al. 2014; Paulussen et al. 2016; El Malti et al. 2016; Li et al. 2018). These include 18 missense mutations, 8 nonsense, 9 frameshift, 6 in-frame deletions/duplications and a single intronic mutation (Figure 1.3). Just over half of the mutations (22/42) occur within the highly conserved zinc finger domain (ZFD) of the ZIC3 protein. However, there are eight nonsense and frameshift mutations that occur prior to the ZFD that are predicted to cause protein truncation either before or within the ZFD, thus increasing the total number of mutations affecting the ZFD (either directly or indirectly) to 30/42 (71%). The ZIC ZFD has been assigned many functional properties including DNA binding, protein-protein interaction and nuclear localisation (Bedard et al. 2007; Lim et al. 2010; Pourebrahim et al. 2011), and therefore, mutations that occur within this domain are predicted to negatively affect or even ablate protein function. It is unknown, however, if all 42 identified ZIC3 human mutations are indeed pathogenic, nor is it known the exact mechanism by which pathogenic mutations lead to Heterotaxy.
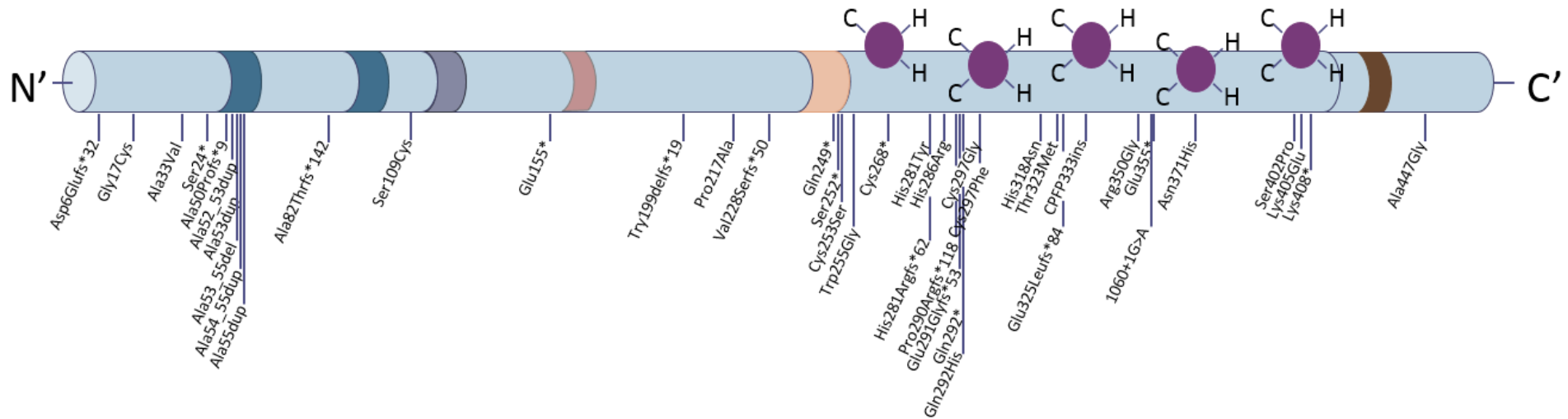
Figure 1.3: *ZIC3* alleles in human Heterotaxy. The location of all 42 known human mutations in *ZIC3* that are associated with Heterotaxy in humans is shown. These include 18 missense, 9 frameshift 8 nonsense, 6 in-frame expansions or deletions and 1 intron mutation. Of these mutations, 22 (52%) occur within the highly conserved zinc finger domain of the ZIC3 protein. Experimental analysis is needed to determine if all 42 mutations are pathogenic. C: Cysteine, H: Histidine, N': N-terminus, C': C-terminus.

1.4 ZIC3 is associated with Heterotaxy in the mouse

Zic3 null mice recapitulate human Heterotaxy phenotypes and therefore provide a good model of ZIC3 protein function. In addition to Heterotaxy, *Zic3* null mice display gastrulation defects, neural tube defects and axial patterning abnormalities (Purandare et al. 2002). Targeted construction of conditional mouse alleles has enabled the deletion of *Zic3* in specific cells and facilitated the analysis of the temporal and spatial requirements of *Zic3* during development. Genetic data from two independent research groups (summarised in Table 1.1) demonstrated ZIC3 is not required in the developing heart or node cells to prevent Heterotaxy (Table 1.1) (Jiang et al. 2013; Sutherland et al. 2013). For example, despite the broad range of cardiac defects observed in *Zic3* null models, removal of *Zic3* from cardiac progenitors using Cre strains *Mesp1-Cre*, *Nkx2.5-Cre*, *Mef2c-Cre*, *Wnt1-Cre*, or *βMyHC-Cre* did not result in cardiac defects. Similarly, deletion of *Zic3* in the lateral crown cells or the ciliated node pit cells did not result in Heterotaxy (Table 1.1, *Nodal-dependant enhancer-Cre,* [*NDE-Cre*], *Foxj1-Cre*). Instead, *Zic3* is required in the ectoderm or nascent mesoderm during gastrulation to prevent L-R axis defects. Deletion of *Zic3* in cells expressing Brachyury (*T-Cre*) displayed increased lethality, heart looping defects and situs abnormalities. *T-Cre* is expressed in the migrating primitive streak before node development, which suggests *Zic3* is required prior to node development to prevent L-R axis defects (Perantoni 2005; Sutherland et al. 2013). Laterality defects were also observed when *Zic3* was deleted from cells in the pluripotent epiblast (*Sox2-Cre*) (Jiang et al. 2013). Ultimately this identifies a role for ZIC3 during early gastrulation before node formation to prevent Heterotaxy.

Table 1.1 Temporal and spatial requirement of ZIC3 in the prevention of L-R axis defects occurs prior to node formation. L-R: left-right

Structure	Allele symbol; Cre target	Cells targeted	Temporal expression of Cre target	L-R axis defects	Reference
Ectoderm	*Zic3$^{del/y}$; Sox2-Cre+*	Cells in the pluripotent ectoderm	Prior to and during gastrulation	Present	(Jiang et al. 2013)
Nascent mesendoderm	*Zic3$^{del/y}$; T-Cre+*	Cells migrating through the wings of mesoderm	During gastrulation and mesoderm migration	Present	(Jiang et al. 2013; Sutherland et al. 2013)
Node	*Zic3$^{del/y}$; NDE-Cre+*	Lateral crown cells of the node	During L-R symmetry breaking	Absent	(Sutherland et al. 2013)
	Zic3$^{del/y}$; FoxJ1-Cre+	Ciliated node pit cells	During L-R symmetry breaking	Absent	(Sutherland et al. 2013)
Lateral plate mesoderm	*Zic3$^{del/y}$; Mesp1-Cre+*	Lateral plate mesoderm anterior to the node	Early cardiac development	Absent	(Jiang et al. 2013)
Heart	*Zic3$^{del/y}$; Mef2c-Cre+*	Anterior heart field	Cardiac development	Absent	(Sutherland et al. 2013)
	Zic3$^{del/y}$; Nkx2.5-Cre+	Cardiac cresent and cardiomyocytes	Cardiac development	Absent	(Jiang et al. 2013; Sutherland et al. 2013)
	Zic3$^{del/y}$; Wnt1-Cre+	Cardiac neural crest cells	Cardiac development	Absent	(Jiang et al. 2013; Sutherland et al. 2013)
	Zic3$^{del/y}$; βMyHC-Cre+	Embryonic and fetal cardiomyocytes	Late cardiac development	Absent	(Sutherland et al. 2013)

1.5 The ZIC family of transcription factors

The Zinc finger of the cerebellum (*Zic*) genes encode a family of multifunctional proteins that regulate gene expression (Ali et al. 2012; Houtmeyers et al. 2013). The *Zic* genes were first identified by their homology to the *Drosophila* gene *odd-pared* (*opa*; Aruga et al. 1994), while *opa* itself was isolated in a genetic screen for mutations that disrupt segmentation of the *Drosophila* embryo (Nüsslein-Volhard and Wieschaus 1980). Analysis of mutant flies identified *opa* as a pair-rule gene and subsequent cloning revealed *opa* also regulates the expression of genes required for the development of visceral mesoderm and midgut morphogenesis (Benedyk et al. 1994; Cimbora and Sakonju 1995). Sequence analysis showed that *opa* contains a zinc finger domain (ZFD) made up of five C_2H_2 zinc fingers. As these domains are commonly associated with DNA binding, *opa* was assumed to be a transcription factor. The first vertebrate *Zic* gene (*Zic1*) was identified by cDNA screening for genes differentially expressed in the cerebellum (Aruga et al. 1994). Subsequently, *Zics2-4* were identified in the mouse on the basis of their homology to *Zic1* (Aruga et al. 1996a, b). We now know that in mice, as in man, there are five *ZIC* genes (Furushima et al. 2000), each of which encodes a transcriptional regulator protein containing a ZFD.

1.5.1 Genome organisation

The genomic arrangement, as well as the gene and protein structure of the Zic molecules, has been conserved throughout evolution in both vertebrate and invertebrate species (Aruga et al. 2006). Genome sequencing of a large number of vertebrate organisms found that the *Zic* genes are arranged in tandem pairs. In the human genome, the five *ZIC* genes reside at three locations. *ZIC1* and *ZIC4* are arranged as a divergently transcribed tandem gene pair on chromosome 3, as are *ZIC2* and *ZIC5* on chromosome 13, while *ZIC3* is an X-linked singleton (Figure 1.4) (Ali et al. 2012; Houtmeyers et al. 2013). Furthermore, sequencing and phylogenetic analysis have revealed that the *ZIC* genes can be divided into two subgroups (Subgroup A containing *ZIC1-3* and Subgroup B containing *ZIC4-5*), based on similarities in nucleotide sequence and protein structure (Aruga et al. 2006; Houtmeyers et al. 2013).

Aruga *et al.*, 2006 provides a model of *Zic* gene evolution, where a single ancestral *Zic* gene (with a single intron) underwent a tandem duplication event followed by sequence modification. This generated two isoforms in which one acquired an additional intron and the other underwent sequence divergence with the loss of structural domains (including the ZOC domain) and the divergence of zinc finger 1 (Houtmeyers et al. 2013). This was followed by two genome duplication events that would
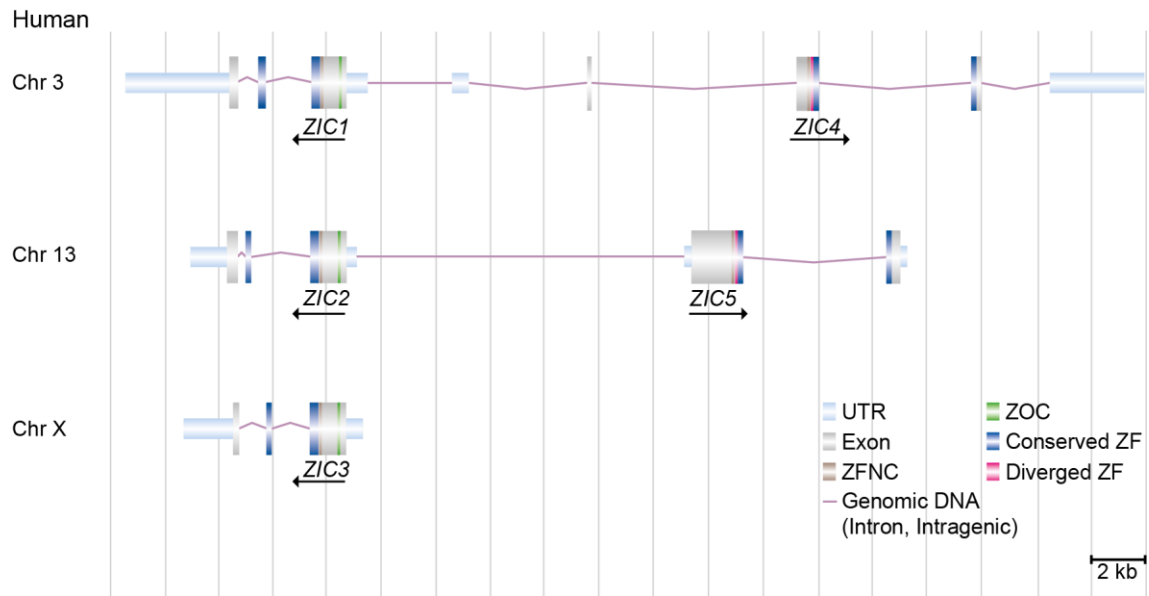
Figure 1.4: Genomic arrangement of human ZIC genes. The arrangement of *ZIC* genes on the human genome, as shown in the Ensembl genome browser. *ZIC1* and *ZIC4* are located on chromosome (Chr) 3, *ZIC2* and *ZIC5* are located on chromosome 13, while *ZIC3* is on the X-chromosome. Each bi-gene clusters contains a *ZIC* gene from Subgroup A (*ZIC1, ZIC2, ZIC3*) and Subgroup B (*ZIC4, ZIC5*). UTR: untranslated region; ZFNC: zinc finger N-terminal conserved region; ZOC: Zic-Opa conserved region; ZF: zinc finger. This figure was taken from Houtmeyers et al. 2013.

have produced eight *Zic* genes organised in four bi-gene clusters, with each containing a *Zic* gene from Subgroup A and Subgroup B. Due to gene loss or de-novo duplication events, the number of *Zic* genes vary in number between species, from the single *opa* gene found in *Drosophila* to seven *zic* genes in Zebrafish and five *Zic* genes present in *Xenopus*, mouse and man.

1.5.2 Zic Pleiotropy

Although the vertebrate *Zic* genes were named based on the tissue from which they were first isolated (murine cerebellum), animal and human genetics show the five genes often function outside of brain development and instead are necessary during early embryonic development (Table 1.2) (Aruga et al. 1994; Houtmeyers et al. 2013; Diamand et al. 2018). Analysis of multiple murine *Zic* alleles yielded an extensive list of processes, tissues and organs that require the function of one or more murine *Zic* genes for normal development. Processes affected by the absence of *Zic* gene function include axis formation, as well as the patterning and morphogenesis of the central nervous, visual, musculo-skeletal and vascular systems (Table 1.2) (Diamand et al. 2018). Analysis of these alleles not only illustrates the diverse function of the proteins, but highlights the fact that different *Zic* alleles give rise to the same phenotype. For example, mutation of *Zic1*, *Zic2*, *Zic3* and *Zic5* all produce skeletal abnormalities, while all ZICs are involved in forebrain development, particularly in cerebellum size and patterning.

Moreover, the catalogue of *Zic* related phenotypes shown in Table 1.2 makes clear the remarkable extent of *Zic* gene pleiotropy, a term which refers to the production of two or more apparently unrelated effects by a single gene. For example, *Zic1* and *Zic4* mouse mutants are small at birth and have cerebellar abnormalities. Similarly, mutation of *Zic5* affects diverse systems or processes such as neurulation, neural crest development, skeletal and eye development. *Zic3* mutation, like *Zic2,* affects a broad range of tissues or processes, some of which are linked by their ontogeny (for example, the limb and tail defects may both reflect a role in skeletal development) and others which arise from a role in multiple tissues (such as craniofacial, cerebellar, neural tube and ventral closure defects). Overall, the extent of *Zic* gene pleiotropy implies multiple sites of expression and overlapping expression of *Zic* family members during embryonic development, and a multitude of molecular roles for the ZIC proteins during this time.

Table 1.2: Murine *Zic* loss-of-function phenotypes.

	Zic phenotypes	Genotypes	References
Gastrulation	Primitive Streak Dysgenesis	$Zic3^{-/-}$, $Zic3^{-/Y}$	(Ware et al. 2006a)
	Failure of Anterior Notochord Production	$Zic2^{Ku/Ku}$	(Warr et al. 2008)
	Left-right axis defect	$Zic2^{Ku/Ku}$ $Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$, $Zic3^{Bn/+}$, $Zic3^{-/-}$, $Zic3^{-/Y}$ $Zic3^{Ka/Ka}$, $Zic3^{Ka/Y}$	(Carrel et al. 2000; Purandare et al. 2002; Ahmed et al. 2013; Barratt et al. 2014)
	Node cilia defects	$Zic2^{Ku/Ku}$	(Barratt et al. 2014)
Patterning and Morphogenesis	Exencephaly	$Zic2^{Ku/Ku}$, $Zic2^{kd/kd}$, $Zic2^{m1Nisw/m1Nisw}$ $Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$, $Zic3^{-/-}$, $Zic3^{-/Y}$, $Zic5^{-/-}(a)$, $Zic5^{-/-}(b)$	(Zhang and Niswander; Nagai et al. 2000; Klootwijk et al. 2000; Purandare et al. 2002; Elms et al. 2003; Inoue et al. 2004; Furushima et al. 2005)
	Spina Bifida (and curled tail)	$Zic2^{Ku/+}$, $Zic2^{kd/kd}$, $Zic2^{m1Nisw/m1Nisw}$ $Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$	(Zhang and Niswander; Garber 1952; Nagai et al. 2000; Klootwijk et al. 2000; Elms et al. 2003)
	Omphalocele	$Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$	(Klootwijk et al. 2000)
	Cleft lip	$Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$	(Klootwijk et al. 2000)
	Cranial Neural crest deficit	$Zic2^{Ku/Ku}$, $Zic2^{kd/kd}$ $Zic5^{-/-}(a)$, $Zic5^{-/-}(b)$	(Nagai et al. 2000; Elms et al. 2003; Inoue et al. 2004; Furushima et al. 2005)
	Trunk Neural Crest Deficit	$Zic2^{Ku/Ku}$	(Elms et al. 2003)
	Hindbrain patterning	$Zic2^{Ku/Ku}$	(Elms et al. 2003)
Skeletal Defects	Craniofacial	$Zic5^{-/-}(a)$, $Zic5^{-/-}(b)$	(Inoue et al. 2004; Furushima et al. 2005)
	Sternum/ribs	$Zic5^{-/-}(a)$, $Zic1^{-/-}$	(Aruga et al. 1999; Inoue et al. 2004)
	Vertebral arches	$Zic1^{-/-}$, $Zic2^{kd/kd}$, $Zic2^{kd/+};Zic3^{Bn/Y}$ $Zic3^{Bn/X}$, $Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$, $Zic3^{-/+}$, $Zic3^{-/-}$, $Zic3^{-/Y}$, $Zic5^{-/-}(a)$	(GrüNeberg 1963; Aruga et al. 1999; Nagai et al. 2000; Purandare et al. 2002; Inoue et al. 2004, 2007)
	Tail (reduced number and or dysmorphic caudal vertebrae)	$Zic2^{kd/kd}$, $Zic3^{Bn/X}$, $Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$, $Zic3^{-/+}$, $Zic3^{-/-}$, $Zic3^{-/Y}$ $Zic3^{Ka/+}$ $Zic5^{-/-}(a)$	(GrüNeberg 1963; Purandare et al. 2002; Inoue et al. 2004; Ahmed et al. 2013)
	Limb	$Zic2^{kd/kd}$	(Nagai et al. 2000)
	Classical HPE	$Zic2^{Ku/Ku}$,	(Warr et al. 2008)
	Roof-plate abnormality/MIH HPE	$Zic2^{kd/kd}$, $Zic5^{-/-}(b)$	(Nagai et al. 2000; Furushima et al. 2005)
	Dorsal spinal cord hypoplasia	$Zic1^{-/-}$	(Aruga et al. 1998)

Organogenesis			
	Forebrain		
	Dysgenesis of medial structures	$Zic1^{-/-}$, $Zic1^{-/-};Zic3^{Bn/Y}$ $Zic3^{-/+}$, $Zic3^{-/-}$, $Zic3^{-/Y}$, $Zic3^{Bn/Bn}$ $Zic5^{-/-}$ (a)	(Purandare et al. 2002; Inoue et al. 2004; Furushima et al. 2005)
	Enlarged ventricles	$Zic5^{-/-}$ (b)	(Furushima et al. 2005)
	Microcephaly	$Zic2^{kd/kd}$, $Zic5^{-/-}$ (a)	(Nagai et al. 2000; Furushima et al. 2005)
	Cerebellum: Reduced Size	$Zic1^{-/-}$, $Zic1^{-/+}; Zic4^{-/+}$ $Zic1^{-/-};Zic4^{-/-}$ $Zic3^{Bn/Y}$, $Zic3^{Bn/Bn}$	(Aruga et al. 1998; Aruga 2004; Grinberg et al. 2004; Blank et al. 2011)
	Cerebellum: Folia Patterning	$Zic1^{-/-}$, $Zic1^{-/+};Zic2^{kd/+}$ $Zic1^{-/+};Zic4^{-/+}$ $Zic1^{-/-};Zic4^{-/-}$, $Zic1^{-/+};Zic2^{Kd/+}$ $Zic3^{Bn/Y}$, $Zic3^{Bn/Bn}$	(Aruga et al. 1998, 2002; Aruga 2004; Grinberg et al. 2004; Blank et al. 2011),
	Non-forebrain		
	Hydrocephalus	$Zic1^{-/+};Zic2^{kd/+}$ $Zic5^{-/-}$ (a)	(Aruga et al. 2002; Inoue et al. 2004)
	Eye: Micro and anopthalmia	$Zic2^{kd/kd}$, $Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$ $Zic5^{-/-}$ (b)	(Klootwijk et al. 2000; Herrera et al. 2003; Furushima et al. 2005)
	Eye: Optic chiasm: morphology	$Zic2^{kd/kd}$	(Herrera et al. 2003)
	Eye: aberrant ipsilateral projection of retinal ganglion cells	$Zic2^{kd/+}$, $Zic2^{kd/kd}$	(Herrera et al. 2003)
	Limb	$Zic2^{kd/kd}$ $Zic3^{Bn/Y}$, $Zic3^{Bn/Bn}$,	(Garber 1952)
	Small size at birth	$Zic1^{-/+};Zic4^{-/+}$, $Zic1^{-/-}; Zic4^{-/-}$ $Zic2^{kd/kd}$, $Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$ $Zic5^{-/-}$	(Garber 1952; Nagai et al. 2000; Inoue et al. 2004; Grinberg et al. 2004)
	Abnormal intestine morphology	$Zic2^{m1Nisw/m1Nisw}$	(Zhang and Niswander)
	Abnormal enteric neural crest morphology	$Zic2^{m1Nisw/m1Nisw}$	(Zhang and Niswander)
	Abnormal neurite morphology	$Zic2^{m1Nisw/m1Nisw}$	(Zhang and Niswander)

Post Birth	Embryonic lethality	$Zic2^{Ku/Ku}$, $Zic3^{Bn/Bn}$, $Zic3^{Bn/Y}$, $Zic3^{-/-}$, $Zic3^{-/Y}$	(Garber 1952; Purandare et al. 2002; Elms et al. 2003)
	Peri-natal lethality	$Zic2^{kd/kd}$, $Zic2^{m1Nisw/m1Nisw}$ $Zic5^{-/-}$(a)	(Zhang and Niswander; Nagai et al. 2000; Inoue et al. 2004)
	Juvenile lethality	$Zic1^{-/+}$;$Zic2^{kd/+}$ (4 weeks), $Zic1^{-/+}$;$Zic4^{-/+}$ (4 weeks), $Zic1^{-/-}$;$Zic4^{-/-}$, $Zic1^{-/-}$, $Zic2^{kd/kd}$ $Zic5^{-/-}$ (4 weeks) (a)	(Nagai et al. 2000; Aruga et al. 2002; Inoue et al. 2004; Grinberg et al. 2004; Blank et al. 2011)
Behavioural	Pre-pulse inhibition	$Zic2^{+/kd}$	(Ogura et al. 2001)
	Motor control	$Zic1^{-/+}$, $Zic1^{-/-}$, $Zic1^{-/+}$;$Zic4^{-/+}$, $Zic1^{-/-}$;$Zic4^{-/-}$ $Zic3^{Bn/Y}$, $Zic3^{Bn/Bn}$,	(Ogura et al. 2001; Aruga 2004; Grinberg et al. 2004)
	Abnormal gait and posture	$Zic1^{-/+}$; $Zic2^{kd/+}$ $Zic5^{-/-}$ (a)	(Aruga et al. 2002; Inoue et al. 2004)

(a) $Zic5^{Tm1Jaru}$ strain, (b) $Zic5^{Tm1Sia}$strain

34

1.5.3 Expression during early embryo development

The temporal-spatial distribution of murine *Zic* gene transcripts has been documented across a range of developmental stages and (as reviewed in Houtmeyers et al. 2013) the murine *Zic* genes can be divided into two groups based on their overlapping expression patterns during early embryo development. *Zic2, Zic3* and *Zic5* are the first of the gene family to be expressed, with transcription of each initiating prior to implantation. During gastrulation, *Zic2, Zic3* and *Zic5* are co-expressed in the ectoderm and some of the newly formed mesoderm that ingresses through the primitive streak, whilst *Zic1* and *Zic4* are not yet detected (Figure 1.5) (Furushima et al. 2000; Elms et al. 2004; Houtmeyers et al. 2013). By the early head-fold stage, all three expressed genes are restricted to the neurectoderm and later become limited to the most dorsal neurectoderm that will form the future neural crest cells and dorsal neurons, as well as pre-somitic and somatic populations of the lateral mesoderm. In contrast, *Zic1* and *Zic4* expression initiate during early organogenesis in the neurectoderm and somitic mesoderm. By 9.5 dpc, all five murine *Zic* genes are co-expressed in the dorsal spinal cord, the dorsal cranial neural tube and in the somites, with small differences in expression domains such as the absence of *Zic4* expression in the roof-plate of the neural tube (Figure 1.5) (Nagai et al. 1997; Furushima et al. 2000; Elms et al. 2004). One day later, the expression of *Zic2*, *Zic3* and *Zic5* is initiated in newly developing eye and in the limb buds. At later stages of development, *Zic* gene expression has not been assayed in the entire embryo, but brain expression patterns have been reported. The progenitor cells of the thalamus, preoptic area, the septum, cortical hem and the retina all express the *Zic* genes in largely overlapping domains at mid-gestation. *Zic* expression in the central nervous system becomes more restricted by embryonic day 12.5 at which time all five *Zic* genes are expressed in the dorsal midline neural tissues (including the cortical hem, the septum, and the ventricular and sub-ventricular zone) and ventral neural tube (Figure 1.5). Interestingly, as shown in Figure 1.5, these expression groups mirror the genomic arrangement of the genes. The *Zic1/4* bi-gene pair form one expression group, whilst the single gene *Zic3* and the *Zic2/5* bi-gene pair form another. It is possible that the bi-gene pairs share regulatory regions as a direct result of their evolved tandem gene arrangement, resulting in similar gene expression patterns (Houtmeyers et al. 2013).

7.5 dpc

Zic1 *Zic4*

8.5 dpc

Zic1 *Zic4*

9.5 dpc

Zic1 *Zic4*

12.5 dpc

Zic1 *Zic4*

Subclass A

Subclass B

Zic5

Zic2 *Zic5*

Zic2 *Zic5*

Zic2 *Zic5*

Zic3

Zic3

Zic3

Zic3
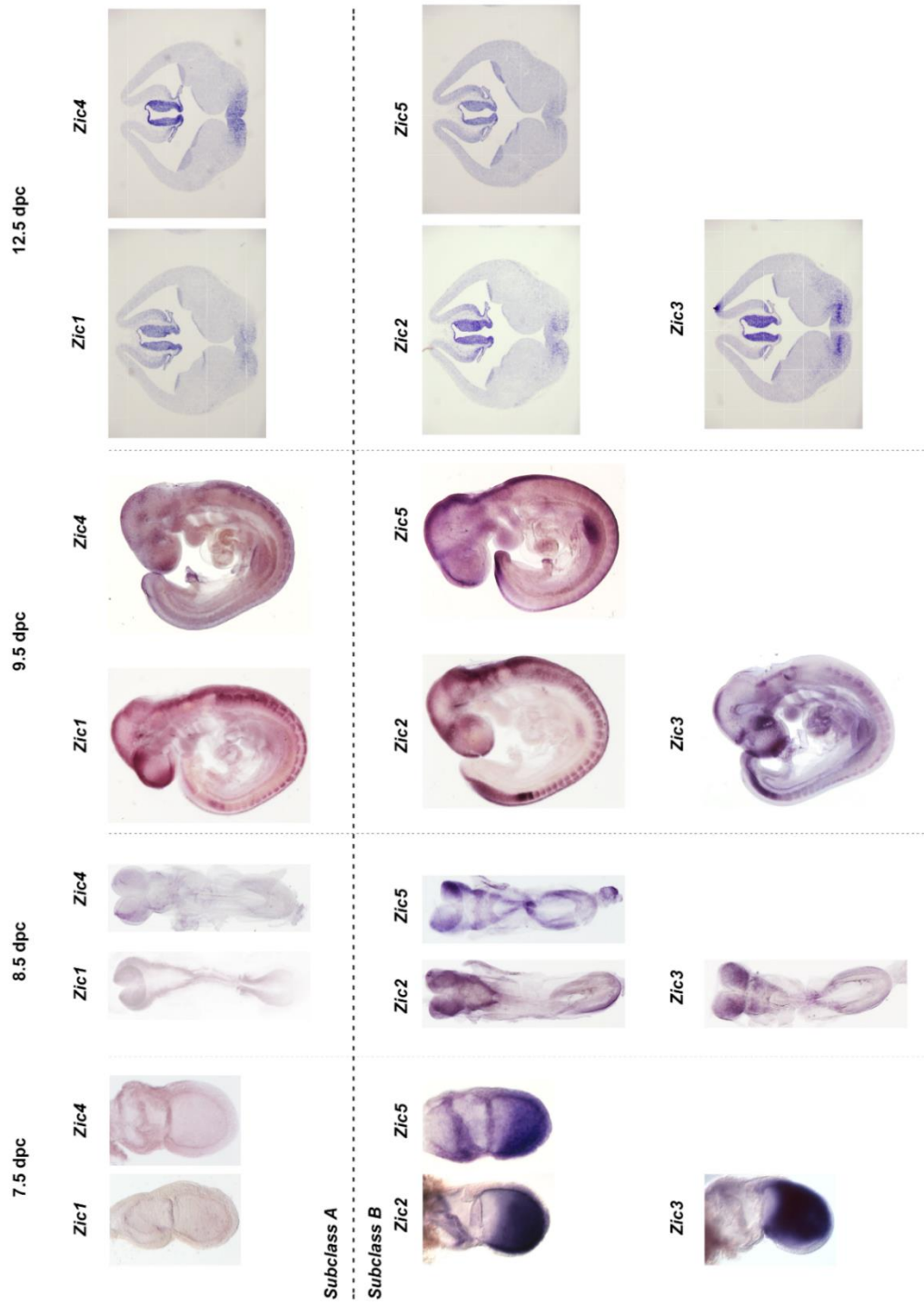
Figure 1.5: Co-expression of the murine *Zic* genes during mouse gastrulation and organogenesis. The murine *Zic* gene expression subclasses reflect their genomic arrangement. Subclass B genes are expressed prior to and throughout gastrulation, during neurulation, and in organogenesis. Subclass A gene expression does not initiate until after neurulation and is absent from the optic cup, the forelimb bud and the pre-somitic mesoderm at early organogenesis. **(A)** Lateral view of 7.5 dpc mouse embryos (anterior is to the left), **(B)** dorsal view of 8.5 dpc mouse embryos (anterior is to the top) and **(C)** lateral view of 9.5 dpc mouse embryos (anterior is to the left) following whole mount in situ hybridization to *Zic1-5* (courtesy of Kristen Barratt). **(D)** Coronal sections of 12.0 dpc mouse forebrains through the hippocampus and habenular nucleus following whole-mount in situ hybridization to *Zic1-5* (courtesy of Jun Aruga). Figure taken from Diamand et al. 2018.

1.5.4 ZIC molecular functions

1.5.4.1 ZIC proteins function as transcription factors

As previously described, the *ZIC* genes are orthologs of the *Drosophila* transcription factor encoding pair-rule gene, *odd-paired* (*opa*). The defining feature of all ZIC proteins (including opa) is the highly conserved ZFD, which is able to bind DNA, meaning the *ZIC* genes are presumed to encode DNA binding transcription factors (Figure 1.6). Transcription factors play essential roles in gene regulation by controlling the activation and repression of target genes. To achieve this, they contain a DNA binding domain that recognises and binds to short but specific DNA sequences in the enhancer or promotor region of a target gene (Wingender et al. 2013). Eukaryotic transcription factors are classified into 10 different superclasses based on the DNA binding domains they possess, with the largest superclass being the zinc coordinating DNA binding domain which encompasses 52% of all transcription factors investigated, which include the ZIC proteins (Wingender et al. 2013, 2018). Transcription factors can also bind within nucleosomal DNA and trigger chromatin remodelling and nucleosome positioning (which can increase accessibility for other proteins), which also influences gene expression (Spitz and Furlong 2012). Moreover, transcription factors contain other regulatory domains, called transactivation domains, which interact with the basal transcription machinery, recruit co-factors via protein binding domains and form transcriptional complexes at enhancers to co-operatively modulate the transcription of gene targets. Ultimately transcription factors have a bipartite structure requiring both a DNA binding domain and transactivation domain to regulate target gene expression.

Nuclear localisation

Transcription factors recognise and bind to specific DNA binding sites, a process that occurs in the nucleus. Transcription factors therefore need a mechanism for nuclear entry. This process is usually facilitated by the presence of specific amino acid sequences within the protein known as nuclear localisation signals (NLSs), which are recognised by and interact with nuclear import receptors present on the nuclear membrane (Jans et al. 2000). Generally, NLSs are composed of multiple monopartite or bipartite clusters of positively charged basic amino acids such as arginine and lysine (Hicks and Raikhel 2003). Sequence analysis of the ZIC proteins, however, shows that none contain a canonical nuclear localisation signal. Instead, experimental analysis of human ZIC3 via two independent studies found the ZFD is necessary for nuclear import. For example, Bedard et al. 2007 reported that 90% of wildtype ZIC3 transfected in NIH3T3 cells was localised to the nucleus. When putative NLS mutations were made in the ZFD, nuclear localisation of the mutant proteins was reduced to 20-30%. Similarly,
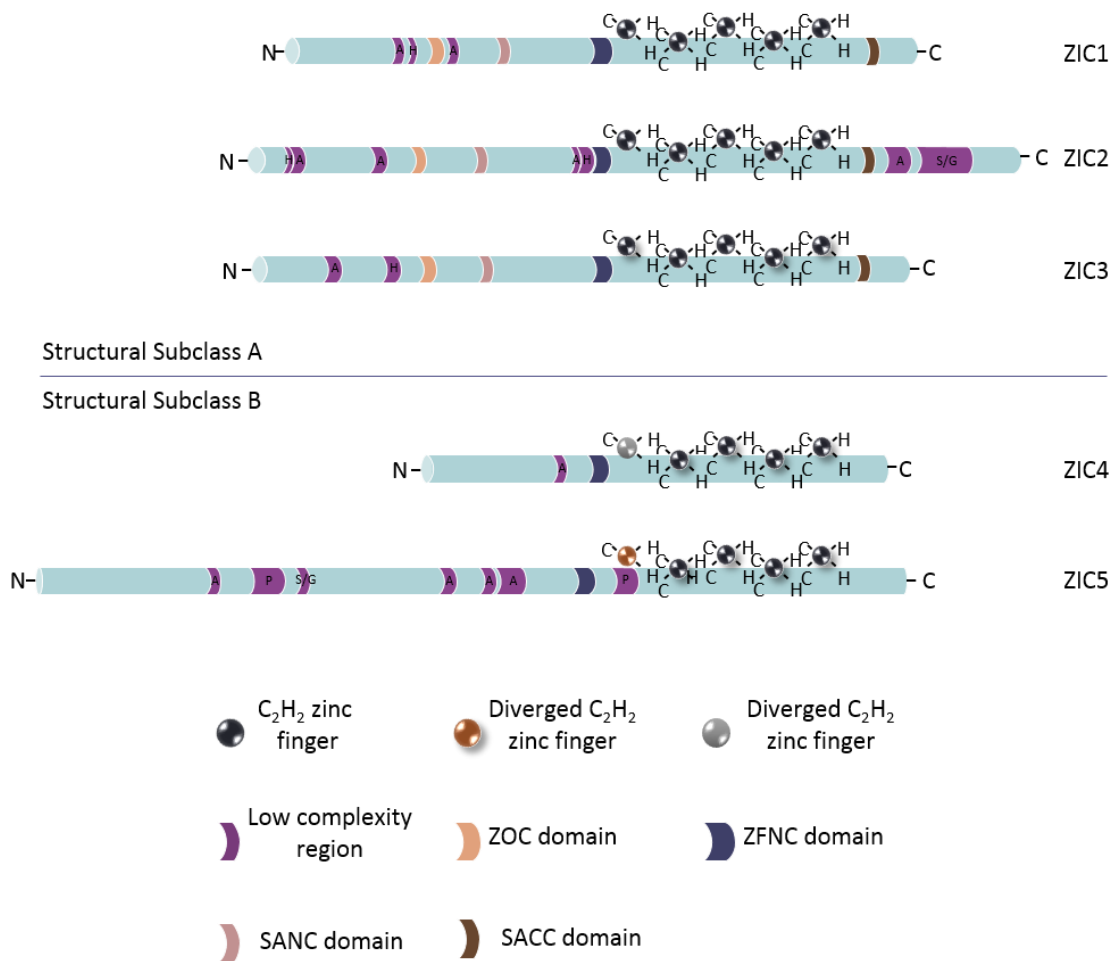
Figure 1.6: Structural features of the human ZIC proteins. All five ZIC proteins contain a zinc finger domain comprised of five tandem C_2H_2 type zinc fingers. This domain is highly conserved however ZIC4 and ZIC5 each contain a diverged first zinc finger. All proteins contain a conserved domain (14-21 amino acids) directly upstream of the zinc finger domain termed the ZFNC. ZIC1-3 contain the highly conserved ZOC domain, a short (9-10 amino acids) region in the N-terminus. Functions of the ZOC and ZFNC are not yet clearly understood. Several low complexity regions (stretches of alanine, proline, histidine, serine and glycine) also exist with the type and number varying between all five proteins. A: alanine, H; histidine, P: proline, S/G: serine/glycine, ZOC: Zic-opa conserved, ZFNC; zinc finger N-terminus conserved, C_2H_2: Cysteine-two, Histidine-two.

Hatayama et al. 2008 made several deletion constructs of ZIC3, lacking various zinc fingers of the ZFD, that when transfected into NIH3T3 cells, perturbed nuclear localisation. Together it is suggested the ZFD (fingers 2-5) is essential for nuclear localisation and presumably harbours an interspersed nuclear localisation signal (Bedard et al. 2007; Hatayama et al. 2008).

DNA binding domain

The ZIC DNA binding domain is most closely related to the ZFD of the GLI, GLIS and NKL families and is highly conserved between ZIC family members both within an individual species and across species (Aruga 2004). C_2H_2 type zinc fingers contain two cysteine (C) and two histidine (H) residues that tetrahedrally co-ordinate a zinc (Zn^{2+}) ion and are often involved in DNA binding (Iuchi 2001). The ZIC ZFD is, however, distinguished by an atypical first zinc finger. Often, one to five amino acid residues are located between the two cysteines of a C_2H_2 zinc finger. In the ZIC proteins, however, the first zinc finger can contain between 6 to 38 amino acids, varying between all species so far examined (Aruga et al. 2006; Layden et al. 2010). Additionally, structural analysis of the ZIC3 protein indicates that the first two zinc fingers of the ZFD may not be canonical (DNA binding) C_2H_2 type zinc fingers and instead form a single structural unit called the tandem $CWCH_2$ motif. This motif is characterised by the presence of a tryptophan residue located between the two canonical cysteines of each zinc finger (Hatayama and Aruga 2010). Additionally, the $CWCH_2$ motif is conserved across a wide range of metazoan species (Aruga et al. 2006), which is indicative of biological significance. Moreover, a missense mutation of the tryptophan in the first zinc finger of ZIC3 is associated with congenital heart malformation in humans. Both protein stability and nuclear localisation are affected by this mutation when overexpressed in NIH3T3 cells (Chhin et al. 2007). The function of the $CWCH_2$ motif is not yet fully understood, however, it has been suggested that it may modulate DNA binding by zinc fingers 3-5 or be involved in protein-protein interactions (Hatayama and Aruga 2010; Houtmeyers et al. 2013).

Despite the presence of a ZFD, little is known about ZIC *in vivo* downstream targets. Studies using embryonic stem cells have provided some evidence for the DNA binding properties of ZIC3 and its involvement in the regulation of stem cell pluripotency. For example, using ChIP-on-chip experiments, Lim et al. 2010 found 379 *Zic3*-enriched promotor targets including those of *Oct4*, *Sox2* and *Nanog*. Further analysis by Lim et al (2010) identified ZIC3 is capable of binding and activating the *Nanog* promoter using *in vitro* DNA binding assays, demonstrating its potential as a transcriptional activator. A protein binding microarray identified the same consensus sequence identified by Lim et al (2010), to be bound by the ZFD of ZIC1-3 (Badis et al. 2009). Furthermore, mutations located within the ZFD of many ZIC proteins can result in a loss of DNA binding and are associated with severe congenital defects (Ware et al. 2004; Brown et al. 2005). In many cases, however, the DNA binding ability of these alleles was tested with a purified protein corresponding only to the ZFD portion of the ZIC protein

(indeed, purification of a full-length ZIC protein has never been reported). While it remains possible that the full-length variant proteins may perform differently in the same DNA binding assay, overall there is good evidence that the ZIC proteins can bind DNA to exert transcription control and that the ZFD is required for this function

Transactivation

Aside from the DNA binding domain, transcription factors also require a transactivation domain. This region of the protein is separate to the DNA binding domain and interacts with other proteins and transcriptional machinery necessary to regulate transcription of target genes. To date, the ZIC transactivation domain has not yet been identified. It is expected an important functional domain such as this would be conserved between the ZIC proteins, similar to the ZIC DNA binding domain conservation. Although a specific transactivation domain has not yet been identified within the ZIC proteins, there are many evolutionary conserved domains within the ZICs with unknown function (Figure 1.6), that may have a role in transactivation.

The best characterised evolutionary conserved domain is called the *ZIC opa* conserved (ZOC) domain, which is a small (9-10 amino acids) region located in the N-terminus of the protein conserved in ZIC1-3 (Figure 1.6). This domain is one of two shown to be involved in transcriptional activation (the other being the ZFD), and has also been shown to be involved in protein-protein interactions, such as binding the myogenic repressor protein, I-mfa (Mizugishi et al. 2004). Additionally, a small (14-21 amino acids) conserved region, located directly prior to the ZFD called the zinc finger N-flanking conserved (ZFNC) domain is found in all five ZIC proteins and is subject to the post-translational modification SUMOylation in human ZIC3 (Aruga et al. 2006; Chen et al. 2013). As reviewed in Houtmeyers et al. 2013 phylogenetic comparison of the ZIC protein sequences has led to the classification of ZICs into two distinct structural subclasses based on the presence of the ZOC domain and the variation within the first zinc finger. The structural subclass A contains the more highly conserved ZIC1, 2 and 3 proteins and subclass B contains the less conserved ZIC4 and ZIC5 proteins (Figure 1.6). Notably, the subclass division is reflected in the genome arrangement and evolution of the *Zic* genes, since each gene pair (i.e. *Zic1/Zic4* and *Zic2/Zic5*) contains one subclass A and one subclass B protein (Figure 1.5, 1.6).

Recently, two previously unannotated domains were identified in the Arkell laboratory based on sequence conservation (Arkell lab, unpublished). These domains are only present in subclass A (ZICs1-3) proteins and were subsequently named SANC (Subclass A N terminally conserved) and SACC (Subclass A C terminally conserved). The SANC domain is made of 10 amino acids (amino acid position 159-168 within ZIC3), while the SACC domain is made of 14 amino acids (amino acid position 418-431 within ZIC3). The function of these domains, however, is not yet known.

In addition to these evolutionary conserved domains, each murine ZIC protein contains several low complexity regions (including poly-alanine, -histidine, -proline, and -serine/glycine tracts), with the number and type of sequence varying between the five ZIC proteins (Figure 1.6). Low complexity regions can undergo expansion and contraction mutations that alter protein function, and poly-alanine tract expansions in *ZIC2* and *ZIC3* are associated with human disease pathogenesis. Expansion of the alanine tract in *ZIC2* is implicated in Holoprosencephaly, while expansion mutations of this domain in *ZIC3* are associated with Heterotaxy (Brown et al. 2001; Frank et al. 2015). Cell-based overexpression assays have shown that expansion of the ZIC2 alanine tract from 15As to 25As results in a near-complete loss of transactivation ability, dependent on the promoter that is used (Brown et al. 2005), suggesting a role for the alanine tract in the modulation of ZIC transactivation. It is unclear, however, if poly-alanine tract expansions affect the function of other ZIC family members. Previous analysis of ZIC3 found no significant change in transactivation ability upon modification of the alanine tract (Cowan et al. 2014), a result which may be attributed to the differences in promoters and experimental conditions (including cell lines, reporters and time at which luciferase activity was measured) in the two studies (Brown et al. 2005; Cowan et al. 2014).

1.5.4.2 ZIC proteins function as co-factors

Although C_2H_2 type zinc fingers were originally identified as DNA binding domains, it is now clear that these domains are also involved in protein-protein interactions (Brayer and Segal 2008). The ZICs were amongst the first proteins to show that C_2H_2 ZFDs could bind proteins in addition to DNA (Koyabu et al. 2001b) demonstrating the bi-functional properties of ZFDs. In addition to its role as a classical DNA binding transcriptional activator, ZIC proteins have also been implicated as a protein binding co-factor in the Shh, Nodal and Wnt signalling pathways during embryonic development. In all cases, the ZIC proteins have been shown to physically interact with the key transcriptional mediators of each signalling pathway (i.e. GLI mediators of HH signals, SMAD mediators of Nodal signals and TCF/LEF mediators of canonical Wnt signals).

ZIC in the Shh pathway

ZIC3 can directly bind the ZFD of the Shh-mediator protein GLI3 in glutathione-S-transferase (GST) pull-down assays using overexpressed protein in mammalian cell lines (Koyabu et al. 2001b). The ZIC proteins can bind GLI DNA binding sites with lower affinity than GLI proteins themselves, however, they do not stimulate transcription at these GLI binding sites (Mizugishi et al. 2001). However, ZIC3 synergises with GLI3 to stimulate transcription at GLI DNA binding sites at higher levels than GLI protein would alone in cell-based transactivation assays (Mizugishi et al. 2001; Koyabu et al. 2001b; Quinn et al. 2012). Additionally, whole mount gene expression analysis in wildtype murine embryos has shown that *Zic3* expression overlaps with that of *Gli3* in the somites, forebrain, midbrain, hindbrain and limb buds at 10.5-11.5 dpc. Moreover, the polydactyly phenotype exhibited by $Gli3^{+/-}$ heterozygotes is rescued when crossed with a *Zic3* null allele ($Zic3^{-/-}$) (Quinn et al. 2012). Although this implies a functional relationship between ZIC3 and Gli3 and suggests a role for ZIC3 in limb development and regulation of Shh signalling, conditional null alleles for *Zic3* generated by a variety of Cre strains and ZIC3 loss-of-function models demonstrate a requirement for ZIC3 in the nascent mesoderm early on during gastrulation to prevent L-R axis defects (Ahmed et al. 2013; Sutherland et al. 2013). This loss-of-function phenotype occurs prior to the time of Shh directed limb patterning and before Hh ligands are expressed in the embryo. Ultimately, this suggests that the primary function of ZIC3 in Heterotaxy occurs upstream of Shh signalling at gastrulation.

ZIC in the Nodal pathway

Recent studies have shown that ZIC2 can physically bind to SMAD2 and SMAD3 proteins which are transcriptional mediators of Nodal signalling (Shen 2007; Houtmeyers et al. 2016). This interaction was demonstrated by co-immunoprecipitation following overexpression of proteins in HEK293T cells and the interaction was localised to the N-terminal domain of ZIC2. ZIC2 was also found to complex

with endogenously expressed phosphorylated SMAD2 protein (Houtmeyers et al. 2016). Furthermore, transactivation assays in HEK293T cells showed that ZIC2 can antagonise SMAD-dependent transactivation in the absence of any known ZIC binding sites, suggesting ZIC2 does not itself bind DNA, but acts in complex with other proteins to exert transcriptional control. Complementing this cell work, *in vivo* studies have shown ZIC2 intersects the Nodal pathway during murine gastrulation (Houtmeyers et al. 2016). For example, halving the dose of Nodal activity on the mutant *Zic2^{Ku/Ku}* background precipitates anterior truncation in ~80% of *Zic2^{Ku/Ku}* embryos, a known consequence of loss of Nodal signalling (whereas *Nodal^{+/-}* embryos exhibit no phenotype) (Houtmeyers et al. 2016). This apparent *Nodal* haploinsufficiency in the absence of ZIC2 function implies that ZIC2 potentiates the Nodal signal during murine gastrulation.

ZIC in the Wnt pathway

Two independent studies have demonstrated that ZIC proteins can inhibit β-catenin mediated Wnt signalling by interacting with the TCF family of transcriptional repressors. For example, overexpression of wildtype β-catenin in *Xenopus* embryos results in the formation of a secondary axis (indicative of an elevated Wnt signalling phenotype). This phenotype can be rescued if *Zic2* or *Zic3* RNA is also overexpressed (via injection into the embryos; Pourebrahim et al. 2011; Fujimi et al. 2012), demonstrating *Zic2* and *Zic3* can inhibit the β-catenin induced secondary axis in *Xenopus* embryos. Currently, there is no *in vivo* evidence of this interaction in the mouse. Fujimi et al. 2012 also demonstrated that all five *Xenopus* Zics (Zic1-5) can inhibit Wnt signalling in a gastrula based TOPflash assay at varying degrees. Similarly, ZIC2 has also been shown to inhibit β-catenin mediated Wnt signalling in an overexpression cell-based TOPflash assay (Pourebrahim et al. 2011).

Importantly, it has been demonstrated ZIC2 physically binds to TCF7L2 (previously TCF4) but does not itself bind to DNA to inhibit Wnt-dependent TCF/β-catenin mediated transcription (Pourebrahim et al. 2011). The precise domain within the ZIC proteins that interacts with the TCF proteins is unknown, although the authors localised this activity to the ZIC2 ZFD via a set of domain mapping experiments in transformed mammalian cell lines (HEK293T and COS7 cells) using overexpression of large ZIC2 deletion mutants. ZIC2 protein truncated before the ZFD was unable to interact with TCF7L2, thus the authors suggest that the ZFD is required for this interaction. However, the study did not perform subcellular localisation studies on the proteins and consequently failed to demonstrate whether the truncated forms of ZIC2 used localise to the nucleus (the site of ZIC/TCF interaction). Furthermore, due to the fact the ZIC proteins contain an interspersed nuclear localisation signal within the ZFD, the lack of ZIC2-TCF7L2 interaction observed by Pourebrahim et al. 2011 in mutants lacking a ZFD could simply reflect a lack of nuclear ZIC protein. Further domain mapping was performed with truncated forms of TCF7L2 protein co-immunoprecipitated with full-length ZIC2 to determine where ZIC2 binds

TCF7L2. It was found that TCF7L2 variants lacking the HMG domain and C-terminus were unable to co-immunoprecipitate ZIC2. An *in vitro* protein binding assay also confirmed this variant could not bind ZIC2 and the authors determined the HMG domain of TCF7L2 is necessary for the ZIC2-TCF7L2 interaction (Pourebrahim et al. 2011).

The Arkell laboratory has also demonstrated that all five human ZIC proteins (ZIC1-5) inhibit Wnt signalling in a cell-based TOPflash assay (Figure 1.7A, unpublished data). In this experiment, the addition of wildtype β-catenin in this system is representative of a low Wnt environment. In contrast, β-catenin-ΔN89 is used in the TOPflash assay to represent a high Wnt environment. β-catenin-ΔN89 contains an N-terminal deletion of the first 89 amino acids. The protein consequently lacks the N-terminal phosphorylation sites, which are required for the normal degradation of the protein. Therefore, this form of β-catenin is able to evade the degradation complex and be constitutively expressed in the cell nucleus. All five human ZIC proteins are able to inhibit Wnt signalling in a low Wnt environment (in the presence of wildtype β-catenin), however, this inhibition is lost in a high Wnt environment (in the presence of β-catenin-ΔN89) (Figure 1.7).
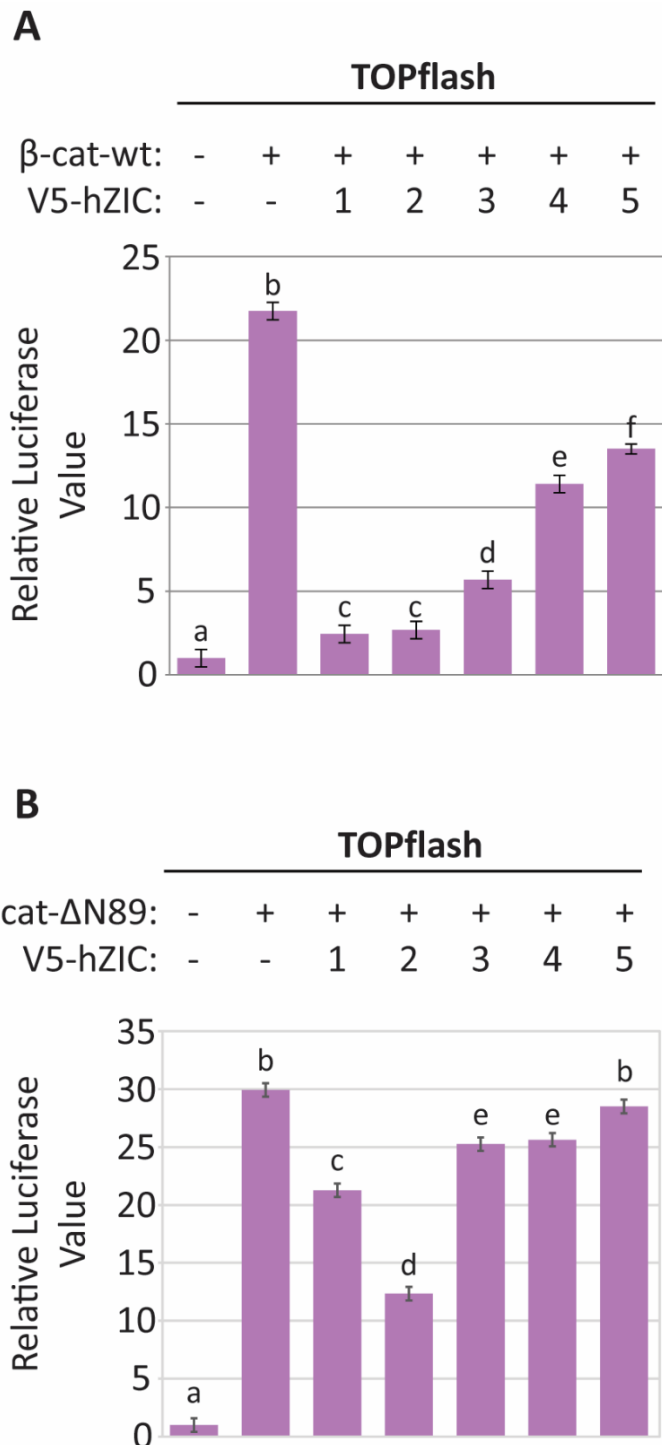
Figure 1.7: ZIC proteins can inhibit Wnt signalling in a low Wnt environment. HEK293T cells were transfected with a TOPflash reporter (TOP-β-gl-luc2), either **(A)** V5-tagged β-catenin-wt (to stimulate a low Wnt environment [called β-cat-wt]) or **(B)** β-catenin-N89 (to stimulate a high Wnt environment [called β-cat-N89]) and a V5-tagged human ZIC expression plasmid (hZIC1-5) shown. Luminescence was measured 24 hours post-transfection in each of three internal replicate samples. Values shown depict the mean relative luciferase values pooled from three independent repeats. Error bars represent SEM. a, b, c, d, e, f: *p<0.05* two-way ANOVA. Experiment performed by H. Bellchambers.

1.6 The Wnt signalling pathway

Wnt signalling is a key pathway involved in many critical developmental processes such as cell differentiation, proliferation and migration. Wnt activity must be tightly controlled during both embryo development and adult homeostasis, as evidenced by severe developmental defects associated with the pathway and by the many forms of cancer that are caused by dysregulated Wnt signalling (Haegel et al. 1995; Morkel et al. 2003; Zhan et al. 2017). The ligands that simulate Wnt signalling can be divided into two distinct classes: classical Wnts (Wnt1, -3a, -8, and -8b) generally activate the canonical Wnt pathway, whilst non-classical Wnts (Wnt4, -5a, and -11) typically activate the non-canonical pathways, which include the planar cell polarity (PCP) pathway and the Wnt/Calcium signalling pathway (Rao and Kühl 2010). Both the canonical and non-canonical pathways involve a Wnt ligand binding to a cell surface receptor, called Frizzled, which results in a cascade of events that stimulates several intracellular processes.

Non-canonical Wnt signalling is independent of the β-catenin protein. The PCP pathway is named as such because it is involved in establishing cell polarity and regulates cell movement during events that include gastrulation and neural tube closure (Komiya and Habas 2008). Wnt/Calcium signalling activates calcium-dependent pathways to mediate cell adhesion and controls genes responsible for cell migration (Croce and McClay 2008; Gó Mez-Orte et al. 2013). In contrast, the canonical Wnt pathway is dependent on the co-activator protein β-catenin to drive gene expression (Figure 1.8). In the absence of WNT ligand stimulation, β-catenin levels are depleted by its degradation in the cytoplasm by a destruction complex composed of adenomatous polyposis coli (APC), glycogen synthase kinase 3 (GSK3), and kinases casein kinase 1 (CK1), which prevent β-catenin entering the nucleus (Behrens et al. 1998; Liu et al. 2002). In this state, the T-cell Factor/Lymphoid Enhancer-binding Factor (TCF/LEF) transcription factors interact with co-repressor proteins, such as the Groucho/transducin like Enhancer of split (Gro/TLE) family, to inhibit the transcription of Wnt target genes (Brannon et al. 1999; Brantjes 2001). In a Wnt activated state, a WNT ligand binds to Frizzled and β-catenin is released from the destruction complex and accumulates in the cytoplasm. β-catenin can then translocate into the nucleus where it displaces the Groucho/TLE co-repressor complexes and physically interacts with the TCF/LEF transcription factors to activate expression of Wnt target genes (He et al. 1998; Tetsu and McCormick 1999; Rao and Kühl 2010).

It is well recognised that in *Drosophila* and *Xenopus* models, TCF proteins interact with the Gro/TLE co-repressors (Cavallo et al. 1998; Roose et al. 1998; Brantjes 2001). Using overexpressed protein in
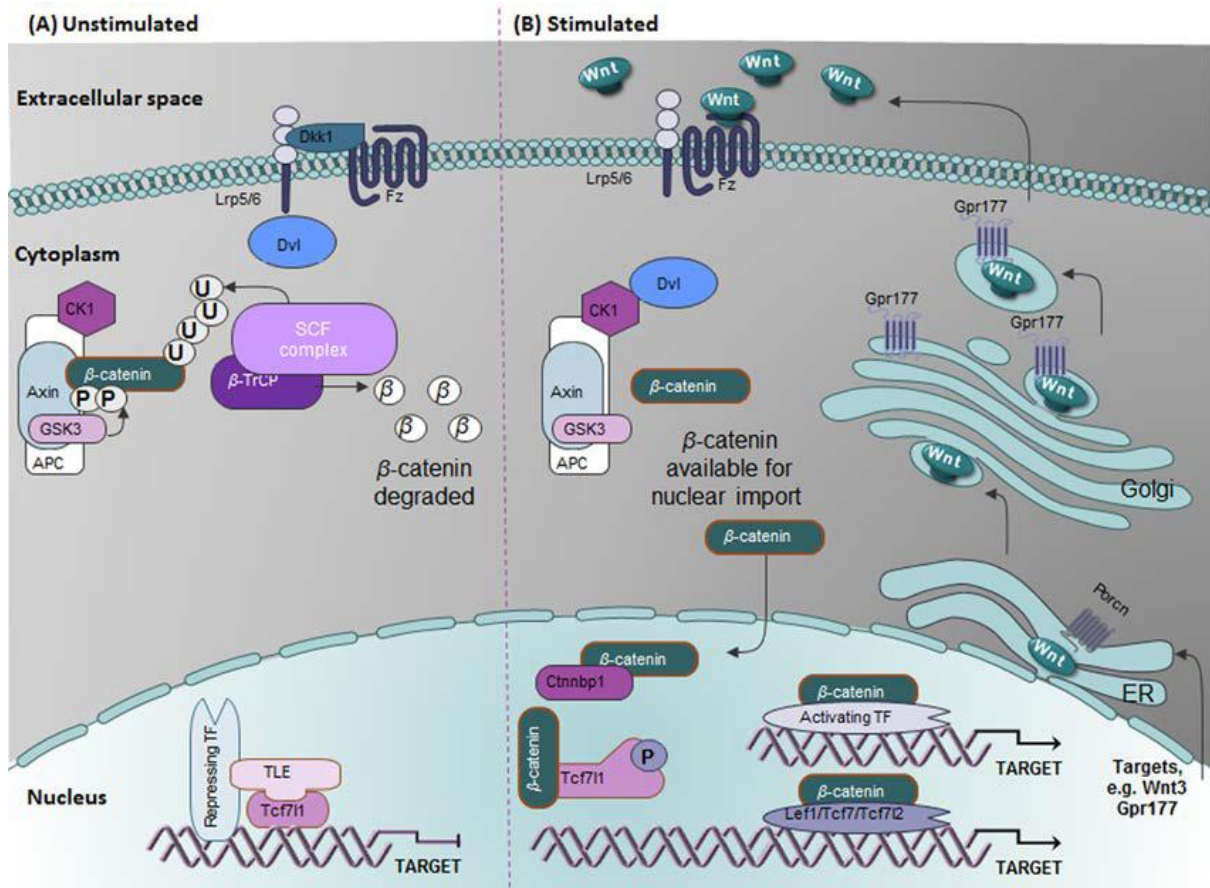
Figure 1.8: Diagram of the canonical Wnt signalling pathway. (A) In the absence of a Wnt ligand, cytoplasmic levels of β-catenin are restrained. β-catenin is phosphorylated (P) by the degradation complex (Axin, GSK3, APC, CK1) and ubiquitinated (U) by the SCF complex which targets β-catenin for ubiquitin dependent degradation via the ubiquitin ligase β-TrCP (Cadigan and Peifer 2009). In the nucleus, Wnt gene targets are repressed by TCF proteins, TLE co-repressors and repressing transcription factors (R-TF). **(B)** Upon binding of a Wnt ligand to a receptor complex (Fz/Lrp5/6) on the cell membrane, a host of intracellular signalling cascades are triggered, releasing β-catenin from the degradation complex resulting in an accumulation of cytoplasmic β-catenin. This prompts nuclear import of β-catenin, where it associates with TCF proteins to drive transcription of Wnt target genes. (Figure adapted from Arkell et al. 2013).

cell-based TOPflash assays, it has been demonstrated that all mammalian Groucho members can inhibit reporter activation when co-expressed with TCF proteins. This transcriptional repression is dependent on the interaction between histone deacetylases and the Gro/TLE proteins (Brantjes 2001). *In vivo* evidence of this interaction is not as well documented and there is no current mouse *Gro/TLE* allele that disrupts this TCF mediated repression to produce an elevated Wnt phenotype during embryo development. Consequently, it is unclear whether the Gro/TLE proteins act as TCF co-repressors during murine gastrulation. More recently, murine studies have demonstrated the HESX1 transcription factor interacts with TLE proteins (Carvalho et al. 2010) and results in elevated Wnt phenotypes including anterior truncation when mutated. Furthermore, the *Hex1* phenotype is exacerbated when *Tcf7l1* is also mutated, suggesting a genetic interaction between the two proteins to inhibit Wnt signalling.

A key regulator in the canonical Wnt pathway is the β-catenin protein, encoded by the *Ctnnb1* gene. β-catenin is the crucial mediator in gene transactivation, interacting with the TCF/LEF proteins to drive expression of Wnt target genes. As the β-catenin protein does not contain a DNA binding domain, it instead interacts with other DNA bound proteins to regulate the expression of Wnt specific genes. The protein contains a centrally located and conserved domain consisting of 12 armadillo repeats, which are involved in binding a variety of protein partners, flanked by N-terminal and C-terminal domains that interact with several co-activators and are involved in chromatin remodelling (Xu and Kimelman 2007; Mosimann et al. 2009).

1.7 Canonical Wnt signalling must be tightly regulated at mouse gastrulation

Canonical Wnt signalling is one of the key pathways activated during early embryogenesis and is necessary during gastrulation in establishing the anterior-posterior axis (Huelsken et al. 2000; Hikasa and Sokol 2013). This pathway must be tightly controlled as dysregulation of Wnt signalling during embryonic development has been shown to cause a myriad of severe morphological defects. Wnt signalling is also important for the development of the primitive streak as loss-of-function mutations of Wnt activators block primitive streak formation (Huelsken et al. 2000; Kelly 2004). Gain-of-function alleles of Wnt activators and loss-of-function alleles of Wnt repressors lead to similar phenotypes due to the over-activation of Wnt signalling. The two most distinct phenotypes associated with elevated canonical Wnt signalling are anterior truncation and posterior axis duplication in mouse and other vertebrates (Zeng et al. 1997; Fossat et al. 2011). For example, ectopic activation of Wnt signalling by injection of β-catenin mRNA into *Xenopus* embryos or loss of Wnt repressors such as Axin leads to a duplicated axis phenotype (Funayama et al. 1995; Zeng et al. 1997). Murine studies have also shown that *Dkk1* null embryos, lacking the DKK1 Wnt repressor, causes loss of head structures and partial or

complete truncation of the forebrain and midbrain (Mukhopadhyay et al. 2001; Fossat et al. 2011). Similarly, *Tcf7l1*[-/-] mouse embryos demonstrate an elevated Wnt phenotype during gastrulation and present with many morphological defects including duplicated axis structures such as the node and notochord and all *Tcf7l1*[-/-] embryos have anterior truncations (Merrill et al. 2004). Collectively, these studies demonstrate the essential role for the correct regulation of the Wnt pathway during embryonic development.

Additionally, these mutant phenotypes indicate that the anterior of the embryo requires low levels of Wnt signalling during gastrulation and that elevation of Wnt signalling causes the posteriorisation of the anterior neuroectoderm and produces additional body axes. It is therefore crucial that Wnt levels are tightly regulated during development to prevent these phenotypes. More specifically, it is suggested that in all animal models investigated so far, activators of Wnt signalling and negative regulators are distributed at opposite ends of the embryo, resulting in Wnt signal distributed as a gradient along the anterior-posterior axis, with greater activity at the posterior side (Lu et al. 2011; Arkell et al. 2013).

1.8 The TCF/LEF family of transcription factors

Both humans and mice have four *TCF* genes: *TCF7*, *TCF7L1*, *TCF7L2* (formally known as *TCF1*, *TCF3* and *TCF4,* respectively) and *LEF1*. All TCF/LEF proteins contain a HMG box which recognises a specific DNA binding motif (known as a Wnt responsive element), and a highly conserved β-catenin binding domain in the N-terminus of the protein required for the transcriptional activation of target genes (Arce et al. 2006; Hrckulak et al. 2016). TCF7L1 and TCF7L2 have an additional C-terminal Binding Protein (CtBP) binding domain in their C-terminus, a short amino acid sequence that is absent in TCF7 and LEF proteins, which mediates the interaction with CtBP proteins. CtBPs are well known as transcriptional co-repressors and can inhibit transcription in a histone deacetylase-dependent or -independent manner (Chinnadurai 2002). Although CtBP has been shown to interact with TCF proteins (Brannon et al. 1999), it is not clear whether this interaction has any true functional significance *in vivo* as the current evidence is contradictory. The interaction between CtBP and TCF proteins has been shown to mediate repression of transcription in cell-based TOPflash assays as well as the repression of Wnt target genes in *Xenopus* embryos (Brannon et al. 1999; Valenta et al. 2003). Conversely, there are reports that demonstrate that CtBP-dependent repression acts independently of TCF proteins in cell assays, and instead functions by binding to APC and sequesters cytoplasmic β-catenin, resulting in less freely available β-catenin for transcription activation (Hamada and Bienz 2004). *Ctbp2* loss-of-function in mouse embryos is embryonic lethal and results in axial truncations, delayed neural development and reduced *Brachyury (T)* expression, consistent with decreased Wnt activity.

The TCF/LEF protein family have been shown to have overlapping expression patterns during embryo development which become more specialised in adult tissues. The exception is TCF7L1, which is the only family member that has been detected at the onset of gastrulation in mouse embryos (Merrill et al. 2004; Hrckulak et al. 2016). Despite the functional redundancy between the TCF/LEF protein family, it has become apparent that each protein possesses unique functions. For example, to date, loss-of-function of *Tcf7l1* in mice results in a gastrulation phenotype that is not seen in *Tcf7*, *Tcf7l2* or *Lef1* mutants (Table 1.3). As stated above, the null allele of *Tcf7l1* (*Tcf7l1$^{-/-}$*) results in duplicated axis and abnormal node development, a phenotype commonly associated with elevated Wnt signalling at gastrulation (Pöpperl et al. 1997; Zeng et al. 1997). Other defects include enlarged hearts, foregut defects and large blood vessels (Merrill et al. 2004). Additionally, loss of *Tcf7l1* in mice is embryonic lethal as *Tcf7l1$^{-/-}$* embryos do not survive past embryonic day 10.5. Overall, the primary role of TCF7L1 during gastrulation is to repress Wnt signalling (Merrill et al. 2004).

In vivo studies have shown that TCF7L1 acts independently of β-catenin interaction during gastrulation (Wu et al. 2012). Mouse embryos containing a *Tcf3ΔN* knock-in mutation that prevents the interaction between TCF7L1 and β-catenin proteins progress through gastrulation without defects and do not display the elevated Wnt phenotype associated with *Tcf7l1$^{-/-}$* embryos. Evidently, the role of TCF7L1 during gastrulation does not directly involve binding to β-catenin and this model instead suggests the primary role of TCF7L1, at gastrulation, is that of a repressor rather than an activator of Wnt signalling.

1.9 Zic3 loss-of-function phenocopies elevated Wnt alleles

Zic3 null embryos are classified into four categories based on the severity of their phenotype and ability to initiate gastrulation (Ware 2006). The most severe phenotype is designated Type I in which embryos fail to gastrulate and resemble the null phenotype of *Nodal* (Conlon et al. 1994; Ware et al. 2006b). In contrast, Type II mutant embryos initiate gastrulation but display a variable phenotype. These embryos have abnormal mesoderm and primitive streak formation. Type II mutants also have a duplicated axis phenotype. Type III embryos proceed through gastrulation without defects, however, show evidence of caudal truncation and abnormal somite development. Type IV embryos show no defects at the ages examined by the authors.

In addition to L-R axis defects, murine embryos null for *Zic3* (described by Sutherland et al. 2013) have abnormal node morphology, a phenotype seen in elevated Wnt mutants (Merrill et al. 2004). Moreover, recent studies have shown that a *Zic3* loss-of-function mutation caused by a premature stop codon also shows signs of anterior truncation, axis defects and node abnormalities (Alzahrani 2017; Arkell lab, unpublished), which are similarly hallmarks of an elevated Wnt phenotype. These mutant mouse models show that dysregulation of *Zic3* causes elevated Wnt signalling and ultimately

Table 1.3: TCF/LEF family loss-of-function phenotypes

Gene	Allele	Phenotype	Reference
Tcf7	Tcf7[tmCle], null	Blocked thymocyte development. Mice are viable and fertile	(Verbeek et al. 1995)
Tcf7l1	Tcf7l1[tm1.1Bjme], null	Duplicated mesoderm structures (node and notochord), embryonic lethal (Tcf7l1$^{-/-}$ embryos do not survive past embryonic day 10.5)	(Merrill et al. 2004)
Tcf7l2	Tcf7l2[tm1Cle], null	Neonatal lethality, decreased number of villi and reduced amount of epithelial cells in the small intestine. Heterozygous mice are viable and fertile	(Korinek et al. 1998)
Lef1	Lef1[tm1Rug], null	Postnatal lethality, impaired development of hair, teeth and mammary glands	(Van Genderen et al. 1994)

identify a role for ZIC3 in the repression of Wnt signalling during gastrulation. **Due to the similarities between *Zic3* loss-of-function mutants and *Tcf7l1* null embryos (as well as other Wnt inhibitor mutants), it is hypothesised that ZIC3 acts with TCF7L1 as a co-repressor of Wnt signalling during early embryonic development. It still remains to be shown, however, whether ZIC3 directly functions with TCF7L1 and whether altered Wnt signalling is responsible for human Heterotaxy phenotypes**.

1.10 Scope of the thesis

Mutation of ZIC3 is implicated in human Heterotaxy patients and loss-of-function alleles of *Zic3* phenocopy elevated Wnt alleles. This introduces a central hypothesis that questions whether ZIC3 interacts with TCF7L1 during gastrulation to repress canonical Wnt activity and prevent Heterotaxy. This thesis aims to investigate whether:

1. **ZIC3 can physically interact with TCF7L1 *in vitro***
2. **ZIC3 Heterotaxy associated variant proteins have altered Wnt pathway function**
3. **the co-factor and transcription factor functions of ZIC3 be separated**

Chapter 2: Materials and Methods

2.1 Plasmids

The following reporter and expression constructs were available within the Arkell laboratory. Constructs were generated by myself, unless otherwise stated.

2.1.1 General Reporter and Expression Constructs

- **p(ZM2)6-β-gl-luc2 reporter plasmid** (constructed by Jerry Ahmed; Arkell laboratory). Six tandem copies of a ZIC DNA binding motif (CCC AGC GGG G) inserted upstream of the human *β-globin* minimal promoter and the coding sequence (CDS) for the *Photinus pyralis* luciferase gene. The plasmid provides a read-out of ZIC transactivation ability (see Appendix A for plasmid map).

- **pTOP-β-gl-luc2 reporter plasmid** (constructed by Helen Bellchambers; Arkell laboratory). Three tandem copies of a TCF DNA binding motif (AAG ATC AAA GGG) inserted upstream of the human *β-globin* minimal promoter and the CDS for the *P. pyralis* luciferase gene. The plasmid provides a read-out of TCF dependent Wnt transactivation ability (see Appendix A for plasmid map).

- **pENTR™3C (pENTR-3C) vector** (commercial vector, Invitrogen™; Cat. No. 11817-012). Used to create 'Entry' clones for Gateway® Recombination Cloning Technology (Life Technologies). Contains a multiple cloning site (MCS) that surrounds the negative selection gene *ccd*B. The MCS is flanked by *att*L sites required for Gateway® LR recombination cloning. A kanamycin resistance gene is included for selection in *Escherichia coli (E. coli)* (see Appendix A for plasmid map).

- **V5-Dest (pcDNA3.1-nV5-DEST) Gateway Destination plasmid** (commercial vector; Invitrogen; Cat. No. 12290-012). Used for Gateway recombination cloning. Contains a human cytomegalovirus (CMV) promoter for expression in mammalian cells upstream of the CDS for the V5 epitope tag. Recombination based cloning sites (*attR*) are included to enable in-frame fusion of an inserted CDS with the V5 tag followed by the bovine growth hormone polyadenylation (bGH polyA) stop sequence. The V5 epitope tag will be fused to the N-terminus of the inserted CDS (see Appendix A for plasmid map).

- **V1-ORF, ORF-V1 Gateway Destination plasmids** (a gift from Darren Saunders, Garvan Institute of Medical Research, Australia). Destination vector used for Gateway Cloning. Contains the CMV promoter upstream of part of the Venus CDS, coding for the N-terminal 158 amino acids of the Venus fluorescent molecule (termed V1) tag followed by two *attR* sites surrounding a *ccdB* gene for LR recombination based cloning (see Section 2.2.5). Venus is a

modified version of the Yellow Fluorescent Protein (YFP) (which itself is a variant of the Green Fluorescent Protein (GFP) derived from the jellyfish *Aequorea Victoria*). The V1 tag itself is not fluorescent and will only produce a fluorescent signal when interacting with V2 (the C-terminal section of the Venus protein).V1 tag will be attached to the N-terminus (V1-ORF) or C-terminus (ORF-V1) of the expressed protein (see Appendix A for plasmid map).

- **V2-ORF Destination vector** (a gift from Darren Saunders, Garvan Institute of Medical Research, Australia) used for Gateway Cloning. Contains the CMV promoter upstream of part of the Venus CDS, coding for the C-terminal 81 amino acids of the Venus fluorescent molecule (termed V2) tag followed by two *attR* sites surrounding a *ccdB* gene for LR recombination based cloning (see Section 2.2.5; Appendix A for plasmid map).

- **pFLAG-CMV-TCF7L2:** Originally pFLAG-CMV-TCF4 (Idogawa et al., 2005). Mammalian expression construct containing human *TCF7L2* cDNA. A gift from Sabine Tejpar (Department of Oncology, Katholieke University Leuven, Belgium).

- **Myc-Tcf7l1 (pBM58-Tcf7l1):** pcDNA3 vector containing full length murine *Tcf7l1* cDNA tagged with a Myc epitope. A gift from Bradley Merrill (Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago; see Appendix A for plasmid map).

- **pENTR3C-hZIC3-wt:** Gateway entry vector (constructed by Jerry Ahmed; Arkell laboratory). An Entry clone containing the human *ZIC2* CDS flanked by *att* recombination sites to catalyse transfer of the h*ZIC2* CDS to any Gateway Destination vector.

- **pUC57-hTCF7L1** (commercial vector, Genscript): Contains the entire human *TCF7L1* CDS within the pUC57 vector backbone with *att* recombination sites on either side of the TCF7L1 CDS.

- **pUC57-mZic3** (commercial vector, Genscript): Contains the entire mouse Zic3 CDS within the pUC57 vector backbone with *attR* recombination sites on either side of the *Zic3* CDS.

- **pENTR3C-mTcf7l1:** The 1943 bp murine *Tcf7l1* CDS was amplified from Myc-Tcf7l1 (pBM58-Tcf7l1) plasmid DNA with oligonucleotides containing *Xho*I and *BamH*I restriction sites (Ark_1578 and Ark_1579, Table 2.8). Following a gel extraction clean-up, the *Tcf7l1* fragment and the Gateway Entry vector pENTR3C were digested with *Xho*I and *BamH*I (section 2.3.2), releasing the *ccdB* fragment from pENTR-3C. The *Tcf7l1* fragment was ligated into the digested pENTR30 via traditional T4 ligation methods (section 2.3.4). The final vector is an entry clone containing the mouse *Tcf7l1* CDS flanked by *attR* recombination sites to catalyse transfer of the *Tcf7l1* CDS to any Gateway destination vector (see Appendix A for plasmid map).

- **V2-Tcf7l1 Expression plasmid:** Contains a CMV promoter for expression in mammalian cells upstream of the C-terminal section of the CDS for Venus fluorescent protein in-frame with

wildtype *Tcf7l1* CDS for expression of a V2-Tcf7l1 fusion protein. (See Appendix A for plasmid maps).

- **pβ-catenin-V5-HIsA (V5-β-catenin) Expression plasmid:** A gift from Yoshitaki Sekido (Usami et al. 2003). Contains a CMV promoter for expression in mammalian cells upstream of the human *β-catenin* CDS in-frame with the CDS for the V5 epitope tag and 6 tandem Histidine residues followed by the bGH polyA stop sequence, for expression of V5-tagged wildtype β-catenin protein. (See Appendix A for plasmid map).

- **pΔN89-β-catenin Expression plasmid:** A gift from Sabine Tejpar; Department of Oncology, Katholieke University Leuven, Belgium); originally described in (Munemitsu et al. 1996). Contains a CMV promoter for expression in mammalian cells upstream of the Myc epitope tag CDS in-frame with a mutated human *β-catenin* CDS (N terminal 89 residues deleted) for expression of Myc-tagged ΔN89 β-catenin protein. The deleted residues contain phosphorylation sites necessary for β-catenin ubiquitination and degradation, and thus the ΔN89 β-catenin protein is stabilised and has elevated transactivation ability relative to wildtype β-catenin. (See Appendix A for plasmid map).

- **pV1-hZIC3-wt Expression plasmid** (constructed by Helen Bellchambers): Contains a CMV promoter for expression in mammalian cells upstream of the N-terminal section of the CDS for Venus fluorescent protein in-frame with wildtype h*ZIC3* CDS for expression of a V1-ZIC3 fusion protein. (See Appendix A for plasmid maps).

- **pnV5-hZIC3-wt Expression plasmid** (constructed by Jerry Ahmed): Contains a CMV promoter for expression in mammalian cells upstream the V5 epitope tag CDS in-frame with the h*ZIC3* wildtype CDS followed by the bGH polyA stop sequence, for expression of V5-tagged wildtype ZIC3 protein (see Appendix A for plasmid map).

2.1.2 ZIC3 variant constructs

All ZIC3 variant plasmids were produced via site directed mutagenesis of the pENTR3C-ZIC3-wt plasmid (section 2.2.2). The following constructs were generated:

- **Premature termination codon (PTC) containing mutants: pENTR-ZIC3-C268X, pENTR-ZIC3-Q292X, pENTR-ZIC3-E325LfsX84, pENTR-ZIC3-K408X, pENTR-ZIC3-E155X, pENTR-ZIC3-Y199delfsX19, pENTR-ZIC3-Q249X, pENTR-ZIC3-S252X and pENTR-ZIC3-E291GfsX53** all contain a nonsense mutation, except pENTR-ZIC3-E325LfsX84, pENTR-ZIC3-Y199delfsX19 and pENTR-ZIC3-E291GfsX53, which contain a frame-shift mutation. Construct names show the specific amino acid (C268, Q292, K408, E155, Q292, and S252) that is mutated to a stop codon (X), with numbers indicating amino acid number of the translated protein product. ZIC3 variants pENTR-ZIC3-C268X, pENTR-ZIC3-Q292X, pENTR-ZIC3-E325LfsX84, and pENTR-ZIC3-K408X were made by Jerry Ahmed

- **Human missense mutations: pENTR-ZIC3-C253S, pENTR-ZIC3-W255G, pENTR-ZIC3-H286R, pENTR-ZIC3-G17C, pENTR-ZIC3-S109C, pENTR-ZIC3-P217A, pENTR-ZIC3-H318N and pENTR-ZIC3-A447G.** Each mutant contained a missense mutation altering a single amino acid within ZIC3 Variants C253S, W255G and H286R affect a conserved residue in ZF1

- **Targeted Zinc finger mutations: pENTR-ZIC3-C268S, pENTR-ZIC3-C302S, pENTR-ZIC3-C335S, pENTR-ZIC3-C365S, and pENTR-ZIC3-C393S.** Each construct contained a missense mutation that converted the second cysteine (C) of ZF1 (C268), ZF2 (C302), ZF3 (C335), ZF4 (C365) or ZF5 (C393) to serine (S). Mutations were installed using site-directed mutagenesis (section 2.2.2). pENTR-ZIC3-C365S was constructed by Jerry Ahmed

- **Low complexity region deletion mutants: pENTR-ZIC3-Aladel, pENTR-ZIC3-Hisdel.** Constructs were missing either the Alanine repeat or the Histidine repeat of ZIC3

- **Conserved domain deletion mutants: pENTR-ZIC3-ZOCdel, pENTR-ZIC3-ZFNCdel, pENTR-ZIC3-N'del, pENTR-ZIC3-SANCdel, pENTR-ZIC3-SACCdel.** Mutants were missing either the ZOC domain (ZIC3-ZOCdel), the ZFNC domain (ZIC3-ZFNCdel), the whole N-terminus region preceding the ZFD (ZIC3-Ndel), the SANC domain (ZIC3-SANCdel) or the SACC domain (ZIC3-SACC del).

All ZIC3 'Entry' clones (pENTR-ZIC3-) were transferred to the V5-DEST expression vector via a Gateway® LR Clonase reaction (Life Technologies; section 2.2.5) to generate **V5-ZIC3-C268X**, **V5-ZIC3-Q292X**, **V5-ZIC3-E325LfsX84**, **V5-ZIC3-K408X**, **V5-ZIC3-E155X**, **V5-ZIC3-Y199delfsX19, V5-ZIC3-Q249X, V5-ZIC3-S252X, V5-ZIC3-E291GfsX53**, **V5-ZIC3-C253S**, **V5-ZIC3-W255G**, **V5-ZIC3-H286R**, **V5-ZIC3-G17C, V5-ZIC3-S109C, V5-ZIC3-P217A, V5-ZIC3-H318N, V5-ZIC3-A447G, V5-ZIC3-C268S, V5-ZIC3-**

C302S, **V5-ZIC3-C335S**, **V5-ZIC3-C365S**, **V5-ZIC3-C393S**, **V5-ZIC3-Aladel**, **V5-ZIC3-Hisdel**, **V5-ZIC3-SANCdel**, **V5-ZIC3-SACCdel**, **V5-ZIC3-ZOCdel**, **V5-ZIC3-ZFNC** and **V5-ZIC3-N'del**.

Similarly, all Entry' clones (pENTR-) were transferred to the V1-DEST expression vector via a Gateway® LR Clonase reaction (Life Technologies; section 2.2.5) to generate: **V1-ZIC3-C268X**, **V1-ZIC3-Q292X**, **V1-ZIC3-E325LfsX84**, **V1-ZIC3-K408X**, **V1-ZIC3-E155X**, **V1-ZIC3-Y199delfsX19**, **V1-ZIC3-Q249X**, **V1-ZIC3-S252X**, **V1-ZIC3-E291GfsX53**, **V1-ZIC3-C253S**, **V1-ZIC3-W255G**, **V1-ZIC3-H286R**, **V1-ZIC3-G17C**, **V1-ZIC3-S109C, V1-ZIC3-P217A, V1-ZIC3-H318N, V1-ZIC3-A447G, V1-ZIC3-C268S**, **V1-ZIC3-C302S**, **V1-ZIC3-C335S**, **V1-ZIC3-C365S**, **V1-ZIC3-C393S**, **V1-ZIC3-Aladel, V1-ZIC3-Hisdel, V1-ZIC3-SANCdel, V1-ZIC3-SACCdel, V1-ZIC3-ZOCdel**, **V1-ZIC3-ZFNC** and **V1-ZIC3-N'del**.

2.1.3 Yeast two-Hybrid vectors

- **pDEST-32:** Gateway Destination vector containing the GAL4 DNA binding domain (commercial vector; provided with ProQuest Two-Hybrid System, Invitrogen). This vector is used to clone a gene of interest in-frame with the GAL4 DNA binding Domain to be used in Yeast Two-Hybrid assays. The vector contains the *ADH1* promotor and *TRP1* CDS for selection on plates lacking tryptophan.

- **pDEST-22:** Gateway Destination vector containing the GAL4 Activation domain (commercial vector; provided with ProQuest Two-Hybrid System, Invitrogen). This vector is used to clone a gene of interest in-frame with the GAL4 Activation domain to be used in Yeast Two-Hybrid assays. The vector contains the *ADH1* promotor and *LEU2* CDS for selection on plates lacking leucine.

- **p32-ZIC3:** ProQuest Two-Hybrid expression construct containing human *ZIC3* CDS fused with the GAL4 DNA Binding Domain. Construct was generated via a Gateway LR recombination reaction using plasmids pENTR-ZIC3-wt and pDEST-32.

- **p32-Tcf7l1:** ProQuest Two-Hybrid expression construct containing *Tcf7l1* CDS fused with the GAL4 DNA Binding Domain. Construct was generated via a Gateway LR recombination reaction using plasmids pENTR-Tcf7l1-wt and pDEST-32.

- **p22-ZIC3-wt:** ProQuest Two-Hybrid expression construct containing human *ZIC3* CDS fused with the GAL4 Activation Domain. Construct was generated via a Gateway LR recombination reaction using plasmids pENTR-ZIC3-wt and pDEST-22.

- **p22-Tcf7l1:** ProQuest Two-Hybrid expression construct containing *Tcf7l1* CDS fused with the GAL4 Activation Domain. Construct was generated via a Gateway LR recombination reaction using plasmids pENTR-Tcf7l1-wt and pDEST-22.

- **pEXP32-Krev1:** ProQuest Two-Hybrid expression construct (commercial vector; provided with ProQuest Two-Hybrid System, Invitrogen) containing full length wildtype rat *Krev1* CDS fused with the GAL4 DNA Binding Domain. Used as a control vector (Bait plasmid) in conjunction with pEXP22-RALGDS to demonstrate a positive protein interaction.

- **pEXP22-RalGDS-wt:** ProQuest Two-Hybrid expression construct (commercial vector; provided with ProQuest Two-Hybrid System, Invitrogen) containing the wildtype Ras association domain of the RalGDS protein. Used as a control vector (Prey plasmid) in conjunction with pEXP32-Krev1 to demonstrate a strong positive protein interaction.

- **pEXP22-RalGDS-m1:** ProQuest Two-Hybrid expression construct (commercial vector; provided with ProQuest Two-Hybrid System, Invitrogen) containing a mutant form of the Ras association domain of the RalGDS protein. Used as a control vector (Prey plasmid) in conjunction with pEXP32-Krev1 to demonstrate a weak positive protein interaction.

- **pEXP22-RalGDS-m2:** ProQuest Two-Hybrid expression construct (commercial vector; provided with ProQuest Two-Hybrid System, Invitrogen) containing a mutant form of the Ras association domain of the RalGDS protein. Used as a control vector (Prey plasmid) in conjunction with pEXP32-Krev1 to demonstrate a negative protein interaction.

2.2 Molecular Cloning

2.2.1 Buffer and Solutions

- Luria Broth (LB) media: 25 g/L Difco™ LB broth, Miller; Bacto.

- LB Agar: 40 g/L Difco™ LB agar, Miller; Bacto.

- Lysis Solution: 50 mM tris(hydroxymethyl)aminomethane (Tris)-HCl (Sigma-Aldrich) pH 8.5, 1 mM EDTA (ethylenediaminetetraacetic acid; Ajax Chemicals), 5% Tween-20 (Sigma-Aldrich) and 0.4 mg/mL of Proteinase K (Roche).

- Loading Dye 10x: 20% glycerol (Merck), 19.2% 0.5 M Na_2EDTA (Sigma-Aldrich) and 0.001% bromophenol blue (Sigma-Aldrich).

2.2.2 Site directed mutagenesis

Mutations in the *ZIC3* CDS (C253S, W255G, H286R, E155X, C268S, C302S, C335S, C365S, C393S C268X, Q292X, E325LfsX84 and K408X) were created using the QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies; Cat. No. 210518) by following manufacturer's instructions. Mutagenesis PCRs (polymerase chain reaction) were performed in a total volume of 50 µL using 100 ng of pENTR-ZIC3-wt and 0.25 µM of each primer (Table 2.1, 2.2). Cycling parameters are shown in Table 2.3. Upon PCR completion, samples were treated with 2 µL of *Dpn*I enzyme (NEB; Table 2.4) by incubation at 37°C for 2 h. The reaction was precipitated (yeast precipitation, section 2.3.5.2) and electro-transformed into DH5α electrocompetent cells, or transformed using L10-Gold® (Agilent) ultracompetent bacteria cells (supplied with kit) via heat-shock according to manufacturer's instructions.

All other mutations within the *ZIC3* CDS were created using site-directed mutagenesis with a proof-reading *Taq* polymerase. For deletion mutations, primers were constructed to amplify the entire plasmid sequence (Table 2.1, 2.2), except the region to be deleted (Pérez-Pinera et al. 2006). The forward primer binds at the C-terminal side of the target region, while the reverse primer binds at the N-terminal side (Figure 2.1). The primers bind the template via their 3'-ends, while their 5'-ends contain complimentary overhangs. Ultimately the 5'-end of each primer contains a region of homology to each other (at least 15-16 bp) that allows annealing into a double-stranded DNA fragment.

Mutagenesis PCRs were performed with the *PfuUltra* II Hotstart PCR Master Mix (Agilent Technologies; Cat. No. 600850) in a total volume of 50 µL using 10 ng of pENTR-ZIC3-wt, 0.2 µM of each primer and the cycling conditions shown in Table 2.5. The reaction was then precipitated (section 2.3.5.2) followed by digestion with *Dpn*I (NEB; Cat. No. R0176) for 2 h. DNA was then precipitated and electro-transformed into *E. coli* (section 2.2.3). The resulting bacterial colonies were screened for the correct plasmid via colony PCR (section 2.2.4) and/or DNA sequencing (section 2.3.6).

Table 2.1: Primers for site directed mutagenesis of pENT-ZIC3-wt to create human *ZIC3* Heterotaxy associated variants. F: forward, R: reverse

Mutation	Primer #	Sequence (5'-3')
ZIC3-C268X	Ark1399_F	CGG CCC AAG AAG AGC TGA GAC CGG ACC TTC AGC
	Ark1400_R	GCT GAA GGT CCG GTC TCA GCT CTT CTT GGG CCG
ZIC3-Q292X	Ark1401_F	GTG GGG GGC CCG GAG TAG AAC AAC CAC GTC TGC
	Ark1402_R	GCA GAC GTG GTT GTT CTA CTC CGG GCC CCC CAC
ZIC3-E325LfsX84	Ark1403_F	CCG AGT GCA CAC GGG CTT GAG AAG CCC TTC CCA
	Ark1404_R	TGG GAA GGG CTT CTC AAG CCC GTG TGC ACT CGG
ZIC3-K408X	Ark1405_F	CTG CGC AAA CAC ATG TAG GTT CAT GAA TCT CAA
	Ark1406_R	TTG AGA TTC ATG AAC CTA CAT GTG TTT GCG CAG
ZIC3-Y199delfsX19	Ark1806_F	GCTGACCCAGCCCGCGCACGGACCC
	Ark1807_R	CGGGCTGGGTCAGCACGGCCGAAC
ZIC3-Q249X	Ark1810_F	GCCAATCAAGTAGGAGCTGTCGT
	Ark1811_R	ACGACAGCTCCTACTTGATTGGC
ZIC3-S252X	Ark1812_F	CAGGAGCTGTAGTGCAAGTGGATC
	Ark1813_R	GATCCACTTGCACTACAGCTCCTG
ZIC3-E291GfsX53	Ark1808_F	TGGAGCATGTGGGGGGGCCCGGAGCA
	Ark1809_R	TGCTCCGGGCCCCCCCACATGCTCCA
ZIC3-E155X	Ark1655_F	GGC ATC CCC TAG CCC CCT AGC TAC TTG
	Ark1656_R	CAA GTA GCT AGG GGG CTA GGG GAT GCC
ZIC3-C253S	Ark1649_F	CAG GAG CTG TCG TCC AAG TGG ATC GAC
	Ark1650_R	GTC GAT CCA CTT GGA CGA CAG CTC CTG
ZIC3-W255G	Ark1651_F	CG TGC AAG GGG ATC GAC GAG GCT C
	Ark1652_R	G AGC CTC GTC GAT CCC CTT GCA CG
ZIC3-H286R	Ark1653_F	GTC ACC ATG GAG CGT GTG GGG GGC

	Ark1654_R	GCC CCC CAC ACG CTC CAT GGT GAC
ZIC3-G17C	Ark1794_F	CTGGGAGTGTGCAGCTTCGG
	Ark1795_R	CCGAAGCTGCACACTCCCAG
ZIC3-S109C	Ark1796_F	CGCTGCCTGTGCCGCCTTC
	Ark1797_R	GAAGGCGGCACAGGCAGCG
ZIC3-P217A	Ark1798_F	CGCTCAGTTTGCTAACTACAGCC
	Ark1799_R	GGCTGTAGTTAGCAAACTGAGCG
ZIC3-H318N	Ark1689_F	TAC AAA CTG GTC AAC AAC ATC CGA GTG C
	Ark1690_R	GCA CTC GGA TGT TGT TGA CCA GTT TGT A
ZIC3-A447G	Ark1804_F	CCCCTTCTGGAGTTCAAACTAGC
	Ark1805_R	GCTAGTTTGAACTCCAGAAGGGG

Table 2.2: Primers for site directed mutagenesis of pENT-ZIC3-wt to create *ZIC3* structural variants.

Mutation	Primer	Sequence (5'-3')
ZIC3-SANC-del	Ark1753_F	CCC CCT AGT GGC GCT GGG CAC CCG TC
	Ark1752_R	CCC AGC GCC ACT AGG GGG CTC GGG GAT GC
ZIC3-SACC-del	Ark1739_F	TCA GAC TCC GCT ATA GCT TCA GCA AAC AGT AAA GAT ACC
	Ark1738_R	AGC TAT AGC GGA GTC TGA CCC TTG AGA TTC ATG AAC
ZIC3-ZOC-del	Ark1644_F	CCT TCA ACT CAA GCT CCG GGC TCA GTG A
	Ark1643_R	CGG AGC TTG AGT TGA GGG CGG CAG AG
ZIC3-ZFNC-del	Ark1648_F	AC GGT CCT CTG TCG TGC AAG TGG ATC G
	Ark1647_R	A CGA CAG AGG ACC GTG GTG GGC CG
ZIC3-N'-del	Ark1675_F	CGA ATT CAT GTG CAA GTG GAT CGA CGA G
	Ark1676_R	CTT GCA CAT GAA TTC GGT ACC GGA TCC AG
ZIC3-Ala-del	Ark1692_F	CCC ACT TCA AGC TGA GCC CTG CC
	Ark1691_R	TCA GCT TGA AGT GGG TTG AGT CCC CGA A

ZIC3-His-del	Ark1694_F	GGG CAC CAG CCA GGT CCC CAG C
	Ark1693_R	CCT GGC TGG TGC CCA GGG CGT TGG CGT A
ZIC3-C268S	Ark1608_F	CCCAAGAAGAGCTCCGACCGGACCTTCAGC
	Ark1609_R	GCTGAAGGTCCGGTCGGAGCTCTTCTTGGG
ZIC3-C302S	Ark1610_F	GCTACTGGGAGGAGTCCCCCCGGGAG
	Ark1611_R	CTCCCGGGGGGACTCCTCCCAGTAGC
ZIC3-C335S	Ark1612_F	CCTTCCCGGGCTCCGGGAAGATCTTTGCC
	Ark1613_R	GGCAAAGATCTTCCCGGAGCCCGGGAAGG
ZIC3-393S	Ark1614_F	CCCTATATCTGCAAAGTGTCCGACAAGTCCTACACG
	Ark1615_R	CGTGTAGGACTTGTCGGACACTTTGCAGATATAGGG

Table 2.3: Thermal cycling parameters using the QuikChange Lightning Site-Directed Mutagenesis kit.

Program	Temperature	Time	Cycles
Initial denaturation	95°C	2 min	1
Denature	95°C	20 sec	18
Anneal	60°C	10 sec	
Extend	68°C	115 sec	
Polish	68°C	5 min	1

Table 2.4: Reaction conditions for restriction enzyme digests. All restriction enzymes sourced from New England Biolabs (NEB). *m: methylated*

Enzyme	Recognition Sequence (5' to 3') / = denotes cut site	Incubation Temperature (°C)	Incubation Time (h)	NEB Buffer
*Eco*RI	G/AATTC	37	2	NEB1,2 or CutSmart
*Bam*HI	G/GATCC	37	2	CutSmart
*Xho*I	C/TCGAG	37	2	CutSmart
*Dpn*I	GA^m/TC	37	2	NEB2 or CutSmart

Table 2.5: Thermal cycling parameters for site directed mutagenesis using PfuUltra II Hotstart polymerase.

Program	Temperature	Time	Cycles
Initial denaturation	95°C	2 min	1
Denature	95°C	20 sec	30
Anneal	60°C	20 sec	
Extend	72°C	57 sec	
Polish	72°C	3 min	1
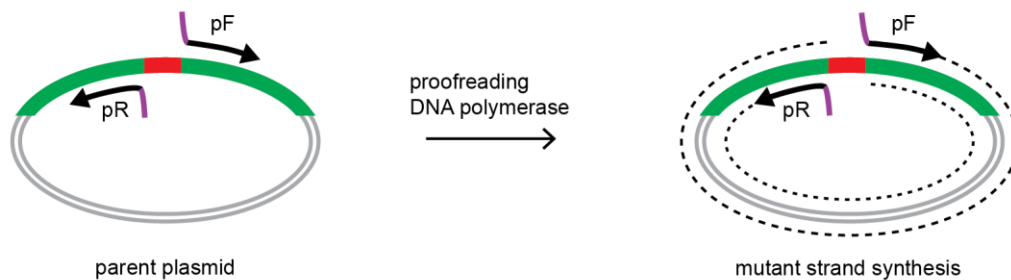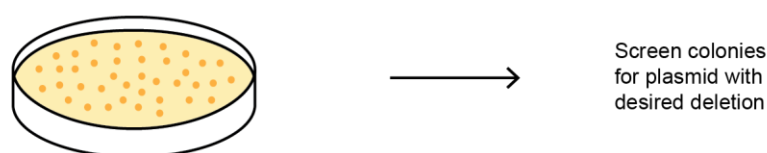
Figure 2.1: Creation of deletion mutants by overlap extension PCR. 1: Primers (pF: forward primer; pR: reverse primer) are designed to have their 5'-ends directed towards each other and 3'-ends oriented such that the DNA polymerase extends the entire plasmid, except the region to be deleted. The PCR is performed with a proofreading DNA polymerase. **2:** PCR products are precipitated (to isolate DNA from the reaction mixture) and **3:** treated with *Dpn*I to digest the methylated template plasmid (Lacks and Greenberg 1975). **4:** Following another round of DNA precipitation, bacteria are electro-transformed with the newly synthesised mutant plasmid. Screening of colonies can be done via colony PCR and/or DNA sequencing.

2.2.3 Bacterial transformation and culture

Bacterial transformation

For electroporation, ~20 ng of plasmid DNA, or 1 µL of a precipitated ligation reaction or LR reaction, were added to ice-cold *E. coli* cells (stored at -80°C). The cells and DNA were transferred to a chilled 0.2 cm electrocuvette (BioRad) and electroporated in an *E.coli* Gene Pulser (BioRad) set to deliver an electric field of 12.5 kVcm^{-1} (2.5 Kv, 25 µF and 200 Ω) to allow entry of the plasmid DNA. Immediately after pulsing, 500 µL of LB media (Bacto; Cat. No. 244620) was added to the electrocuvette to revive the bacteria and the solution was transferred to a 1.5 mL Eppendorf tube and incubated for 1 h on a rotating platform at 37°C. The bacterial solution was then plated on to LB Agar (Bacto; Cat. No. 244520) containing 0.1 mg/mL of the appropriate antibiotic (ampicillin: Sigma-Aldrich; Cat. No. A9518-5G, or kanamycin: Sigma-Aldrich; Cat. No. 60615-5G) and incubated for 16 h at 37°C.

2.2.4 Colony PCR

To enable direct analysis of the resulting bacterial colonies without isolation of plasmid DNA, colony PCRs were performed. Single colonies were hand-picked from the agar plate using a sterile pipette tip and the tip placed into 20 µL of AnalaR H$_2$O for 5 min to allow resuspension of the colony. Subsequently, 5 µL of the bacterial/H$_2$O solution was used in a standard 20 µL PCR using the primers listed in Table 2.6. The remaining bacterial suspension was stored at 4°C until analysis was complete. Colonies positive for the required DNA were used to inoculate 5 mL of LB broth containing 100 µg/mL of either ampicillin or kanamycin in a vented 15 mL tube. The tubes were incubated in a shaking 37°C incubator for 16 h.

Table 2.6: Primers used for screening of bacterial colonies containing the desired plasmids.

Plasmid	Primer #	Sequence (5'-3')	T$_A$ (°C)
pENTR3C-ZIC3-Aladel	Ark1152_F	ATCCGGTACCGAATTCACCCTCTCTCACTTCGG	60
	Ark1009_R	ATGCAGGCTGCTCGCCGAGC	
pENTR3C-ZIC3-Hisdel	Ark1225_F	CTATCTTCAGGCCAGAGC	57
	Ark1009_R	ATGCAGGCTGCTCGCCGAGC	
pENTR3C-ZIC3-SANC	Ark1225_F	CTATCTTCAGGCCAGAGC	50
	Ark1019_R	TGACCAGTTTGTACTTCGCC	
pENTR3C-ZIC3-SACC	Ark1173_F	TGGTCAACCACATCCGAGTGC	55
	Ark1256_R	GACTGATAGTGACCTGTTCGTTG	
pENTR3C-ZIC3-ZOC	Ark1225_F	CTATCTTCAGGCCAGAGC	50
	Ark1019_R	TGACCAGTTTGTACTTCGCC	
pENTR3C-ZIC3-ZFNC	Ark1225_F	CTATCTTCAGGCCAGAGC	50
	Ark1019_R	TGACCAGTTTGTACTTCGCC	

2.2.5 Gateway Left-Right Recombination reaction

DNA sequences were transferred from Entry vectors (pENTR) into Destination expression vectors (V5-DEST, V1-ORF, V2-ORF) using a Gateway LR recombination (Invitrogen™) reaction (Figure 2.2). Gateway cloning utilises the *Lambda* phage *att*achment (*att*) site-specific recombination system to transfer DNA sequences between two vectors with high efficiency. This system is based on that used by bacteriophage *lambda* during integration/excision of phage DNA into a bacterial chromosome (Craig and Nash 1984). In bacteria, integration of DNA occurs through specific *att*P sites on the phage chromosome and *att*B sites on the bacterial chromosome and is mediated by intergrase (Int) and intergrase host factor (IHF) proteins, which bind to the specific *att* sites to cleave and join target DNA. After recombination, the integrated phage DNA is now flanked by hybrid *att*L and *att*R sites in the bacterial chromosome (*att*P x *att*B → *att*L + *att*R). The desired plasmid is selected by a combination of antibiotic resistance and the negative selection of the original Destination vector based on the presence of the lethal *ccdB* gene.

Each LR reaction (performed according to modified manufacturer's instructions) contained equal amounts (150 ng) of 'Entry' and 'Destination' vectors, in a total volume of 9 μL of Tris-EDTA (10 mM Tris-Cl, pH 8.0 1 mM EDTA) and half the recommended amount of LR Clonase II Mix (1 μL). The reaction products were precipitated in the presence of 5 μL of yeast RNA (1 μg/μL; Sigma-Aldrich; see section 2.3.5.2) before being electro-transformed into DH5α bacteria (section 2.2.3) and grown on ampicillin agar plates to select for the desired plasmid.
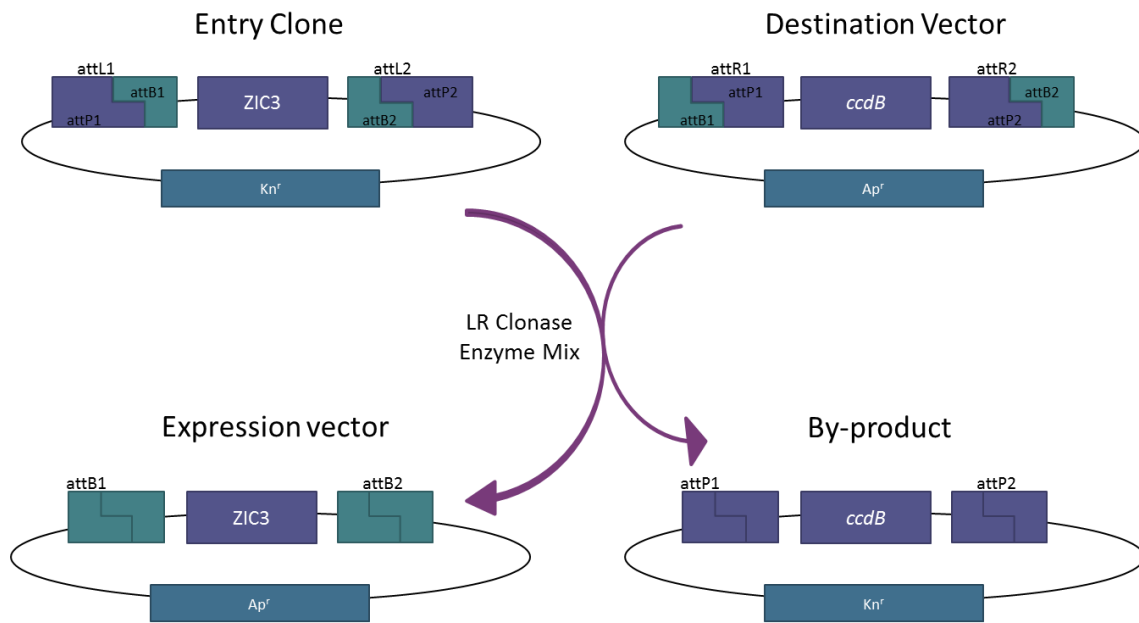
Figure 2.2: The Gateway Left-Right (LR) Recombination Reaction. Gateway cloning utilises the *Lambda* phage site-specific recombination system to transfer DNA sequences between two vectors with high efficiency. An 'Entry' Clone contains the gene of interest (*ZIC3*), flanked by *att*L1 and *att*L2 sites and the kanamycin resistance CDS (KnR). The 'Destination' vector contains the CDS for the cell lethal *ccdB* gene flanked by *att*R1 and *att*R2 sites, and the ampicillin resistance CDS (AmpR). The 'Entry' clone and Destination Vector undergo a site-specific recombination reaction mediated by the LR Clonase Enzyme Mix to form the desired Expression Clone containing the gene of interest and an unviable by-product. The reaction is transformed into DH5α bacteria (in which the *ccdB* gene product is lethal) and grown on ampicillin+ media to select bacteria containing the desired expression clone.

Small scale and large scale plasmid isolation (Mini and Midi prep)

Depending on the amount of DNA required, plasmid DNA was isolated using two methods. To isolate small quantities of plasmid DNA (<5 μg), single colonies were handpicked from LB-agar plates using a sterile tip and placed in a 14 mL bacterial culture tube containing 3-5 mL LB broth with an appropriate antibiotic (0.1 μg/mL) and incubated overnight (16 h; 37°C with shaking at 160 rpm). The plasmid DNA was then isolated using the High Pure Plasmid Isolation Kit (Scientifix; Cat. No. 740588.25) following the manufacturer's instructions. The resulting plasmid DNA was suspended in 100 μL of AnalaR H_2O. For larger quantities of DNA (>5 μg), a bacterial starter culture was made by inoculating 3 mL of LB broth containing the appropriate antibiotic (0.1 μg/mL) with a single bacterial colony. The starter culture was incubated for 5-8 h (37°C with shaking at 160 rpm) and 100 μL of this culture was then added to a further 100 mL of LB Broth and incubated overnight (16 h; 37°C with shaking at 160 rpm). The bacterial cells were pelleted by centrifugation at 3,200 x g for 20 min at 4°C and DNA was isolated using the NucleoBond Xtra Midi Plasmid DNA purification kit (Scientifix; 740410.5) following the manufacturer's instructions. The resulting plasmid DNA was suspended in an appropriate amount of AnalaR H_2O. The concentration and purity of the plasmid DNA was determined using a spectrophotometer (NanoDrop 1000 v8.1) and the DNA stored at -20°C until required.

2.3 General Molecular Biology

2.3.1 Polymerase Chain Reaction (PCR)

PCR was carried out using a Mastercycler® Vapo.protect™ or Mastercycler® Epgradient S™ (Eppendorf). PCR primers were designed for sequencing, colony PCR and general PCR amplification (Table 2.6 and Table 2.8). To promote specificity, primers were designed such that they had a G/C clamp at the 3' end where three of the five 3' most bases were G or C. The melting temperature (T_m) of the primers was designed to be within 45–65°C and primer pairs were designed to have a T_m within 5°C of each other. PCR products were amplified with 10-20 ng of plasmid DNA with each of the two of the primers at a final concentration 0.5 μM. Reactions were made in a total volume of 20 μL using a commercial PCR buffer at 1X (ImmoMix [Bioline], ReddyMix [Bioline] or ReddyMixKCl [Thermo Scientific]; Table 2.7). Thermal cycling was performed using a standard PCR program in which, after an initial denaturation step (94°C, 5 min), the DNA is briefly denatured at 95°C (30 s per cycle), the primers annealed to the template DNA at a temperature empirically determined for each primer pair (30 s per cycle) and then the amplicon extended at 72°C (30 s per cycle). Before a final extension at 72°C for 7 min. Success of the PCR was analysed by gel electrophoresis.

Table 2.7: PCR mastermixes used for amplifying plasmid and genomic DNA.

Buffer	Components	Supplier
2X ImmoMix™	IMMOLASE™ DNA Polymerase, Stabiliser, 2 mM dNTPs, 32 mM $(NH_4)_2SO_4$, 134 mM Tris-HCl [pH 8.3], 0.02% Tween 20, 3 mM $MgCl_2$	Bioline
2X PCR ReddyMix™ Master Mix with KCl	ThermoPrime *Taq* polymerase, 75 mM Tris-HCl (pH 8.8), 50 mM KCl, 20 mM $(NH4)2SO4$, 1.5 mM $MgCl2$, 0.01% (v/v) Tween 20, 0.2 mM dNTPs, Inert dye	Thermo Scientific
2X PCR ReddyMix™ Master Mix without KCl	ThermoPrime *Taq* polymerase, 75 mM Tris-HCl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 0.01% (v/v) Tween 20, 0.2 mM dNTPs, Inert dye	Thermo Scientific
PfuUltra II Hotstart PCR Master MIX	*PfuUltra II* Fusion HS DNA polymerase, reaction buffer, magnesium and dNTPs*	Agilent Technologies

component concentrations not disclosed by the manufacturer

2.3.2 Restriction Enzyme Digestion

Plasmid DNA was digested with restriction enzymes for either diagnostic or cloning purposes. The amount of restriction enzyme per reaction was calculated according to the general principle that one unit of a restriction enzyme can digest to completion 1 µg of plasmid DNA in 1 h. A minimum 5 fold excess of restriction enzyme to DNA (5 U/µg of plasmid DNA) was used in each restriction enzyme digest. Additional reaction components (such as restriction enzyme buffer and BSA) were added as recommended by the manufacturer of the restriction enzyme (Table 2.4) and all reactions were conducted in a volume that ensured the restriction enzyme component did not exceed 10% of the total reaction volume. The reactions were incubated at the recommended temperature for 2-3 h and heat inactivated as appropriate for the enzyme used.

2.3.3 Agarose Gel Electrophoresis

To analyse DNA fragments, samples were separated by horizontal gel electrophoresis. Agarose gels were created using Ultrapure Agarose (Life Technologies) and 1X TBE buffer (0.1 M Tris-HCl, 0.09 M boric acid and 0.001 M EDTA; Amresco). Agarose gels contained RedSafe™ DNA gel stain (1:20,000 dilution; Intron Biotechnology) to allow DNA to be visualised. DNA that ranged between 10,000 bp and 500 bp was run on a 1% gel, whilst DNA between 500 bp and 100 bp was run on a 2% gel. Samples were prepared for electrophoresis by the addition of 10X loading dye and 500 ng of the 1 Kb$^+$ DNA ladder (Life Technologies), which was loaded onto all gels to enable sizing of the electrophoresed DNA. Agarose gels were electrophoresed at 6-8 V/cm^{-1} until the DNA was sufficiently resolved (generally 20-40 min). The Gel Doc XR system and the accompanying Quantity One software (Bio-Rad) were used to visualise DNA products on the gel using an ultraviolet (UV) light source, and to photograph the gel.

2.3.4 DNA Ligation

DNA fragments were inserted into a prepared vector using T4 DNA ligase (NEB). To increase the probability of the production of the desired ligation product, two reactions with different vector to insert molar ratios (1:2 and 1:5) were generated for each insert/vector combination, calculated with the following formula:

$$\frac{Insert\ length\ (bp)}{Vector\ length} \times 50\ ng\ vector\ = x\ ng\ of\ insert\ required\ for\ 1:1\ molar\ ratio$$

The insert and vector DNA were mixed in the presence of 200 U of T4 DNA ligase and 1 x T4 Ligase Buffer (NEB) in a total volume of 20 µL AnalaR H$_2$O. Ligation reactions were incubated at room temperature for 1 h or at 16°C overnight (16 h) to increase the number of transformants. The ligated DNA was purified by precipitation in the presence of 5 µL of yeast RNA (1 µg/µL; Sigma-Aldrich; Section 2.3.5). *E. coli* bacteria (DH5-α) were electro-transformed with the purified ligated DNA (1 µL) before growth on LB-agar plates.

2.3.5 Purification of DNA

2.3.5.1 Gel extraction and PCR clean up

DNA fragments were purified from agarose gels using the NucleoSpin® Gel and PCR Clean-up Kit (Scientifix). Upon PCR fragment or digested plasmid DNA separation via agarose electrophoresis the gel was placed under UV light to visualise the position of the DNA bands, and the appropriate band cut from the gel (using a clean scalpel blade) and placed into a clean 1.5 mL micro-centrifuge tube. The weight of the gel slice was determined and Buffer NTI was added (200 μL of buffer per 100 mg of agarose gel). The DNA was purified out of the agarose according to manufacturer's instructions. DNA was eluted in 25 μL AnalaR H_2O and the concentration and purity of the DNA was determined using a spectrophotometer (NanoDrop 1000 v8.1) and stored at -20°C until required.

2.3.5.2 Ethanol precipitation

Plasmid DNA and/or inserts prepared for ligation (via PCR and/or restriction enzyme digest) were purified via ethanol precipitation. To the reaction tube, 0.25X of the reaction volume of 10 M Ammonium Acetate (NH_4OAc) was added, followed by 10X the NH_4OAc volume of 100% ethanol at -20°C. Tubes were briefly vortexed and centrifuged at 16,000 x g for 20 min to pellet the DNA. Supernatant was removed and 180 μL of 70% ethanol at -20°C was added followed by centrifugation at 16,000 x g for 2 min. Supernatant was removed and tubes left to air-dry before 20 μL of Analar H_2O was added to resuspend the precipitated DNA. To purify ligated plasmids from ligation reactions, ethanol precipitation was used with the following changes: in a 20 μL ligation reaction, 5 μL of yeast tRNA (1 μg/μL), 12.5 μL of 7.5 M NH_4OAc and 70 μL of 100% ethanol were added at -20°C. All other steps were followed except DNA was dissolved in 5-7 μL of Analar H_2O.

2.3.6 DNA sequencing

Sanger sequencing of DNA fragments or plasmids was performed at the Biomolecular Research Facility (BRF) at the Australian National University (ANU). Each reaction (20 μL total volume) contained the DNA of interest (250 ng plasmid DNA or 20 ng of PCR product) and 3 pmol of the appropriate sequencing primer (Table 2.8), 1 μL Big Dye (BRF; ANU) in Sequencing Buffer (BRF; ANU) to give final reaction conditions of 70 mM Tris-HCL (pH 9.0 at 25°C), 1.75 mM $MgCl_2$ and 1 μL of Big-Dye Terminating mixture (BDT). The reactions were subjected to cycle sequencing using program outlined in Table 2.9 in a Mastercycler PCR machine (Eppendorf).

The reactions were precipitated by the addition of 125 mM EDTA (pH 8.0) to a final concentration of 10 mM, followed by the addition of 3 M NaOAc (pH 5.2) to a final concentration of 0.12 M and 2.5 times the final reaction volume of 100% ethanol. The reactions were incubated for 20 min at room

Table 2.8: Sequencing primers.

Name (Number)	Sequence (5'->3') and Purpose
Ark_T7 Seq_F (929)	TAATACGACTCACTATAGGG Sequence across 5' vector/insert junction in pcDNA3.1 nV5 clones, V1-ZIC3 clones, V2-TCF7L1 clone and β-catenin-V5-HisA clone
Ark_bGH Seq_R (930)	TAGAAGGCACAGTCGAGG Sequence across 3' vector/insert junction in pcDNA3.1 nV5 clones, V1-ZIC2 clones, V2-TCF7L2 clone and β-catenin-V5-HisA clone
Ark_1152_F	ATCCGGTACCGAATTCACCCTCTCTCACTTCGG Internal to *hZIC3* CDS, use for sequencing *hZIC3* wt and variant clones
Ark_1225_F	CTATCTTCAGGCCAGAGC Internal to *hZIC3* CDS, use for sequencing *hZIC3* wt and variant clones
Ark_1172_R	CACCATCATCACCACACCAGC Internal to *hZIC3* CDS, use for sequencing hZIC3 wt and variant clones
Ark_1153_R	GTGCGGCCGCGAATTCCCGCTCTAGAACTAGTGG Internal to *hZIC3* CDS, use for sequencing *hZIC3* wt and variant clones
Ark_1173_F	TGGTCAACCACATCCGAGTGC Internal to *hZIC3* CDS, use for sequencing *hZIC3* wt and variant clones
Ark_1008_F	TCCAGCTGGCATCCCCGAGC Internal to *hZIC3* CDS, use for sequencing *hZIC3* wt and variant clones
Ark_1009_R	ATGCAGGCTGCTCGCCGAGC Internal to *hZIC3* CDS, use for sequencing *hZIC3* wt and variant clones
Ark_1010_F	GTGCACACGGGCGAGAAGCC Internal to *hZIC3* CDS, use for sequencing *hZIC3* wt and variant clones
Ark_1011_R	TGACCAGTTTGTACTTCGCC Internal to *hZIC3* CDS, use for sequencing *hZIC3* wt and variant clones
Ark_1015_F	GAA AGC CGC TTC TTG TAA TC Internal to β-catenin CDS, use for sequencing β-catenin wt and variants
Ark_1016_R	TGG CCT CAA GGG CTA GTT TCT Internal to β-catenin CDS, use for sequencing β-catenin wt and variants
Ark_1580_F	TTACCCACTGTCTCCAG Internal to Tcf7l1 CDS, use for sequencing mouse Tcf7l1
Ark_1412_R	GTGCAATGTAACATCAG Sequence across 3' vector/insert junction in pENT3C plasmids. Used for sequencing mouse pENTR-Tcf7l1
Ark_1578_F	AGGGATCCATGCCCCAGCTC To PCR amplify *Tcf7l1* fragment. Contains *Xho*I and *BamH*I restriction sites
Ark_1579_R	CCCTCGAGGCCTGAAGATCAGCTACTGG To PCR amplify *Tcf7l1* fragment. Contains *Xho*I and *BamH*I restriction sites
Ark_1411_F	AACTAAGCAGAAGGCCA Sequence across 3' vector/insert junction in pENT3C plasmids. Used for sequencing mouse pENTR-Tcf7l1
Ark_1581_R	GACAGTGGGTAATACGG

	Internal to *Tcf7l1* CDS, use for sequencing mouse *Tcf7l1*	
Ark_1583_R	TCTTGCTCTTAGAGGCC	
	Internal to *Tcf7l1* CDS, use for sequencing mouse *Tcf7l1*	
Ark_1582_F	AGAGCAAGAAGCCATGC	
	Internal to *Tcf7l1* CDS, use for sequencing mouse *Tcf7l1*	

Table 2.9: Thermal cycling parameters for DNA sequencing

Program Step	Temperature (°C)	Time	Cycles
Initial Denature	94	10 min	1
Denature	96	10 sec	x 30
Anneal	50	5 sec	
Extend	60	4 min	
Stop	4	Indefinite hold	1

temperature and centrifuged at 3,200 $x\,g$ for 30 min, 4°C. To remove the supernatant, the reaction plate was inverted onto paper towel and centrifuged upside down until 201 $x\,g$ to remove any remaining supernatant. Samples were briefly air-dried in the dark then wrapped in foil to protect the light sensitive dye. Samples were submitted for sequencing immediately or stored at -20°C until they could be submitted. Samples were analysed on an ABI 3730 sequencer (Applied Biosystems) within the BRF or an ABI 3130 Genetic Analyser in the Research School of Biology (RSB) Sequencing Facility within ANU. The resulting electronic sequence files were analysed using Geneious software version 5.5.9 (Biomatters).

2.4 Culture of mammalian cell lines

2.4.1 Cell culture

Human Embryonic Kidney 293T (HEK293T) cells, a highly transfectable line that are commonly used in ZIC *in vitro* assays, were used to study the effect of ZIC3 mutations. The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM; Sigma-Aldrich; Cat no D6429) supplemented with 10% (v/v) fetal bovine serum (Life Technologies or Interpath Services), 2 mM L-glutamine and 0.1 mM non-essential amino acid solution (Life Technologies; referred to as supplemented DMEM from herein). The cells were grown on sterile tissue culture plasticware (Corning) in a humidified chamber at 37°C with 5% CO_2. All procedures were performed in a laminar flow hood with sterile techniques.

2.4.2 Passaging Cells

Adherent cells grown in culture plasticware (T75 cm^2 or T25 cm^2) were generally passaged once they reached 70-80% confluency. The culture media was removed and cells washed with Phosphate Buffered Saline (PBS; Amresco) using ~0.05 mL/cm^2 culture surface. Cells were detached from the surface of the flask by incubation in a minimal volume of trypsin (0.5 g/L) in PBS for a maximum of 5 min in a humidified incubator (37°C with 5% CO_2). After incubation, supplemented DMEM was added to inactivate the trypsin and the solution repeatedly pipetted to create a homogeneous cell solution. An appropriate amount of the cell suspension, as determined by end use experiments, was transferred to new tissue culture plasticware containing supplemented DMEM. Additional supplemented media was added to the new tissue culture plasticware to reach the minimum volume requirements for each dish/flask.

2.4.3 Transfection of mammalian cell lines

Cells to be transfected were plated into 6-well or 12-well sterile tissue culture dishes (Corning; Table 2.10). If the cells were to be used for immunofluorescence or for a proximity ligation assay, two circular glass coverslips (13 mm, No. 1 thickness; ProSciTech) were added to each well of a 6-well dish before

Table 2.10: Transfection amounts based on culture plate sizes. Separate dilutions were made for Lipofectamine™ 2000 and plasmid DNA e.g. in a 12-well plate, 4 µL of Lipofectamine™ 2000 was diluted in 100 µL of unsupplemented DMEM, whilst 1.6 µg of plasmid DNA was diluted in another 100 µL unsupplemented DMEM.

Multi-well plate	Well diameter (Bottom mm)	Cell number (100% confluency)	Lipofectamine™ 2000 (µL/well)	Total plasmid DNA (µg)	Working volume
6-well plate	34.8	9.5×10^5	10	4.0	2-3 mL
12-well plate	22.1	3.8×10^5	4	1.6	0.75-1.5 mL
96-well plate	6.4	3.2×10^4	-	-	100-200 µL

the cells were plated. Cells were transfected when 60-80% confluent. All cells were transiently transfected with plasmid DNA using a lipid-based transfection agent (Lipofectamine™ 2000; Life Technologies). The Lipofectamine™ 2000 reagent was diluted 1:25 in unsupplemented DMEM and incubated at room temperature (~25°C) for 5 min. The appropriate plasmid DNA was diluted in a separate volume of unsupplemented DMEM equivalent to that used to dilute the Lipofectamine™ 2000. At the end of the 5 min incubation, the Lipofectamine™ 2000 solution was added to the DNA dilution and incubated at room temperature for 20 min to enable DNA-liposome complex formation. During this time, the media on the cells was replaced with fresh supplemented DMEM. The DNA-Lipofectamine™ 2000 mixture was added drop-wise to the well containing cells. The plate was then gently rocked to mix the solutions and returned to the 37°C humidified incubator.

2.4.4 Luciferase reporter assays

The experimental design of the luciferase reporter assays is shown in Figure 2.3. In all cases, HEK293T cells grown to 70% confluency in tissue culture grade 12-well plates were transfected with the appropriate combination of reporter and expression constructs. For the ZIC transactivation assays, cells were transfected with 0.8 µg of reporter plasmid p(ZM2)6-β-gl-luc2 and 0.8 µg of one expression vector (V5-hZIC3-wt or one of the V5-hZIC3-variant expression constructs) or a plasmid from which no protein would be generated (V5-DEST) as a negative control. For the TCF/β-catenin transactivation assays (TOPflash), cells were transfected with 0.5 µg of a reporter plasmid pTOP-β-gl-luc2, a β-catenin expression plasmid (pβ-Catenin-V5-HIsA or pCANΔN89-β-catenin) and a ZIC expression construct (V5-ZIC3-wt or one of the V5-ZIC3-variant expression constructs) or a plasmid from which no protein would be generated (V5-DEST; 0.5 µg) as a negative control. Total DNA concentration did not exceed 1.6 µg.

For all luciferase assays, each well of transfected cells was passaged 6-8 h after transfection and split into three wells of a solid white tissue culture grade 96-well plate (Costar®, Corning; Cat. No. CLS3917) to generate three internal repeats for each experiment. The position of each internal replicate sample was randomised for each experiment to avoid plate reader bias from the luminometer. The remaining cells were plated in one well of a 6-well plate for SDS-PAGE and western blot. The luciferase activity of each sample in the 96-well plate was determined 24 h post-transfection by incubation with a luciferase substrate (Promega ONE-Glo luciferase reagent, Promega, Cat. No. E6110). The culture media was replaced with 100 µL of a 1:1 (v/v) mix of the ONE-Glo luciferase reagent and DMEM, and cells incubated for 15 min at ambient conditions. The intensity of luminescence produced was measured by the GloMax®-96 Microplate Luminometer (Promega), or the TECAN Infinite M1000 Pro plate reader. Upon result analysis, the luciferase activity was normalised to V5-DEST to determine relative luciferase activity. The mean relative luciferase activity and standard deviation were
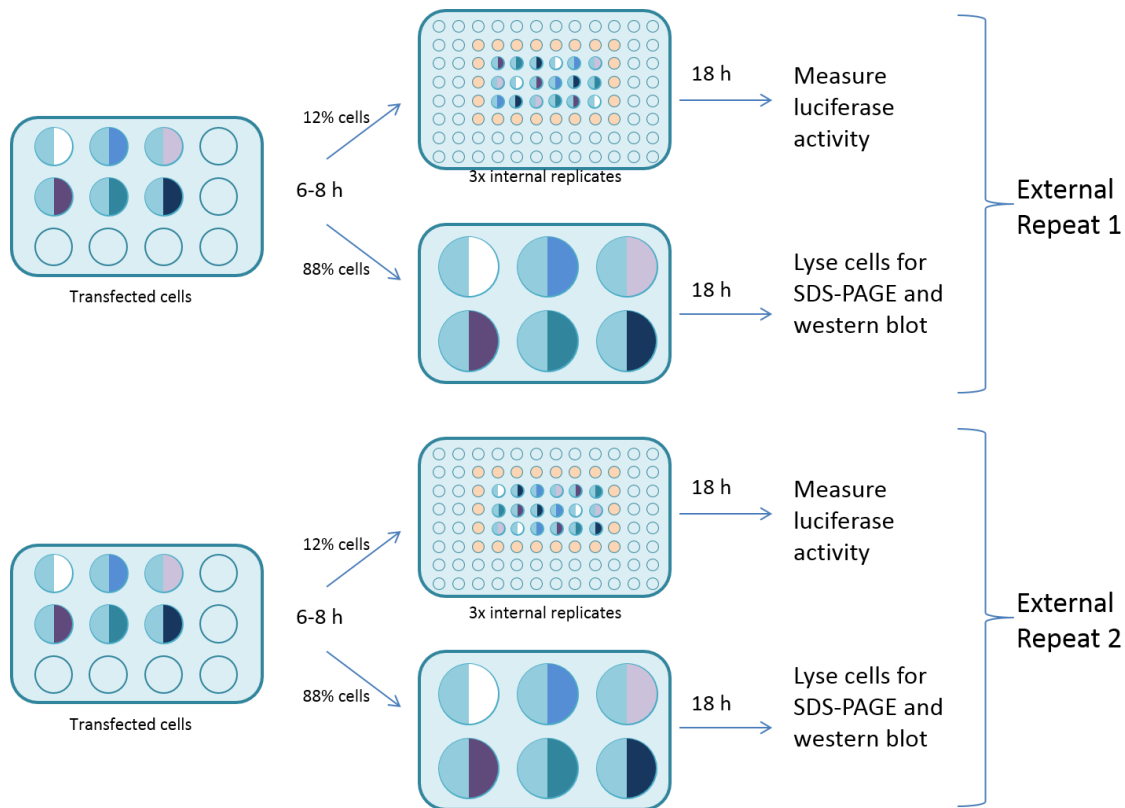
Figure 2.3: Experimental design of luciferase assays. HEK293T cells grown on a 12-well plate were transfected with a reporter construct that contains DNA binding motifs and a mammalian promoter upstream of the *P. pyralis* luciferase cDNA (left blue shaded semi-circle). They are also co-transfected with an expression vector (various colours). In some cases, the expression vector does not contain a cDNA (V5-Dest; white). This serves as a negative control to reveal the background level of luciferase and/or ensures equivalent amounts of DNA are included in each transfection. Approximately 5.5 h post-transfection, each transfection is split as shown. First, 4% of the transfected cells are transferred to each of three wells in a 96-well plate (to generate three internal replicates for the luciferase assay) and the remaining 88% of cells are transferred to one well of a 6-well plate (to be collected for lysis, SDS-PAGE and western blot analysis of the overexpressed protein). The cells for luciferase assay are distributed on the 96-well plate in an order generated by a randomised list of sample numbers to ensure that the plate position of each sample varies between external replicate experiments. Culture media without cells is added to adjacent wells surrounding the ones in use (orange). Luciferase substrate is added to each sample in the 96-well plate 18.5 h after the split (24 h post-transfection) and the plate read in a luminometer. After luciferase activity has been recorded, cells from the 6-well plate are lysed, the nuclear fraction recovered and subjected to SDS-PAGE and western blot analysis.

calculated from three internal repeats (using Microsoft Excel). Where pooled data is shown, data represents at least three independent repeats, with luminescence values from one representative experiment shown. The data was subjected to a two-way ANOVA using Genstat and a Fisher's least significant difference (LSD) test performed ($\alpha = 0.05$) as described in section 2.7.

2.4.5 Bimolecular Fluorescence Complementation

The experimental design of the Bimolecular Fluorescence Complementation (BiFC) assays were similar to the luciferase assays. HEK293T cells were grown to 70% confluency in tissue culture grade 12-well plates and transfected with 800 ng each of V1- and V2-tagged expression constructs. V1 and V2 are N-terminal and C-terminal fragments, respectively, of the F46L form of YFP, that exhibits an increased rate of maturation at 37°C relative to both YFP and GFP (Nagai et al. 2002). For the competition experiments the cells were co-transfected with 1 µg of pV2-Tcf7l1, 1 µg of pV1-hZIC3-wt and increasing amounts (0.5 – 2.0 µg) of pnV5-hZIC3-C268X (along with pnV5-DEST to equalise the amount of transfected DNA).

Each well of transfected cells was split 24 h post-transfection into three wells of a solid white tissue-culture grade 96-well plate (Costar®) to generate the three internal repeats. The position of each internal replicate sample was randomised for each experiment to avoid plate reader bias from the luminometer. The remaining cells were plated in one well of a 6-well plate for SDS-PAGE and western blot. 24 h after splitting (i.e. 48 h post-transfection), media was removed from each sample in the 96-well plate and replaced with 100 µL PBS. The fluorescent intensity from each well was then detected using a TECAN Infinite m1000 Pro plate reader. Venus fluorescence was measured using Excitation wavelength 515 nm and Emission wavelength 528 nm. 16 measurements were taken from each well to give an average fluorescent intensity. A minimum of three independent biological experiments were carried out. Fluorescent intensity data was subjected to a two-way ANOVA using Genstat and a Fisher's LSD test performed ($\alpha = 0.05$) as described in section 2.7.

2.4.6 Proximity Ligation Assay

Proximity ligation assays (PLA) were conducted according to the manufacturer's instructions using the Duolink® In Situ Red Starter Kit Mouse/Rabbit (DUO92101-1KT) which included Anti-Rabbit PLUS and Anti-Mouse MINUS PLA probes. HEK293T cells grown on coverslips in a 6-well plate were transfected with 1 µg of appropriate expression plasmids. 24 h post-transfection, cells were fixed, prepared for immunofluorescent staining and primary antibodies hybridised as per section 2.5.6. Coverslips were mounted cell side down onto glass slides (76 x 26 mm, No. 1 thickness, ProSciTech G300FB) with Duolink® In Situ Mounting Medium with DAPI (Sigma-Aldrich, DUO82040). Nail polish was used to seal

the edges of the coverslip to prevent the cells from drying out. Images were acquired using a Zeiss Axio Observer (1.210) fluorescent microscope.

2.5 Analysis of cultured cells

2.5.1 Nuclear and Cytoplasmic Protein Lysates

HEK293T cells to be processed for western blotting were processed with the NE-PER fractionation kit (ThermoScientific) according to the manufacturer's instructions, but with reagent volumes altered. Media was removed from the cultured HEK293T cells in a 6-well plate and a cell suspension made in 800 µL fresh unsupplemented DMEM. Cell were pelleted by centrifugation at 20,800 $x\ g$ and the media removed and replaced with 100 µL of Cytoplasmic Extraction Reagent (CERI) with 1X cOmplete Mini EDTA-free Protease Inhibitor Cocktail (Roche) added. Cells were vortexed briefly and incubated for 10 min (4°C). The CERII buffer was then added (5.5 µL per well in a 6-well plate, ice-cold) and the mixture vortexed and incubated on ice for an additional 1 min, and then vortexed prior to centrifugation for 15 min at 20,800 x g at 4°C. The supernatant (containing cytoplasmic proteins) was transferred to a pre-chilled micro-centrifuge tube. The pellet was then resuspended in ice-cold Nuclear Extraction Reagent (NER) (50 µL per well of a 6-well plate) with added 1X protease inhibitor to release nuclear proteins and incubated on ice. The cells were vortexed for 15 s every 10 min for a total of 40 min. The nuclear fraction was separated by centrifugation at 20,800 x g (25 min, 4°C). The supernatant containing the nuclear proteins was transferred to a pre-chilled micro-centrifuge tube. Lysates were then either stored at -80°C until required or immediately processed for Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE).

2.5.2 SDS-PAGE

Protein from cytoplasmic and nuclear cell lysates was separated by molecular weight (MW) on an acrylamide gel. Polyacrylamide gels were prepared in a Mini PROTEAN® casting apparatus (BioRad). An 8-10% polyacrylamide 'Running' gel (Table 2.11) was made and overlaid with hydrated butanol (Sigma-Aldrich) and left at room temperature to polymerise for 1 h. Following polymerisation, the butanol was removed and the running gel overlaid with a 3.75% polyacrylamide 'Stacking' gel (Table 2.12) and a comb inserted to create wells for loading samples. The gel was left at room temperature for a further 1 h to polymerise. Once polymerisation was complete, the gels were placed into a BIORAD Mini PROTEAN® Tetra Cell apparatus filled with 1X SDS-PAGE Running Buffer (192 mM glycine, 24.9 mM tris base and 3.47 mM SDS). To prepare lysates for SDS-PAGE, a 1:20 dilution of Dithiothreitol (DTT, Sigma-Aldrich; 20X stock solution) and 1:4 dilution of Sample buffer (NUPAGE® LDS Sample Buffer, a proprietary solution supplied as 4X, Invitrogen) was added and the lysate heated for 5 min

Table 2.11: Composition of 8-10% polyacrylamide 'running' gel used to separate proteins according to molecular weight. Volumes shown produce two gels of 1.5 mm thickness.

Composition	Volume (mL)	
	8%	10%
40% Acryl:Bisacryl (37.5:1; Sigma-Aldrich)	5.0	6.25
1.5M Tris (pH = 6.8; Merck)	6.25	6.25
10% SDS (Roche)	0.25	0.25
Deionised H_2O	13.25	12
10% Ammonium persulphate (Sigma-Aldrich)	0.25	0.25
Tetramethylethylenediamine (TEMED; Sigma-Aldrich)	0.025	0.025

Table 2.12: Composition of 3.75% polyacrylamide 'stacking' gel used for sample loading. Volumes shown produce two gels of 1.5 mm thickness.

Composition	Volume (mL)
40% Acryl:Bisacryl (37.5:1; Sigma-Aldrich)	0.47
0.5M Tris (pH = 6.8; Merck)	1.25
10% SDS (Roche)	0.05
Deionised H_2O	3.18
10% Ammonium persulphate (Sigma-Aldrich)	0.10
TEMED (Sigma-Aldrich)	0.01

at 90°C to denature the proteins. The lysates were cooled and 20-30 μL of each sample was loaded into the pre-made gels. A pre-stained molecular weight ladder (PageRuler™ Prestained Protein ladder; ThermoScientific) was loaded in the first and last well of the gel to enable sizing of the detected proteins. Samples were electrophoresed at 100 V for 80-120 min to separate the proteins, using the separation of the protein ladder as a guide for run time.

2.5.3 Wet transfer

After separation by SDS-PAGE, proteins were transferred to a polyvinylidene fluoride (PVDF) membrane (Millipore ImmobilinTM–P) via wet transfer. The membrane was first activated via soaking in 100% (v/v) methanol for 2 min prior to being assembled for wet transfer. Cold (4°C) 1X Towbins Transfer buffer (192 mM glycine, 24.8 mM Tris, 20% methanol, pH = 7.4) was poured into a transfer tray and thin transfer sponges soaked to eliminate any bubbles. The SDS-PAGE gel was prepared for wet transfer by removing it from the BIORAD Mini PROTEAN® Tetra Cell (BioRad) system and discarding the 'stacking' portion of the gel. The prepared SDS-PAGE gel was positioned over the PVDF membrane and sandwiched together between blotting paper (3 mm Chromatography paper, Whatman®) and the thin transfer sponges before being assembled in the wet transfer apparatus (BIORAD Trans-Blot SD-Semi-dry Transfer Cell apparatus; BioRad). The transfer apparatus was placed within an ice cavity to keep cool and the transfer was run at 15 V overnight (16 h).

2.5.4 Detection of proteins via western blot

Upon completion of the wet transfer, the PVDF membrane was rinsed with 1X PBS once and then blocked in skim milk blocking buffer (1X PBS with 5% (w/v) skim milk and 0.2% (v/v) Tween-20) for 1 h (room temperature with shaking). After blocking, the membrane was incubated with primary antibodies diluted in 3 mL blocking buffer (see Table 2.13 for dilutions) for 2 h (room temperature with rotation). The membrane was then washed with blocking buffer (six washes, 5-10 min each at room temperature with shaking) and then incubated with the appropriate Horseradish Peroxidase (HRP)-conjugated secondary antibodies diluted in 3 mL blocking buffer (See Table 2.14 for dilutions) for 1.5-2 h (room temperature with rotation). The membrane was then washed in 1X PBS with 0.2% Tween-20 (Sigma-Aldrich; six washes, 5-10 min each at room temperature with shaking). Lastly the membrane was rinsed with 1X PBS prior to developing. For western blot to detect the Venus protein (by hybridization with an anti-GFP antibody), the blocking buffer for both primary and secondary post-hybridization washes was 1x TBS (1L: 50 mM Tris-Cl, pH 7.6; 150 mM NaCl) with 0.2% Tween20 (Sigma-Aldrich).

Upon completion of antibody hybridisation, the membrane was treated with SuperSignal® West Pico Chemiluminescent Substrate (Thermo Scientific; ~37 µL per cm^2 of membrane) and incubated for 5 min in the dark. The SuperSignal West Pico Chemiluminescent Substrate contains a substrate upon which the HRP tag of the secondary antibody can act to produce a chemiluminescent signal. Following incubation, excess detection reagent was drained and the membrane wrapped in cling wrap. Under safe light conditions, the membrane was exposed to X-ray film (CL-Xposure™ Film; Thermo Scientific or Amersham Hyperfilm MP; GE Healthcare Ltd.) for various times (10 s to 16 h). The film was developed using a Multigrade Paper Developer solution (Ilford) and fixed using a Rapid Fixer solution (Ilford). The film was then rinsed with H_2O and air-dried. Once dry, the X-ray film was aligned with the membrane and the marker bands from the protein ladder traced onto the film to enable measurement of detected protein bands. Alternatively, following incubation with the Chemiluminescent Substrate, the membrane was immediately exposed to a charge-coupled camera via the ImageQuant™ LAS4000 biomolecular imager (GE Healthcare Life Sciences) for 10 s - 10 min (dependent on the experiment and antibody used) and a digital image produced.

Table 2.13: Primary Antibodies used for western blotting.

Antibody	Dilution
Mouse anti-TCF7L1 (Abcam; cat.no. ab86175)	1/1,000
Goat anti-β-catenin (Santa Cruz; sc-1496)	1/500
Mouse anti-V5 (Life Technologies; cat. no. R960-25)	1/3,000
Mouse anti- β-tubulin (Abcam; cat. no. ab7792)	1/1,000
Mouse anti-TATA Binding Protein (TBP; Abcam; cat. no. ab818)	1/2,000
Rabbit anti-V5 (Abcam; cat. no. ab9116)	1/7,500
Rabbit anti-GFP (Cell signaling technology; no.2555)	1/1,500

Table 2.14: HRP-Conjugated Secondary Antibodies used for western blotting.

Antibody	Dilution
HRP-Rabbit anti-Goat IgG Conjugate (Life Technologies; cat. no. 81-1620)	1/5,000
HRP-Donkey anti-Rabbit IgG Conjugate (Jackson Immunoresearch; cat. no. 711-035-152)	1/5,000
HRP-Rabbit anti-Mouse IgG Conjugate (Life Technologies; cat. no. 61-6520)	1/5,000

2.5.5 Antibody removal and re-hybridisation

When two different proteins on the same membrane with a similar molecular weight need to be detected, it is necessary to sequentially detect the proteins. In this case the first round of primary and secondary antibodies was removed as follows: the membrane was re-hydrated in

1X PBS for 15 min (room temperature with shaking). The PBS was removed and 15 mL of western blot Stripping Buffer (ThermoScientific) applied to the blot and incubated with shaking for 5-10 min. The stripping solution was removed prior to washing the blot three times with 1X PBS for a total of 30 min. To check antibody removal, the blot was treated with SuperSignal® West Pico Chemiluminescent Substrate (Thermo Scientific). If no signal was detected the membrane was hybridised with the next round of antibodies.

2.5.6 Immunofluorescence Staining and Microscopy

HEK293T cells to be transfected for immunofluorescence staining and microscopy were grown on sterilised circular glass coverslips (13 mm, No. 1 thickness; ProSciTech) in 35 mm culture dishes (Iwaki or Nunc) until 70-80% confluent. The transfected cells were grown for 24 h before the coverslips were removed and the cells fixed in 4% paraformaldehyde (PFA; Electron Microscopy Sciences) in 1X PBS (30 min, room temperature). The PFA was removed and the cells washed in 1X PBS (3 x 5 min, room temperature) before being permeabilised by addition of 0.25% Triton X-100 (Sigma-Aldrich) in 1X PBS (5 min, room temperature). The cells were again washed in 1X PBS (3 x 5 min, room temperature) before incubation in blocking solution (5% w/v Diploma brand skim milk powder in 1X PBS) (1 h, room temperature). After blocking, the cells were exposed to a combination of primary antibodies diluted in 30 µL (per coverslip of cells) of blocking solution (see Table 2.15 for dilutions; 1 h, room temperature in a humidified chamber). The cells were washed with 300 µL of the blocking solution (6 x 5 min, room temperature). Fluorophore conjugated secondary antibodies (used to detect the primary antibodies) were diluted in blocking buffer and 30 µL of diluted secondary antibody (see Table 2.16 for dilutions) was then applied to the cells (1 h, room temperature, in a darkened humidified chamber). The cells were then washed in 1X PBS (6 x 5 min, room temperature) and mounted cell side down onto glass slides *(*76 x 26 mm, No. 1 thickness, ProSciTech Cat. No. G300FB) with anti-fade mounting agent (2.4 mM n-propyl gallate (Sigma-Aldrich). Nail polish was used to seal the edges of the coverslip to prevent the cells from drying out. For DAPI (4', 6-diamidino-2-phenylindole) staining (to define the cell nuclei) cells were incubated with 30 µL of DAPI (0.005 mg/mL in 1X PBS; Life Technologies) for 5 min at room temperature. Cells were then washed three times with 1X PBS followed by mounting the coverslips onto glass slides. Prepared slides were viewed and recorded using a DeltaVision Elite fluorescence microscope fitted with a 60X oil immersion objective (Olympus; PlanApo N 60X/1.42 Oil), or a Zeiss Axio Observer (1.210) fluorescent microscope under 40X objective.

Table 2.15: Primary antibodies used for immunofluorescence staining.

Primary Antibody	Dilution
Mouse anti-V5 (Life Technologies; cat. no. R960-25)	1/200
Rabbit anti-V5 (Abcam; cat. no. ab9116)	1/500
Mouse anti-TCF7L1 (Abcam; cat.no. ab86175)	1/200
Mouse anti-TCF7L2 (Abcam; cat.no. 32873)	1/200
Mouse anti-FLAG (Sigma, cat no. F1804)	1/200
Rabbit anti-Lamin B1 (Abcam; cat. no. ab16048)	1/1,000

Table 2.16: Fluorophore-conjugated secondary antibodies used for immunofluorescence staining.

Secondary Antibody	Dilution
Alexa[594]-conjugated anti-mouse (Life Technologies, cat. no. A21203)	1/500
Alexa[488]-conjugated anti-rabbit (Life Technologies, cat. no. A21206)	1/500

2.5.7 Microscopy - Quantification of Relative Localisation Pattern

Image files were renamed by a member of the Arkell laboratory prior to quantification and later decoded to minimise operator bias. Subcellular distribution of transfected ZIC proteins in 30-100 transfected cells was scored blind per experiment. For each cell, the nuclear and cytoplasmic boundaries were traced using the Intuos® 2 graphics tablet (Wacom) and the average fluorescence intensity from each compartment [nuclear (N_F) and cytoplasmic (C_F)] measured using ImageJ (NIH) software. The cytoplasmic boundary was determined by V5-ZIC3 protein that had not entered the nucleus. In cases where cytoplasmic V5-ZIC3 was scarce and did not give a clear cytoplasmic boundary, the brightness of the image was altered to increase the intensity of the cytoplasmic protein. This trace was used for the original V5-ZIC3 channel and the nucleus was excluded to enable measurement of the cytoplasmic fluorescence density. Background intensity (measured in three different parts of each image) was subtracted from N_F and C_F to obtain corrected fluorescence intensities. These corrected N_F and C_F values were added to determine total fluorescence intensity (T_F) from each cell, before calculating the % nuclear (N_F/ T_F) and % cytoplasmic (C_F/ T_F) fluorescence. Percentages from all scored cells were averaged to determine the localization pattern of the protein in the experiment. Three independent experiments were conducted and the % localization in each cellular compartment averaged across the three experiments.

2.6 Yeast Two-Hybrid assay

Yeast two-hybrid experiments were conducted using the ProQuestTM Two-Hybrid System (Invitrogen, Cat no PQ1000101) following the manufacturer's instructions (Invitrogen 2005). *MaV203* yeast were streaked onto a YPAD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose, 20 g/L Bacto Agar, 40 mg/L adenine sulfate) agar plate and incubated at 30°C for ~48 h until single colonies were formed. Transformation of MaV203 competent cells were plated on SC-Leu-Trp plates (6.75 g/L yeast nitrogen base, 20 g/L glucose, 0.55 g/L synthetic complete mixture [Dropout complete], 0.02 g/L adenine 0.02 g/L histidine, 0.05 g/L lysine and 0.02 g/L uracil , 22 g/L Bacto Agar [pH 5.5]) for 48-60 h until single colonies were formed before patching two independent colonies onto a single SC-Leu-Trp plate. Replica cleaning was performed as required using sterile filter paper and a standard replica plating tool under sterile conditions. Colonies were replica plated onto SC-Leu-Trp-Ura (SC-Leu-Trp plates without addition of Histidine) SC-Leu-Trp+5FOA (SC-Leu-Trp plates with addition of 5FOA at 0.2% final concentration, [Formedium]), SC-Leu-Trp+3AT (SC-Leu-Trp plates with addition of 3AT at 100 mM, 50 mM, 25 mM and 10 mM; Sigma-Aldrich, A8056). Extent of colony growth was compared to positive and negative controls that were supplied within the kit.

2.7 Statistical analysis

For luciferase and BiFC assays, the luminescence or fluorescence value from each data point was used to generate a data set for statistical analysis. Nine data points were generated for each treatment (three internal replicates from each of three external repeats). Luciferase and BiFC data were all subjected to a two-way ANOVA using Genstat software and a Fisher's least significant difference (LSD) test performed (α = 0.05) to determine which treatments differ significantly from the others. If the difference between two means is larger than the computed LSD, then the treatments are considered significantly different. The test assigns a unique letter to the treatments that are significantly different to all other treatments (e.g. a, b, c, d etc.). If two treatments are not significantly different then they are assigned the same letter.

Chapter 3: Can ZIC3 physically interact with TCF7L1 *in vitro*?

3.1 Introduction

3.1.1 *Zic3* loss-of-function phenocopies elevated Wnt alleles

Deletion of the murine *Zic3* gene results in Heterotaxy, as do multiple other alleles with *Zic3* loss-of-function mutations, including the *Zic3 katun* mouse strain (Garber 1952; Gebbia et al. 1997; Klootwijk et al. 2000; Ahmed et al. 2013). The *katun* allele was first isolated in an ENU screen (Bogani et al. 2004) and showed phenotypic similarities to the *Bn* allele (which deletes the entire *Zic3* gene). Genetic analysis revealed the *katun* mutation is a result of a guanine (G) to thymine (T) substitution at position 1,283, resulting in the change of Glutamic acid into a termination codon at amino acid 249 (E249X). The mutation introduces a premature stop codon (PTC) two amino acids upstream of the ZFD of ZIC3 and results in the translation of a protein that lacks the ZFD and C-terminus of the protein. Molecular analysis by the Arkell laboratory has demonstrated that the *katun* allele produces a functionally null protein. For example, the *katun* transcript is not subject to nonsense mediated decay (NMD) and, despite lacking nuclear import signals from the ZFD, the truncated protein is able to enter the nucleus via passive diffusion (Bedard et al. 2007; Ahmed et al. 2013). In cell-based reporter assays, the truncated protein was unable to initiate transcription at the *ApoE* promoter. Similarly, the protein could not inhibit canonical Wnt signalling via β-catenin mediated transcription when overexpressed in a HEK293T TOPflash assay (Ahmed et al. 2013).

The *katun* phenotype is incompletely penetrant and highly variable. *In vivo* analysis of the *katun* allele by the Arkell laboratory has shown that heterozygous females (*Zic3*^{Ka/X}) are viable and exhibit minor defects with decreased penetrance. For example, a subset of heterozygous mice exhibit tail kinks (Figure 3.1), a phenotype commonly reported with elevated canonical Wnt signalling (MacDonald 2004; Bogani et al. 2004; Ahmed et al. 2013; Norum et al. 2015). As *Zic3* is located on the X-chromosome, male carriers of the *katun* mutation are hemizygous and display a more severe phenotype that resembles the phenotype of homozygous females (*Zic3*^{Ka/Ka}). Hemizygous embryos exhibit partial dominant lethality, with mice also exhibiting tail kinks. These embryos also have morphological defects characteristic of Type I and Type II *Zic3*^{tm1Bca} null embryos (Purandare 2002), including reduced or abnormal expression of mesoderm and endoderm markers *Cer1* and *Lim1* during gastrulation (Ahmed et al. 2013). A subset of *Zic3*^{Ka/Y} embryos recovered at 9.5 dpc display heart looping defects, including sinistral looping, ventral looping, or hearts with no looping (49%), indicative of abnormal L-R axis formation. Additionally, *katun* embryos exhibit abnormal nodes; the fundamental structure involved in L-R axis determination, and a likely source of these defects (Figure3.1). Of greater
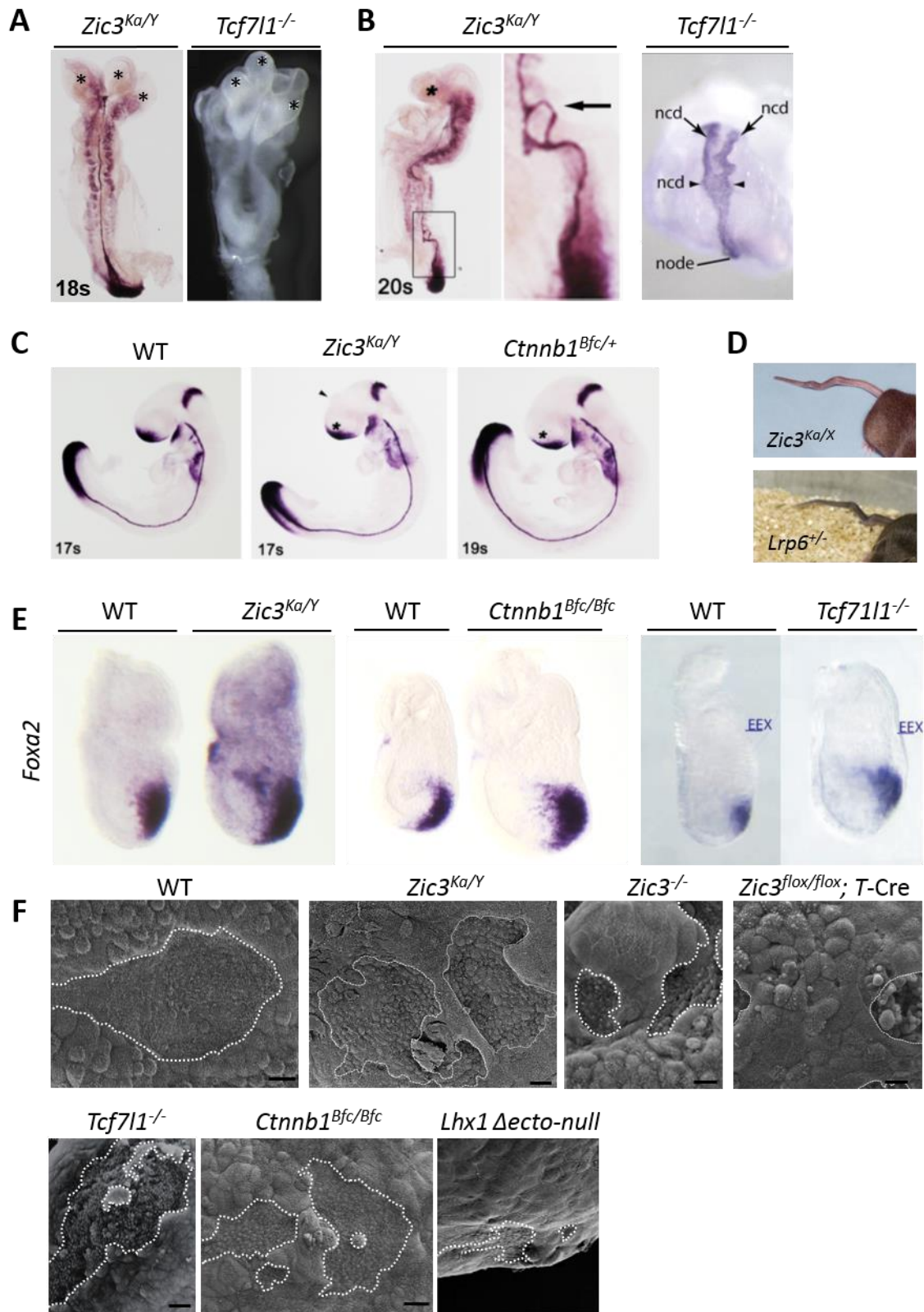
Figure 3.1: ZIC3 loss of function phenocopies elevated Wnt alleles. (A) Dorsal view of a *Zic3$^{Ka/Y}$* embryo (following WMISH to *Brachyury;* Arkell laboratory, unpublished) and *Tcf7l1$^{-/-}$* embryo (Merrill et al. 2004) exhibiting a third head fold (asterisk). **B)** Lateral view of a *Zic3$^{Ka/Y}$* embryo (following WMISH to *Brachyury* and *Follistatin;* Arkell laboratory, unpublished) and ventral view of a *Tcf7l1$^{-/-}$* embryo (following WMISH to *Foxa2;* Merrill et al. 2004) showing abnormal head development (asterisk) and axis duplications. The region within the rectangle of *Zic3$^{Ka/Y}$* is expanded in showing the axis duplication as indicated by the black arrow. **C)** Lateral view of a wildtype, *Zic3$^{Ka/Y}$* or *Ctnnb1$^{Bfc/+}$* embryo following WMISH to *Fgf8* and *Brachyury* showing reduced expression of *Fgf8* in the forebrain (asterisk) and a protrusion anteriorly to the head structure (black arrowhead; Arkell laboratory, unpublished). **D)** *Zic3$^{Ka/X}$* and Wnt gain-of-function mutant *Lrp$^{-/+}$* mice exhibit tail kinks. **E)** Z*ic3$^{Ka/Y}$* (Ahmed et al. 2013), *Ctnnb1$^{Bfc/Bfc}$ (K. Diamand; Arkell lab unpublished)* and *Tcf7l1$^{-/-}$* (Merrill et al. 2004) embryos display ectopic expression of *Foxa2. EEX: embryonic/extra-embryonic boarder.* **F)** SEM images of the surface of the node in Z*ic3$^{Ka/Y}$* (Arkell laboratory, unpublished), *Zic3$^{-/-}$* (Sutherland et al. 2013), *Zic3$^{flox/flox}$; T-Cre* (Sutherland et al. 2013), *Tcf7l1$^{-/-}$* (Merrill et al. 2004), *Ctnnb1$^{Bfc/Bfc}$* (Arkell laboratory, unpublished) and *Lhx1 Δecto-null* (Costello et al. 2015). The node in the wildtype embryo is a distinct tear-drop shaped depression with a clear edge. However, in *Zic3* loss-of-function or Wnt gain-of-function alleles, the node has an abnormal shape and a disorganised border (dotted line).

significance, *Zic3^{Ka/Y}* embryos exhibit anterior (forebrain) truncations, axis duplications and duplicated head folds (Figure 3.1). These severe phenotypes are consistently associated with elevated canonical Wnt signalling, including murine embryos null for the Wnt repressor *Tcf7l1* (Figure 3.1; Zeng et al. 1997; Merrill et al. 2004; Fossat et al. 2011; Alzahrani 2017). *Zic3^{Ka/Y}* embryos also display ectopic expression of primitive streak marker *Foxa2,* consistent with alleles *Tcf7l1^{-/-}* and *Ctnnb1^{Bfc/Bfc}*, which result in a Wnt gain-of-function effect (Figure 3.1). Moreover, *Tcf7l1^{-/-}* embryos, and other embryos that exhibit elevated canonical Wnt signalling, also have defective nodes, suggesting elevated Wnt signalling, defective node formation, and L-R axis establishments are correlated (Figure 3.1) (Merrill et al. 2004; Costello et al. 2015; Arkell unpublished). Overall, current *in vitro* and *in vivo* evidence strongly supports the hypothesis that ZIC3 acts to repress canonical Wnt signalling during murine gastrulation, preventing congenital malformations. Whether the ZIC3 and TCF7L1 proteins physically act together to achieve this outcome remains to be directly verified *in vivo*, but there is circumstantial evidence to support this.

3.1.2 *Zic3* and *Tcf7l1* have overlapping expression patterns during gastrulation

Analysis of various animal models has identified the need for restricted expression of Wnt signalling in the developing embryo. During normal gastrulation, Wnt signalling is distributed as a gradient along the anterior-posterior axis, with signal suppressed in the embryo's anterior. This model suggests that the expression of Wnt antagonists would be greater at the anterior of the embryo, while distribution of Wnt activators and secreted ligands is greater at the posterior of the embryo (Figure 3.2). Mutations that result in the gain-of-function of activators or the loss-of-function of negative regulators of the Wnt signalling pathway can result similar phenotypes as they have a similar effect on the distribution of Wnt signal: Wnt signalling becomes elevated toward the anterior of the embryo.

A prerequisite for ZIC3 to co-repress Wnt signalling in conjunction with TCF7L1 is that the two transcription factors must both be expressed in the same cells. The expression of *Zic3* (during the time this interaction is proposed to occur i.e. gastrulation) in murine embryo development has been well documented by whole mount *in situ* hybridisation (WMISH) (Elms et al. 2004). Briefly, expression of *Zic3* is initiated prior to gastrulation at 5.0 dpc and is detected throughout the extra-embryonic ectoderm and within some epiblast cells. *Zic3* is not detected in distal epiblast cells or the endoderm at these stages. This pattern of expression is maintained until the onset of gastrulation, when *Zic3* ectoderm expression begins to recede from the anterior half of the embryo. Instead, *Zic3* is expressed in the ectoderm of the embryo lined by the emerging primitive streak mesoderm (6.0 – 6.75 dpc). This mesoderm migrates anteriorly as wings to encircle the embryonic portion of the egg cylinder, or proximally as extra-embryonic mesoderm. The embryonic mesoderm wings and the ectoderm that forms the posterior amniotic fold at the posterior end of the embryo contains *Zic3* transcripts but the

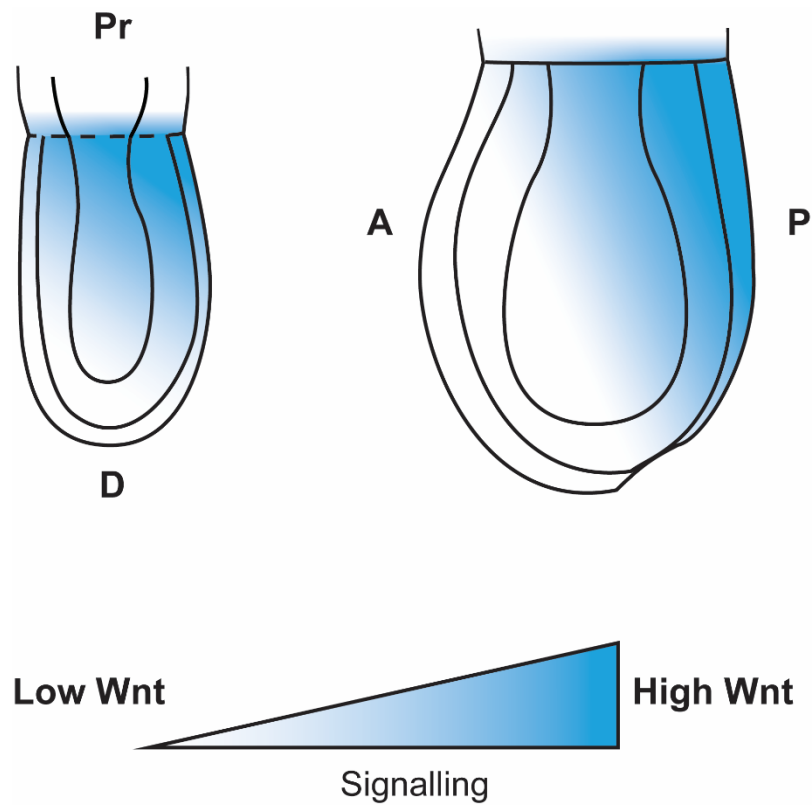**Figure 3.2: Wnt signalling is distributed as a gradient during gastrulation.** Wnt signalling is distributed as a gradient along the anterior-posterior axis during murine gastrulation. Wnt antagonists (shaded in blue) are concentrated at the posterior side, whilst Wnt antagonists are concentrated at the anterior of the embryo. *Pr: proximal, D: distal, A: anterior, P: posterior*. *Figure adapted from* Barratt 2018*.*

extra-embryonic mesoderm does not (Figure 3.3). Mid- and late-streak stage embryos express *Zic3* in the anterior definitive endoderm at the junction between the embryonic and extra-embryonic regions. *Zic3* expression is not observed in the cells of the developing ventral node at the anterior primitive streak, or in the axial mesendoderm, which emerges from the node, but it is observed in ectoderm in this region. This pattern of *Zic3* expression remains relatively constant throughout the rest of gastrulation. By the early headfold stage of development *Zic3* expression in the extra-embryonic ectoderm has ceased and expression in the embryonic ectoderm is restricted to the neuroectoderm.

Currently, whole mount analysis on mouse embryos detailing the expression pattern of *Tcf7l1* during gastrulation has not been published. Instead, protein expression using sectioned mouse embryos has been previously documented (Merrill et al. 2004). More recently, the Arkell laboratory has performed whole mount analysis of *Tcf7l1* during gastrulation (unpublished data, Figure 3.3), which extends the immunohistochemistry analysis performed by Merrill et al. 2004. Briefly, during the onset of gastrulation, *Tcf7l1* is expressed in the ectoderm of the embryo. As gastrulation proceeds and the mesoderm is formed, *Tcf7l1* transcripts are restricted to the anterior and extra-embryonic region, but are not found in the endoderm, nor are transcripts detected in the node when it forms anterior to the primitive streak. During the early allantoic bud stage, *Tcf7l1* expression begins to recede gradually from the allantoic bud and embryonic ectoderm beginning from the posterior-most region. The expression in both the ectoderm and mesoderm continues until head formation begins, establishing the domain of the brain progenitor. Notably, *Tcf7l1* is detected as a gradient along the anterior-posterior axis of the embryo, with expression concentrated at the anterior side and it also overlaps with *Zic3* expression domains (Figure 3.3). For example, during MS-LS stages, *Zic3* and *Tcf7l1* are expressed throughout the ectoderm and migrating wings of mesoderm. Additionally, at the EHF stage, both genes are expressed in the developing head folds.

3.1.3 Difficulties in studying ZIC protein function

There are several issues faced when studying ZIC protein function. These include but are not limited to lack of good commercially available antibodies for ZIC protein detection, lack of *in vitro* expressed ZIC purified protein, lack of bona fide target DNA sequences and protein partners and low expression of ZIC proteins in cell lines (Hruz et al. 2008 Appendix B). The combination of these factors makes it difficult to characterise ZIC protein function *in vivo* or *in vitro*. To overcome these issues, overexpression studies using epitope tagged ZIC proteins and highly sensitive and robust cell based assays were conducted in this thesis.
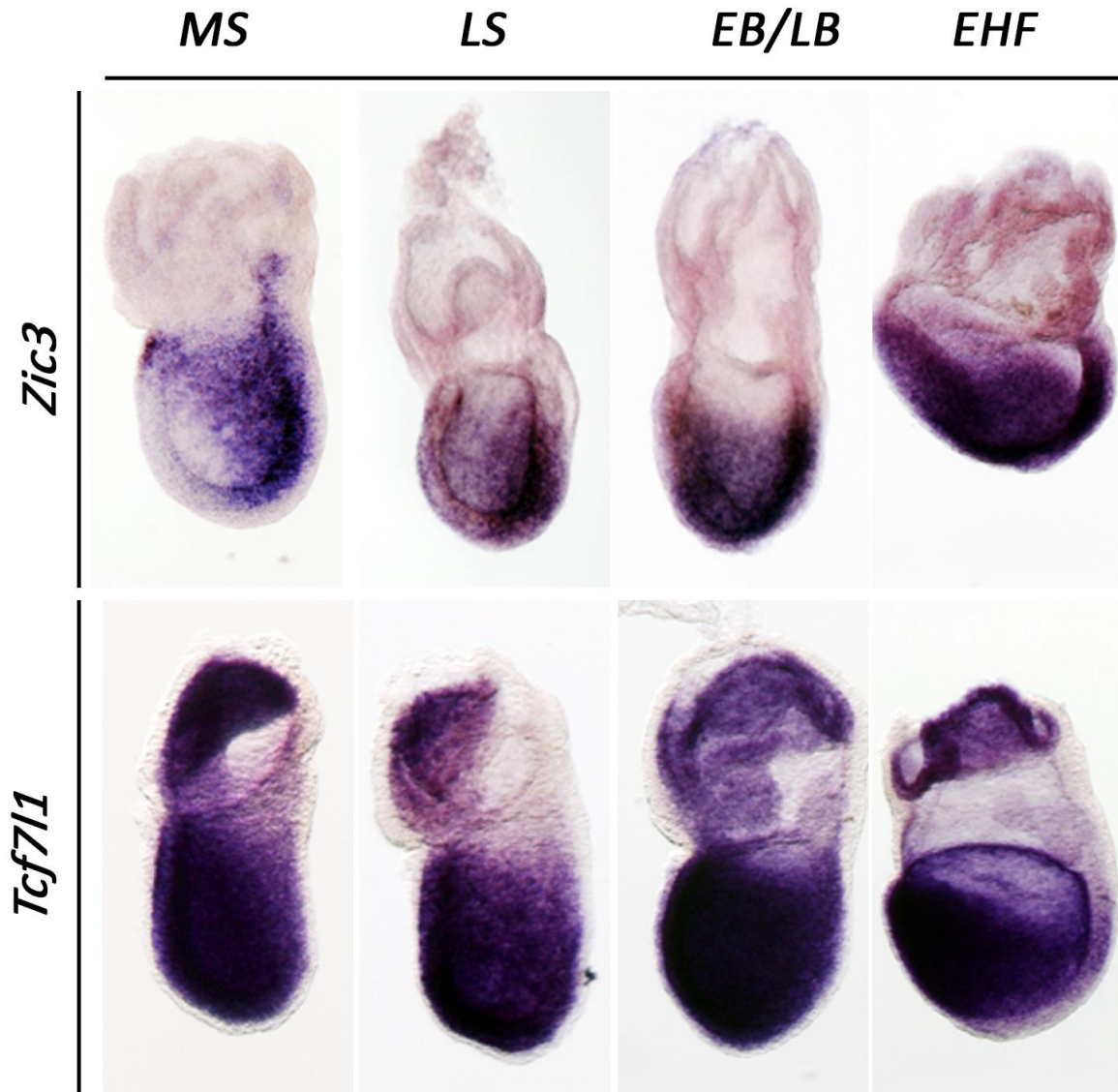
Figure 3.3: Expression of *Zic3* and *Tcf7l1* during mouse gastrulation. Lateral view of whole mount wildtype murine embryos staged from Mid-streak (MS) to early headfold (EHF) stage hybridised to either *Tcf7l1* or *Zic3*. Anterior is to the left in all images. Embryo stages are indicated on the panel. Expression of *Zic3* at MS-LS is detected in the ectoderm and mesoderm throughout gastrulation, but reduced in the endoderm. By LB stage, *Zic3* expression shifts towards the anterior ectoderm, and later becomes localised to the head-folds. Expression of *Tcf7l1* at MS-LS is restricted to the ectoderm and the mesoderm throughout gastrulation and later becomes localised to the head-folds. MS: mid streak, LS: late-streak, EB: early bud, LB: late bud, EHF: early head fold. *In situ* performed by K. Barratt and K. Walsh, Arkell laboratory.

3.1.4 Chapter aims

The TCF7L1 transcription factor acts as a Wnt repressor during early embryonic development and produces an elevated Wnt phenotype when mutated in mice (Merrill et al. 2004). Moreover, *Zic3* loss-of-function mouse models phenocopy the elevated Wnt phenotype seen in *Tcf7l1*$^{-/-}$ embryos, suggesting the ZIC3 protein also functions as a repressor of the pathway. As described in Chapter 1, both published (Pourebrahim et al. 2011; Fujimi et al. 2012) and unpublished data (Arkell laboratory) provides evidence that all ZIC proteins, including ZIC3, inhibit β-catenin mediated transcription in *Xenopus* embryos and cell based assays. Moreover, published data has demonstrated that ZIC2 protein can directly interact with TCF7L2 via its ZFD in a co-immunoprecipitation assay when overexpressed in cell lines (Pourebrahim et al. 2011).

A question arises as to whether ZIC3 and TCF7L1 share the same functional role during gastrulation: to co-operatively repress β-catenin dependent transcription. A model for ZIC3-TCF7L1 Wnt repression is proposed in which ZIC3 functions as a co-repressor of Wnt dependent transcription alongside TCF7L1 in the absence of Wnt stimulation. For this model to hold true, however, these proteins must physically and functionally interact. Currently, evidence demonstrating that these proteins function together is lacking. For example, though loss-of-function studies of both proteins produce common phenotypes, it remains to be shown whether ZIC3 and TCF7L1 are capable of direct protein interaction. **Therefore, this chapter will aim to determine whether ZIC3 and TCF7L1 are capable of physical interaction through multiple molecular assays.**

3.2 Results

The overlapping expression pattern of *Zic3* and *Tcf7l1* during murine gastrulation indicates that these proteins may physically interact. Proof that the proteins can directly interact with each other would further support the hypothesis in which they function together to repress Wnt signalling during gastrulation. A variety of methods are routinely used in molecular biology to demonstrate protein-protein interactions via overexpression of candidate interacting molecules. The use of multiple techniques, also known as orthogonal repeats, provide stronger evidence of the interaction compared to performing the same technique numerous times.

3.2.1 Method 1: Co-immunoprecipitation

A widely used method to investigate protein-protein interactions is co-immunoprecipitation, a technique that uses specific antibodies to 'pull down' and isolate proteins and their binding partners. Briefly, an antibody specific to one protein is incubated with a cell lysate that contains this target protein and its binding partner. This allows an antibody-protein complex to be formed and captured (often through the use of agarose or magnetic beads), while non-bound proteins are washed away. The target protein and binding partner are eluted and analysed via SDS-PAGE electrophoresis and western blot analysis to confirm the presence of both proteins. Successful immunoprecipitations, however, are highly dependent on the method and reagents used. There are several different lysis and wash buffers available that can be applied to immunoprecipitation protocols as protein complexes can be sensitive to the type of buffers used. As such, protocol optimisation is often required for each experiment. That is, there is not a 'one method suits all' approach for precipitating protein complexes.

Previous studies have demonstrated that ZIC2 can bind TCF7L2 in co-immunoprecipitation experiments using overexpressed proteins in HEK293T cells (Pourebrahim et al. 2011). Due to the homology between ZIC2 to ZIC3 and TCF7L2 to TCF7L1, a co-immunoprecipitation experiment was conducted using the protocol outlined by Pourebrahim et al. 2011 and also established protocols in the Arkell laboratory to see if ZIC3 directly binds TCF7L1. HEK293T cells were transfected with V5-ZIC3 and Myc-Tcf7l1 expression plasmids. Additionally, V5-ZIC2 and Flag-TCF7L2 expression plasmids were included as a positive control. An anti-V5 antibody was used to pull down V5-ZIC3/Myc-TCF7L1 or V5-ZIC2/Flag-TCF7L2 protein complexes from the cell lysate, and either anti-Flag, anti-Myc, anti-TCF7L1 or anti-TCF7L2 antibodies were used to detect the Myc-TCF7L1 or Flag-TCF7L2 proteins via western blotting (as well as anti-V5 to detect V5-ZIC proteins). Despite multiple attempts, co-immunoprecipitation experiments with ZIC3 and TCF7L1 were inconclusive. Western blot analysis could not detect either TCF7L1 or TCF7L2 (i.e. the protein binding partner) following co-immunoprecipitation and could only detect V5-ZIC proteins (the original pulled down protein) (Data

not shown). As the positive control (ZIC2 and TCF7L2) failed to detect TCF7L2 in the pull down, it was concluded there was an issue with the experimental technique, therefore the co-immunoprecipitation used could not confirm (or refute) an interaction between ZIC3 and TCF7L1. Due to the vast array of troubleshooting possibilities, other protein binding techniques were prioritised. Expression plasmids were sent to our collaborators (R. Houtmeyers, S. Tejpar; KU Leuven, Belgium), who have well established co-immunoprecipitation protocols within their laboratory. During thesis write-up, a successful Co-IP demonstrating a direct ZIC3 and TCF7L1 interaction was conducted by Rob Houtmeyers (KU Leuven, Belgium). Images, however, were not made available before thesis submission.

3.2.2 Method 2: Yeast two-hybrid

The yeast two-hybrid method was developed by Fields and Song in 1989 and is a simple and cost effective technique for identifying novel protein interactions as well as assessing known strong and weak protein interactions (Fields and Song 1989; Brückner et al. 2009). The system is based on the activation of reporter genes by the GAL4 transcription factor. In the yeast two-hybrid assay, the GAL4 protein is split into two halves; the DNA binding domain and the activation domain. Stimulation of the reporter genes in the assay only occurs when the GAL4 protein is reconstituted into a functional transcription factor, which can drive gene expression (Mehla et al. 2015).

The two proteins to be examined for interaction are termed the 'bait' and 'prey' proteins respectively. The 'bait' protein is expressed as a 'hybrid' as it is fused with the DNA binding domain of GAL4 and can recognise and bind to specific GAL4 DNA binding sites, which are located within a specific reporter construct integrated in the yeast genome. The 'prey' is similarly a fusion of the activation domain of the GAL4 protein and the second protein of interest. Reporter gene expression will be activated if a functional GAL4 protein is generated. This is only achieved if the bait and prey proteins interact to bring both the DNA binding domain and the activation domain of the GAL4 transcription factor together (Figure 3.4). In the absence of bait and prey interaction, expression of the reporter gene will not be induced as neither the bait or the prey are capable of transactivation in isolation (Figure 3.4).
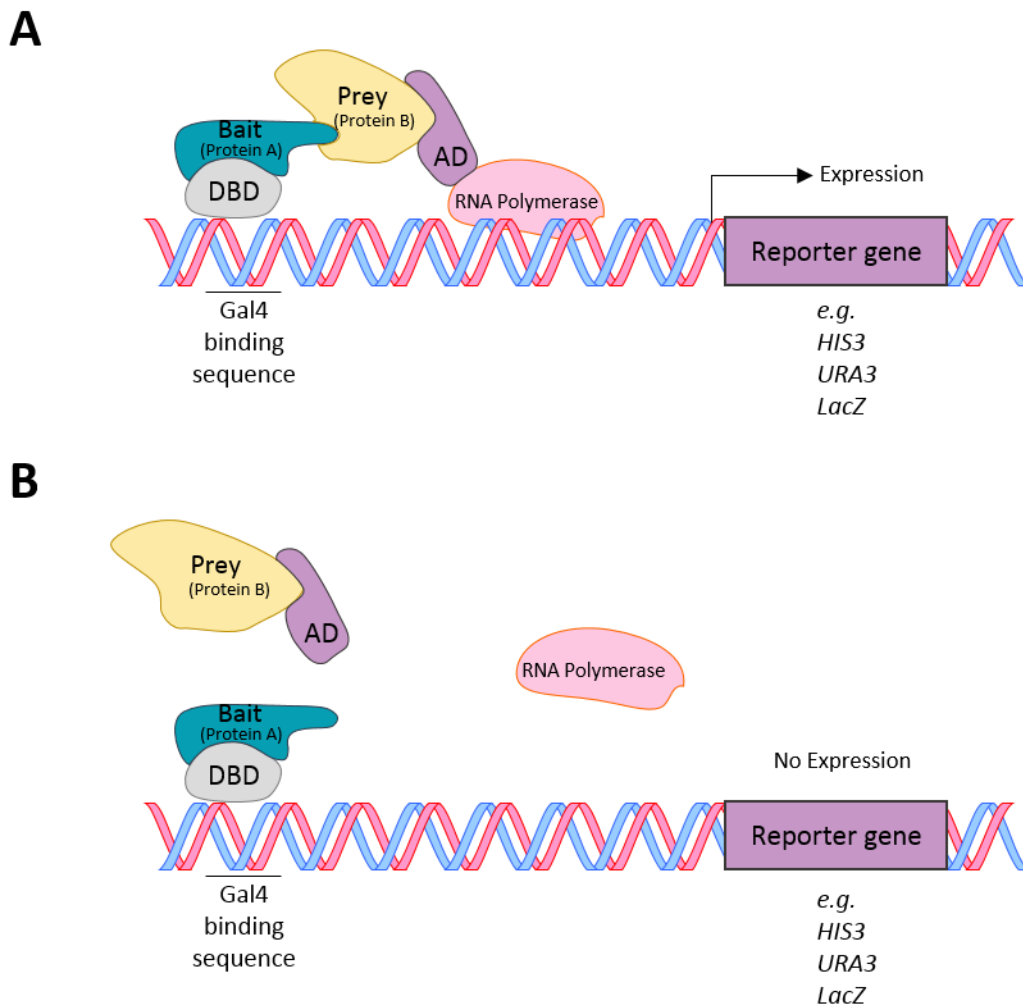
Figure 3.4: Schematic of the yeast two-hybrid system. The yeast two-hybrid system utilises the DNA binding and activation domains of the GAL4 transcription factor. The DNA binding domain recognises and binds specific DNA sequences of a reporter gene and the activation domain is involved in interacting with transcriptional machinery including RNA polymerase. The 'bait' protein (Protein A) is fused to the DNA binding domain of GAL4, while the 'prey' protein (Protein B) is fused to the GAL4 activation domain. **(A)** If the bait and prey proteins interact, a functional GAL4 transcription factor is reconstituted and can drive expression of reporter genes. **(B)** If the bait and prey proteins do not interact, the reporter gene is not activated. DBD: DNA binding domain, AD: activation domain.

The yeast strain used here (MaV203, *S. cerevisiae*) acts as a host strain for the bait and prey plasmids and contains copies of reporter genes *HIS3*, *URA3* and *lacZ* downstream from GAL4-binding sites that have been integrated independently at different loci of the yeast genome. These genes are placed under the control of the GAL4 transcription factor and are essential for yeast growth. For example, the *HIS3* and *URA3* reporters are included for the biosynthesis of Histidine and Uracil respectively, and the yeast are grown on media lacking these essential acids, therefore, only yeast cells that contain interacting bait and prey proteins will drive expression of *HIS3* or *URA3* from the reporter and allow the formation of colonies on selective agar plates. A second means to assess the expression of *HIS3* is the addition of the inhibitor 3-aminotriazole (3AT) to selective medium: 3AT is a competitive inhibitor of HIS3, and when added to selective medium in increasing concentrations, enables direct assessment of *HIS3* expression levels. For example, yeast growth on selective plates containing higher concentration of 3AT would indicate greater *HIS3* expression and consequently suggests a stronger protein-protein interaction, whilst a weak interaction is shown if yeast only grow on low concentrations of 3AT. Similarly, induction of *URA3* can also be measured by the addition of the selective compound 5-Fluoroorotic Acid (5FOA). Interacting proteins that express *URA3* results in the conversion of 5FOA into a toxic compound, 5-fluorouracil, resulting in cell death. Therefore 5FOA selects against positive interactions by killing yeast expressing *URA3*. As such, cell growth on a medium containing 5FOA would suggest absent (or weak) protein interaction.

To determine whether ZIC3 and TCF7L1 proteins interact, both wildtype *ZIC3* and *Tcf7l1* cDNA were transferred into yeast two-hybrid compatible vectors as part of the Life Technologies ProQuest™ Two-Hybrid System. Vector conversion uses Gateway LR recombination cloning technology to insert both *ZIC3* and *Tcf7l1* CDS into the required 'bait' and 'prey' vectors (pDEST-32 and pDEST-22). The cDNA for each protein of interest was cloned into both 'bait' and 'prey' vector backbones to ensure false negatives resulting from the fusion protein physically blocking the binding sites for each protein of interest was avoided. Four new plasmids were thereby generated to enable all conformations of bait and prey proteins: p32-ZIC3-wt, p32-Tcf7l1, p22-ZIC3-wt and p22-Tcf7l1 (Table 3.1).

Table 3.1: Combination of bait and prey plasmids used in Yeast Two-Hybrid assays. *The bait plasmid includes GAL4 DNA binding domain and the prey plasmid includes GAL4 activation domain*.

Plasmid Combination	'Bait' plasmid	'Prey' plasmid
1	p32-ZIC3-wt	p22-Tcf7l1
2	p32-Tcf7l1	p22-ZIC3-wt

3.2.2.1 ZIC3 and TCF7L1 show mild interaction in the Yeast two-hybrid assay

To investigate the interaction between ZIC3 and TCF7L1 proteins, a yeast two-hybrid experiment was performed using the ProQuest™ Two-Hybrid system. Three different controls supplied in the kit were included in the assay based on the interaction between Krev1 (a member of the Ras family of GTP binding proteins) and RalGDS (the Ral guanine nucleotide dissociator stimulator protein) proteins (Herrmann et al. 1996; Serebriiskii et al. 1999): wildtype Krev1 and RalGDS generate a strong interaction control, whereas two mutant forms of RalGDS generate a weak and absent protein interaction, respectively. Two different conformations of ZIC3 and TCF7L1 were included to enable each to act as the 'prey' and 'bait' plasmids (Table 3.1). All plasmids were transformed into the MaV203 yeast strain and subsequently patch and replica plated onto selection plates (Figure 3.5).

The results produced from the yeast two-hybrid assay were inconclusive. ZIC3:TCF7L1 transformants, from both combination 1 and combination 2, showed minimal (but not absent) growth on SC-Leu-Trp-Ura and SC-Leu-Trp-His+3AT selection plates (Figure 3.5). Slight growth was seen on these plates compared to the 'absent' interaction control, which only presented with a faint haze of cells, presumed to be artefacts from replica plating. This is indicative that Histidine and Uracil reporter activation may be slightly induced from interacting ZIC3/TCF7L1 bait and prey proteins. More notably, cell death was present on SC-Leu-Trp+5FOA plates (ZIC3-TCF7L1 Conformation 1) at a level comparable to the weak interaction control, suggesting ZIC3 and TCF7L1 have a weak protein interaction (Figure 3.5).

It should be noted, however, that the controls provided in the ProQuest™ Two-Hybrid kit do not clearly differentiate a strong and weak protein interaction in any experimental repeat conducted. In most cases, the strong and weak interaction controls looked very similar, making it difficult to interpret a true weak protein interaction from a strong one (a criticism of the kit used, not the technique itself). Nonetheless, results from the yeast two-hybrid assay did not convincingly indicate whether ZIC3 and TCF7L1 are protein partners or not. Therefore, other methods to identify protein-protein interactions were investigated to clarify this result.
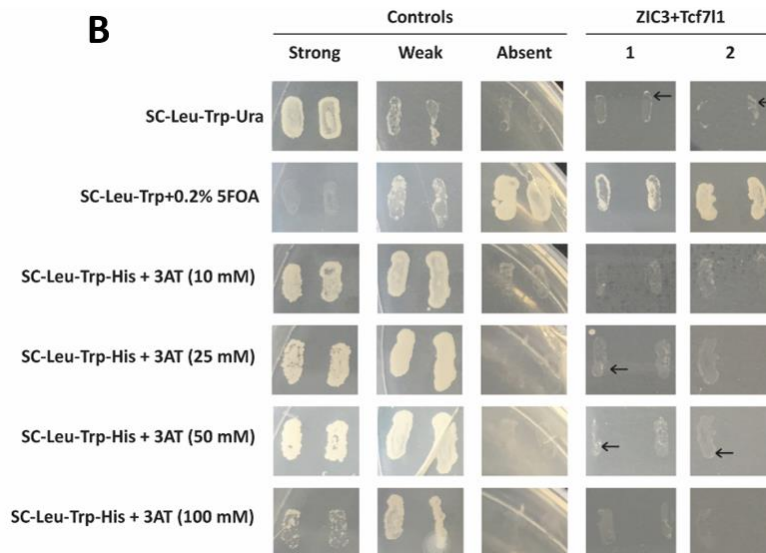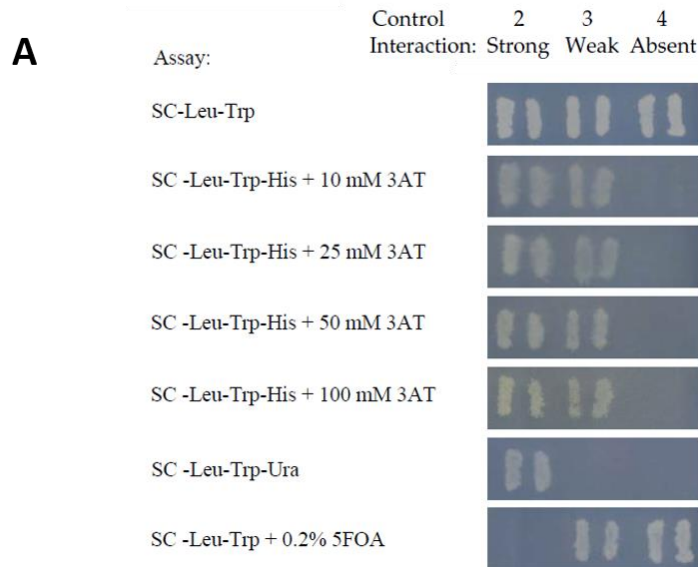
Figure 3.5: Yeast two-hybrid suggests a weak interaction between ZIC3 and TCF7L1 proteins. A yeast two-hybrid assay was conducted using the ProQuest™ Two-Hybrid System (Invitrogen). Colonies were grown on the selection media shown to determine the level of protein interaction. A strong, weak and absent interaction controls were included as supplied with the kit. **(A)** Expected results for investigating protein interaction using the ProQuest Two-Hybrid system. Image sourced from Invitrogen ProQuest User manual. **(B)** Two combinations of ZIC3 and TCF7L1 plasmids were transformed. (1) p32-ZIC3-wt, p22-Tcf7l1 (ZIC3 as bait, Tcf7l1 as prey); (2) p32-Tcf7l1, p22-ZIC3-wt (Tcf7l1 as bait, ZIC3 as prey). Transformed yeast colonies were patched onto a master plate, replica plated onto selection plates and incubated for 24 hours at 30°C. Plates were replica cleaned as necessary and incubated for a further 48 hours. Two independent colonies are shown for each treatment. Black arrows indicate slight colony growth.

3.2.3 Method 3: Proximity ligation assay

A Proximity Ligation Assay (PLA) is a highly specific and sensitive method designed to identify proteins, protein-protein interactions and post-translational modifications using standard immunofluorescence techniques (Weibrecht et al. 2014). PLA uses antibodies to target the protein(s) of interest, which theoretically allows the technique to be used on both tissues and cells (Leuchowius et al. 2011). For example, to detect protein-protein interactions, primary antibodies are first hybridised to the two proteins of interest (Figure 3.6). Secondary antibodies that are conjugated to complementary oligonucleotide probes bind to the primary antibodies attached to the proteins. If the two proteins of interest are in close proximity to each other (within 40 nm), the two oligonucleotides can hybridise and form a circular DNA molecule, which is then amplified by a mechanism known as rolling circle amplification. The amplification step produces several repeating sequences which are recognised by complementary detection oligos coupled to fluorochromes. The resulting fluorescent signal can be detected using a standard fluorescent microscope (Figure 3.6). If the PLA probes are too far apart, indicative of the two proteins not in binding distance, no fluorescence will be detected (Fredriksson et al. 2002; Söderberg et al. 2006; Weibrecht et al. 2014).
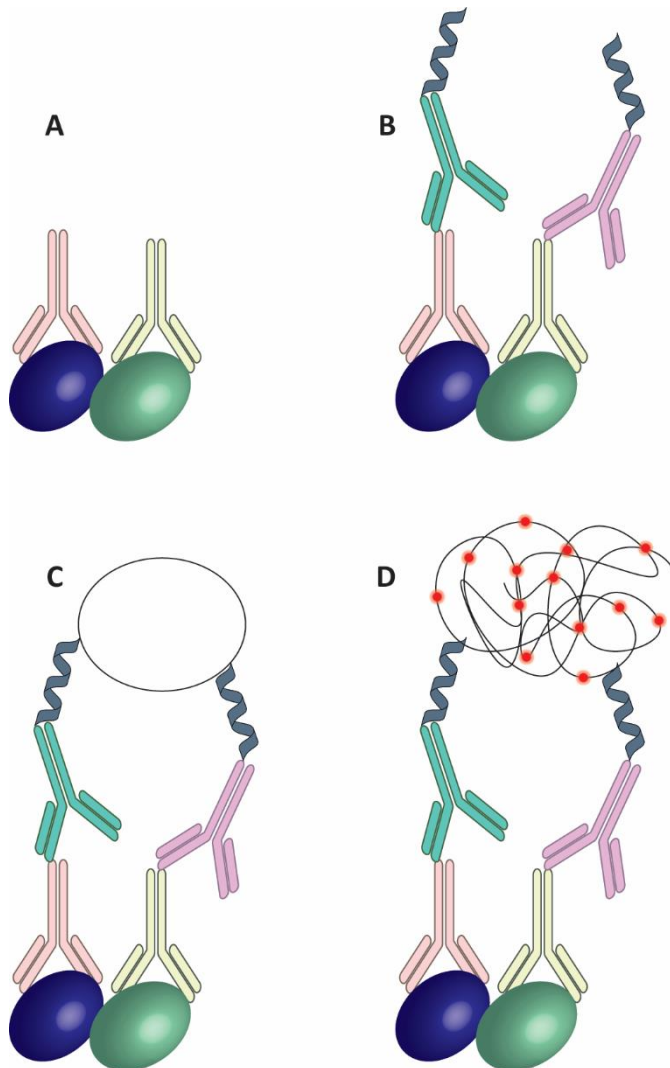
Figure 3.6: Schematic of how fluorescence is produced in a Proximity Ligation Assay. *In situ* Proximity Ligation assay (PLA) can be used to detect protein-protein interactions with high specificity and sensitivity. **(A)** Two primary antibodies recognise and bind to the proteins of interest. **(B)** Secondary antibodies conjugated to short oligonucleotide sequences (PLA probes) hybridise to the primary antibodies. **(C)** If the PLA probes are in close proximity, the probes will ligate and produce a circular DNA template, which is amplified via rolling circle amplification. **(D)** This creates repetitions of the same sequence to be recognised by complementary detection oligonucleotides coupled to fluorochromes which can be visualised through fluorescent microscopy.

3.2.3.1 Proximity ligation assay detects an interaction between ZIC3 and TCF7L1

The DuoLink® PLA (Sigma) was used to demonstrate the interaction between ZIC3 and TCF7L1 as this system can be used to detect stable, weak or transient protein-protein interactions (Fredriksson et al. 2002). HEK293T cells were co-transfected with a combination of either V5-ZIC3, V5-ZIC2, Myc-Tcf7l1, FLAG-TCF7L2 or V5-β-catenin and a PLA was performed according to manufactures instructions (Chapter 2.4.6). As expected , wildtype ZIC2 protein was able to interact and produce fluorescent signal when transfected with FLAG-TCF7L2, confirming this interaction that has been previously documented through co-immunoprecipitation by Pourebrahim et al. 2011 using overexpressed proteins. A secondary positive control V5-β-catenin-wt transfected with Myc-Tcf7l1 similarly produced detectable fluorescent signal when overexpressed in HEK293T cells. Fluorescent signal was also detected following combined overexpression of V5-ZIC3 and Myc-Tcf7l1 expression plasmids, suggesting a positive interaction between these two proteins (Figure 3.7). This assay indicates ZIC3 and TCF7L1 are in close enough proximity to each other to infer the two proteins interact.

Although this system can be used to confirm or investigate protein-protein interactions, limitations are encountered when assessing the strength of such interactions. The output of PLA shown here produces binary data in that an interaction is either present or absent. Although not pursued, it could be possible to use imaging software to trace around cells and quantify the fluorescence intensity for each cell positive for a PLA interaction. This has the potential to be used for downstream analysis to identify mutations that disrupt or hinder the ZIC3-TCF7L1 interaction. However, due to the cost of probes, detection reagents and antibodies for individual transfections, alternative methods were explored.
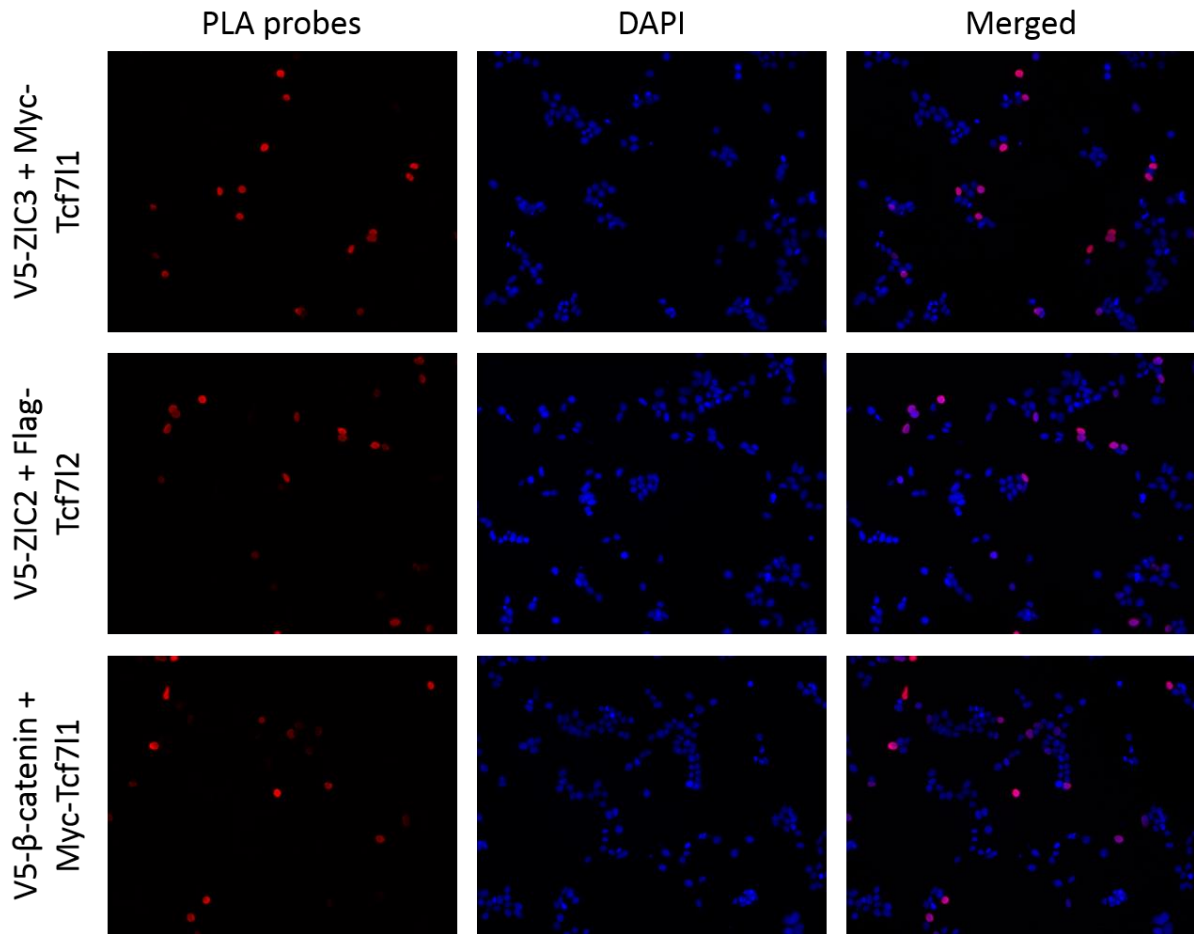
Figure 3.7: ZIC3 and TCF7L1 produce detectable fluorescence in a Proximity Ligation Assay. HEK293T cells were grown on coverslips and co-transfected with the expression plasmids shown. 24 hours post-transfection, cells were fixed and prepared for immunofluorescence. Primary antibodies α-V5, α-TCF7L1 or α-TCF7L2 were hybridised to appropriate cells before addition of PLA probes. Images of cells were captured with a fluorescent microscope. Three independent experiments were performed; one representative image from each PLA is shown.

3.2.4 Method 4: Bimolecular fluorescence complementation

Bimolecular fluorescence complementation (BiFC) is a relatively new technique for identifying protein-protein interactions. The assay relies on the Venus fluorescent protein producing a fluorescent signal when two proteins are in close proximity (Kodama and Hu 2012; Miller et al. 2015). Venus is a modified version of the Yellow Fluorescent Protein (which itself is a modified version of the Green Fluorescent Protein derived from the jellyfish *Aequorea Victoria*) (Nagai et al. 2002). The Venus protein is split into an N' and C' half, and each half is attached to the two test proteins of interest. The N' and C' halves of Venus are themselves non-fluorescent and will only produce a fluorescent signal when the fragments come together, reassemble and fold into their correct structure (Figure 3.8). When these test proteins interact or are in close enough proximity, the two fragments of Venus reassemble and produce a detectable fluorescent signal. BiFC has many advantages over other techniques as it allows direct visualisation of protein-protein interactions in living cells. Unlike other techniques, there is no need for fixation or cell lysis to detect protein interactions (Miller et al. 2015). Furthermore, BiFC is a quantitative technique that allows direct comparisons when analysing mutations (e.g. when identifying hypomorphic mutations).

Though some BiFC compatible expression vectors already existed in the Arkell laboratory (expression plasmid pV1-hZIC3-wt), several expression vectors were also generated for this technique: *mTcf7l1* cDNA was cloned into the pV2-ORF vector backbone, and mouse *Zic3* cDNA and human *TCF7L1* cDNA were transferred to the V1-ORF and V2-ORF vector backbones, respectively, to investigate whether there is a difference in protein-protein interaction between mouse and human ZIC3 and TCF7L1 proteins. The pV1-ZIC3 constructs produce a fusion protein with the N-terminal fragment of the Venus fluorescent protein attached to the ZIC3 protein (V1 tag), whilst the pV2-TCF7L1 constructs produce a fusion protein of TCF7L1 with the C-terminal fragment (V2) of the Venus fluorescent protein. If pV1-ZIC3-wt protein interacts with pV2-TCF7L1, it will cause the V1 and V2 fragments of Venus to reform and give a detectable fluorescent signal (Figure 3.8).

3.2.4.1 Bimolecular fluorescence complementation assay suggests ZIC3 and TCF7L1 interact

To test the relative ability of wildtype ZIC3 to interact with TCF7L1 protein, HEK293T cells were transfected with V1-hZIC3-wt or V1-mZic3-wt along with either V2-hTCF7L1 or V2-mTcf7l1. Fluorescence intensity was measured using a fluorescent plate reader 48 hours post-transfection. Cells were then lysed and assessed by SDS-PAGE electrophoresis and western blot to confirm protein expression (Figure 3.9). When human ZIC3 was transfected with either mouse or human TCF7L1, a 2.8 fold increase in fluorescence compared to background levels was observed (Figure 3.9). Similarly, a
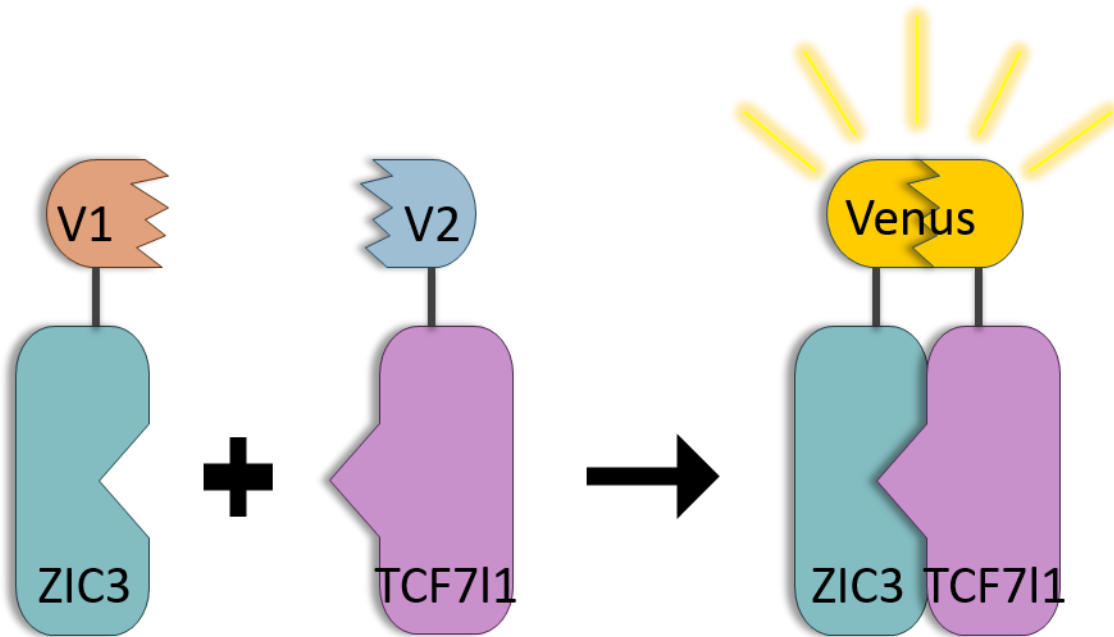
Figure 3.8: Schematic of the production of fluorescence in the BiFC assay. pV1-hZIC3-wt produces a fusion protein with the V1 fragment of the Venus protein attached. pV2-Tcf7l1 produces the TCF7L1 fusion protein with the V2 fragment of Venus attached. Only when the V1-hZIC3 protein is in close enough proximity to the V2-TCF7L1 protein will the Venus protein reform to give a fluorescent signal that can be visualised and/or quantified.
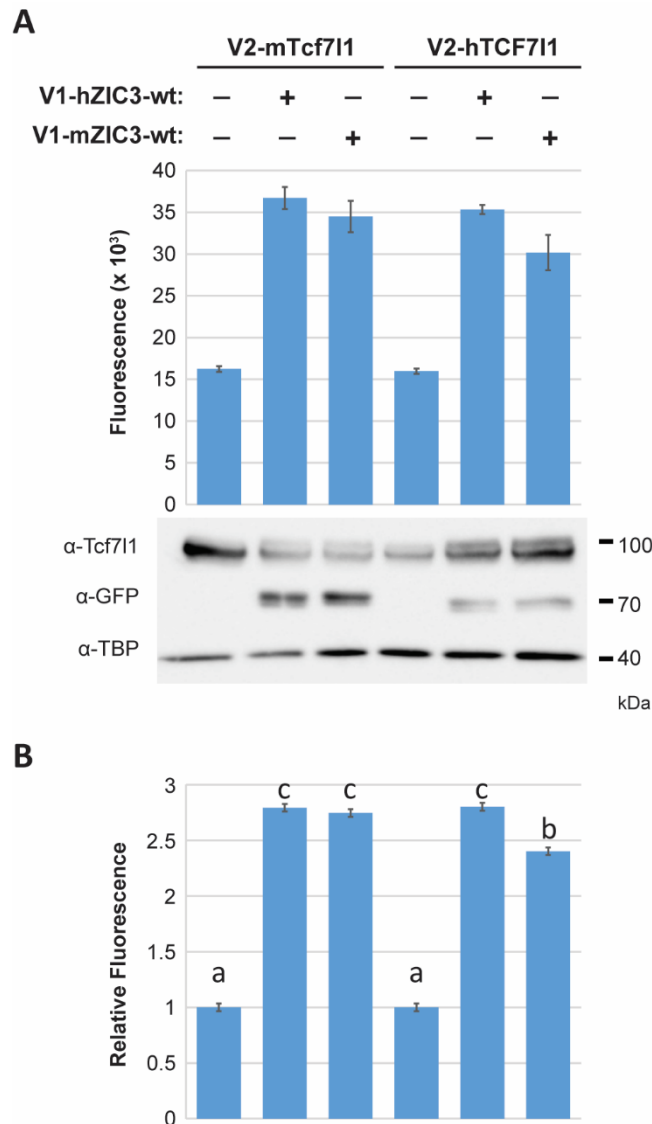
Figure 3.9: ZIC3 interacts with human and mouse TCF7L1 with the same strength. HEK239T cells were co-transfected with V1-hZIC3-wt or V1-mZIC3-wt along with either V2-hTCF7L1 or V2-mTcf7l1 expression plasmids. The mean fluorescence intensity emitted from each of three internal replicate samples was detected 48 hours post-transfection. Three independent experiments were performed. **(A)** Mean fluorescence intensity of three internal replicates from one representative experiment. Error bars represent the SD between the three internal replicates. A western blot was performed to confirm expression of transfected proteins. Primary antibodies α-TCF7L1 was used to detect TCF7L1 protein, α-GFP was used to detect ZIC3 protein and α-TBP was used as a loading control. **(B)** Pooled data from three independent repeats. Values depict the mean relative luciferase values. Error bars depict SEM. Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc.

2.8 fold increase was detected when mZIC3 was transfected with mTCF7L1, but only a 2.4 increase when transfected with hTCF7L1. These fold changes suggest that ZIC3 and TCF7L1 proteins are in close enough proximity to interact and suggests a direct protein-protein interaction. Moreover, no significant difference was found between hZIC3:hTCF7L1 transformants when compared to the interaction between hZIC3 and mTCF7L1 (Figure 3.9). Protein alignment of human and mouse TCF7L1 protein sequences illustrates they share 95.2% similarity (Figure 3.10). As such, these proteins were used interchangeably in downstream BiFC analysis.

A common control for BiFC or other proximity assays that confirm a specific interaction between your two proteins is occurring is to include a protein that contains a loss-of-function mutation or lacks the domain responsible for the protein-protein interaction. Such a mutation is yet to be identified for ZIC3 or TCF7L1 proteins, therefore this control was unable to be included. An alternative control commonly used in BiFC studies is a competition assay. In this assay, increasing amounts of a non-Venus tagged protein is added into the assay to act in competition with the Venus tagged protein. As increasing amounts of this non-Venus tagged protein is added, the total fluorescence is expected to decrease. The work presented in Chapter 4 identifies a human ZIC3 variant, ZIC3-C268X, a mutant protein that interacts with TCF7L1 with greater affinity in the BiFC assay compared to wildtype ZIC3 (Chapter 4, Figure 4.5). As such, this V5 tagged ZIC3 variant was used as the competitor in these experiments to demonstrate the specificity of the assay. HEK293T cells were co-transfected with V1-ZIC3-wt, V2-Tcf7l1 and increasing amounts of the non-fluorescent plasmid V5-ZIC3-C268X. Adding increasing concentrations of V5-ZIC3-C268X decreased overall fluorescence, demonstrating the fluorescence observed is due to the specific interaction between the ZIC3 and TCF7L1 proteins rather than random association of the V1 and V2 protein tags. (Figure 3.11).

```
95.2% identity in 588 residues overlap; Score: 2920.0; Gap frequency: 0.7%

Human        1 MPQLGGGGGGGGGGSGGGGGSSAGAAGGGDDLGANDELIPFQDEGGEEQEPSSDSASAQR
Mouse        1 MPQLGGGRGGAGGG---GGGSGAGATSGGDDLGANDELIPFQDEGGEEQEPSSDTASAQR
               ******* ** ***    **** ***   *************************** *****

Human       61 DLDEVKSSLVNESENQSSSSDSEAERRPQPVRDTFQKPRDYFAEVRRPQDSAFFKGPPYP
Mouse       58 DLDEVKSSLVNESENQSSSSDSEAERRPQPARDAFQKPRDYFAEVRRPQDGAFFKGGAYP
               ****************************** ** ***************** *****   **

Human      121 GYPFLMIPDLSSPYLSNGPLSPGGARTYLQMKWPLLDVPSSATVKDTRSPSPAHLSNKVP
Mouse      118 GYPFLMIPDLSSPYLSNGPLSPGGARTYLQMKWPLLDVPSSATVKDTRSPSPAHLSNKVP
               ************************************************************

Human      181 VVQHPHHMHPLTPLITYSNDHFSPGSPPTHLSPEIDPKTGIPRPPHPSELSPYYPLSPGA
Mouse      178 VVQHPHHMHPLTPLITYSNDHFSPASPPTHLSPEIDPKTGIPRPPHPSELSPYYPLSPGA
               ************************ ***********************************

Human      241 VGQIPHPLGWLVPQQGQPMYSLPPGGFRHPYPALAMNASMSSLVSSRFSPHMVAPAHPGL
Mouse      238 VGQIPHPLGWLVPQQGQPMYSLPPGGFRHPYPALAMNASMSSLVSSRF-PHMVAPAHPGL
               ***********************************************  ***********

Human      301 PTSGIPHPAIVSPIVKQEPAPPSLSPAVSVKSPVTVKKEEEKKPHVKKPLNAFMLYMKEM
Mouse      297 PTSGIPHPAIVSPIVKQEPAAPSLSPAVSAKSPVTVKKEEEKKPHVKKPLNAFMLYMKEM
               ******************** ******** ******************************

Human      361 RAKVVAECTLKESAAINQILGRKWHNLSREEQAKYYELARKERQLHSQLYPTWSARDNYG
Mouse      357 RAKVVAECTLKESAAINQILGRKWHNLSREEQAKYYELARKERQLHAQLYPTWSARDNYG
               ********************************************** *************

Human      421 KKKKRKREKQLSQTQSQQQVQEAEGALASKSKKPCVQYLPPEKPCDSPASSHGSMLDSPA
Mouse      417 KKKKRKREKQLSQTQSQQQIQEAEGALASKSKKPCIQYLPPEKPCDSPASSHGSALDSPA
               ******************* *************** ****************** *****

Human      481 TPSAALASPAAPAATHSEQAQPLSLTTKPETRAQLALHSAAFLSAKAAASSSGQMGSQPP
Mouse      477 TPSAALASPAAPAATHSEQAQPLSLTTKPEARAQLALHSAAFLSAKAAASNSSQMGSQPP
               ****************************** ****************** * *******

Human      541 LLSRPLPLGSMPTALLASPPSFPATLHAHQALPVLQAQPLSLVTKSAH
Mouse      537 LLSRPLPLGSMPAALLTSPPTFPATLHAHQALPVLQAQPLSLVTKSAH
               ************ *** *** ***************************
```

Figure 3.10: Human TCF7L1 and mouse TCF7L1 share 95.2% protein sequence. Amino acid alignments between human TCF7L1 and mouse TCF7L1 using the online protein alignment tool SIM illustrates they share 95.2% of their protein sequence. Identical amino acids are indicated by an asterisk (*) below the aligned sequences.
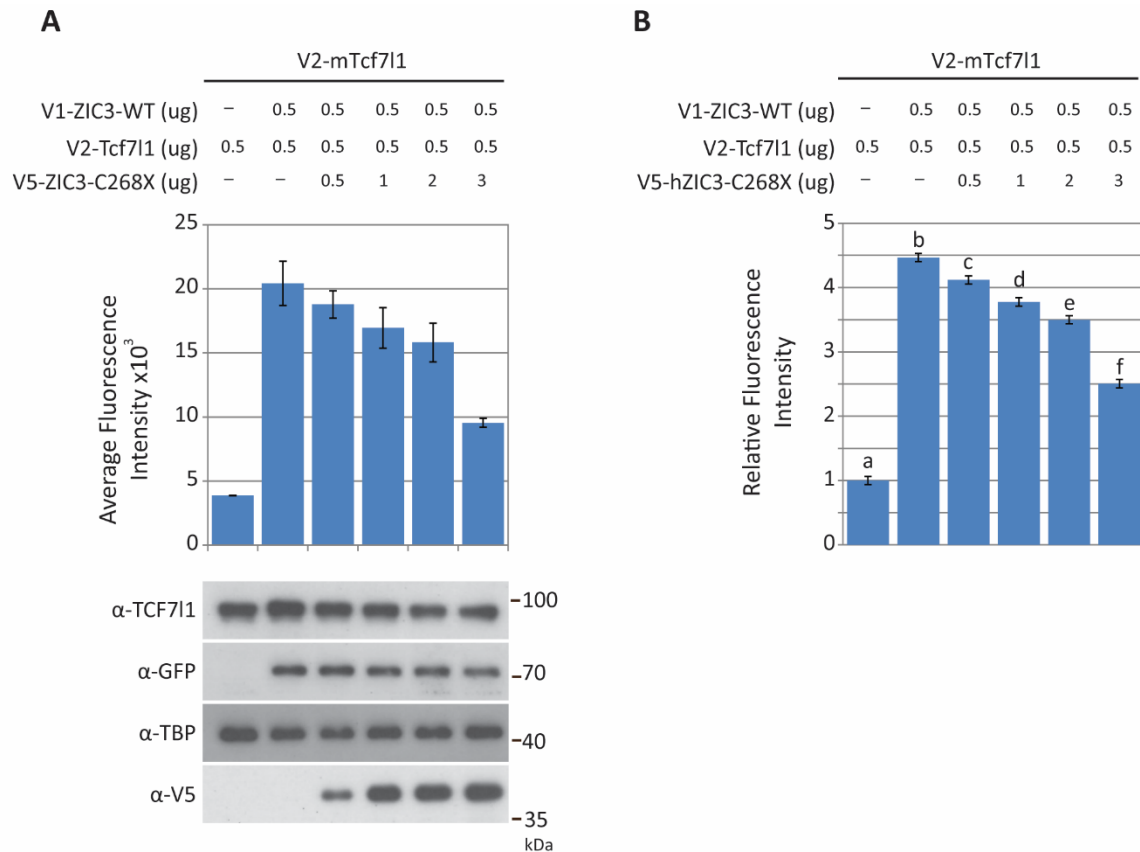
Figure 3.11: Competition assay demonstrates V1-ZIC3 and V2-Tcf7l1 interact. HEK293T cells were co-transfected with V1-hZIC3-wt, V2-mTcf7l1 constructs along with increasing amounts of pnV5-ZIC3-C268X. The mean fluorescence intensity emitted from each of three internal replicate samples was detected 48 hours post-transfection. Three independent experiments were performed. **(A)** Mean fluorescence intensity of three internal replicates from one representative experiment. Error bars represent the SD between the three internal replicates. The corresponding western blot is shown below. α-TCF7L1 antibody was used to detect V2-Tcf7l1 protein, α-GFP antibody used to detect V1-hZIC3-wt protein, α-V5 antibody was used to detect V5-hZIC3 protein and α-TBP was used as a loading control. wt: wildtype. **(B)** Pooled data from three independent repeats. Values depict the mean relative luciferase values. Error bars depict SEM Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc.

113

3.3 Discussion

The findings of this study build upon a novel hypothesis for ZIC3 function in the Wnt pathway during gastrulation. *Zic3* loss-of-function mouse models demonstrate elevated Wnt phenotypes including forebrain truncations and axis duplications, suggesting ZIC3 functions as a repressor of the signalling pathway during embryogenesis. Given that ZIC2 can physically interact with a related protein TCF7L2 in cell assays and that *Zic3* and *Tcf7l1* share expression domains during gastrulation, it is proposed ZIC3 acts with TCF7L1 as a co-repressor. There is, however, no evidence of this interaction occurring in published studies. Notably, there is no single method that is universally used in molecular biology to demonstrate protein-protein interactions. Each available technique offer advantages over others, therefore the use of orthogonal repeats provides greater support of the interaction. As such, four different methods were used to investigate the interaction between ZIC3 and TCF7L1: co-immunoprecipitation, yeast two-hybrid, proximity ligation assay and BiFC. Overall, I have shown through multiple molecular assays that ZIC3 is capable of interacting with TCF7L1, in a context specific manner, when overexpressed in cell lines, consistent with the role of ZIC3 as a canonical Wnt co-repressor during gastrulation and providing a precursor for this interaction occurring *in vivo*.

3.3.1 ZIC3 and TCF7L1 demonstrate a positive interaction in multiple molecular assays

Protein-protein interactions are commonly investigated using biochemical co-immunoprecipitation experiments, however the attempts presented in this thesis could not confirm an interaction via this method. A co-immunoprecipitation between ZIC3 and TCF7L1 was successfully conducted by our collaborators, indicating the two proteins can physically interact, however data was not made available in time for this thesis submission. There are several alternative molecular techniques such as proximity assays that are widely used and increasing in popularity, which I have used to show a positive interaction between ZIC3 and TCF7L1. For example, a PLA successfully demonstrated ZIC3 and TCF7L1 interact following transient transfection in HEK293T cells. Ideally, this assay would be performed on endogenous proteins, however due to the lack of reliable ZIC3 primary antibodies, coupled with the low expression of ZIC3 in HEK293T cells (Genevestigator, Appendix B), overexpression studies were utilised for this primary experiment. This assay will be useful to test future ZIC3-TCF7L1 interaction studies, particularly when screening for mutations that disrupt this protein interaction. Furthermore, a PLA experiment could be performed on embryo sections to demonstrate the *in vivo* interaction of these two proteins following production and optimisation of suitable ZIC3 antibodies. Alternatively, to aid in protein detection, a cell line or mouse model can be created in which a V5 epitope tag is introduced (via CRISPR/Cas9 mutagenesis) into endogenously expressed ZIC3 and a different tag (such as FLAG) can be introduced into endogenous TCF7L1. This would allow the use of the reliable V5/FLAG-antibodies in downstream experiments.

BiFC analysis was also able to confirm the interaction between ZIC3 and TCF7L1 when over expressed in HEK293T cells. BiFC assays offer a significant advantage over co-immunoprecipitation or yeast two-hybrid experiments as quantitative data can be produced easily. Indeed, signal emitted from interacting proteins can be imaged with a fluorescent microscope to visualise the interaction at a cellular level, or with a fluorescent plate reader to determine the average fluorescent intensity from a large collection of cells (Kerppola 2009). The latter method can evaluate vastly greater cell numbers than would be able to be individually photographed and scored. Similar to the PLA, BiFC will be useful for screening mutations that disrupt the ZIC3-TCF7L1 interaction and quantitative measurements will be particularly useful for identifying hypomorphic variants.

Overexpression studies are often criticised for not accurately representing the physiological environment in which they are being tested. To address this, a suite of single-molecule tracking techniques are being developed to visualise the movement and molecular dynamics of transcription factors. Single-molecule tracking utilises tags such as EGFP, mEOS or HaloTag and, coupled with the use of high resolution microscopy and robust computational and statistical algorithms such as pair correlation analysis, allows the researcher to track movement and interactions of single proteins in live cells (Hinde et al. 2010; Presman et al. 2017). This technique is considered suitable to assess molecular mechanisms using overexpressed protein while overcoming potential artefacts associated with exogenous protein studies (Hinde et al. 2016). As such, it could be used to validate the protein interactions between ZIC3 and TCF7L1.

3.3.2 Yeast two-hybrid assays showed weak, but not absent, interaction between ZIC3 and TCF7L1

Yeast two-hybrid experiments are often used to screen for novel protein interactions and can confirm protein-protein interactions in living yeast cells. ZIC3 and TCF7L1 proteins were assessed for their ability to interact using a yeast two-hybrid experiment, however, data produced from the assay did not detect a convincing interaction between the two proteins. Performing this protein interaction assay in isolation may suggest that the proteins are not binding partners or that there is only a relatively weak protein interaction occurring. Indeed, the importance of orthogonal repeats is emphasised as a positive interaction between the two proteins was detected using other methods, and this technique likely produced a false negative result.

Positive interactions in yeast two-hybrid are dependent on a number of factors including protein expression in the yeast cells, correct protein folding and post-translational modifications needed for the interaction (Brückner et al. 2009). For example, the ZIC3 and TCF7L1 proteins may have low or absent protein expression in the heterologous system (i.e. mammalian protein expression in yeast). However due to the lack of epitope tags within the vector backbone of bait and prey constructs, and

the lack of ZIC specific antibodies, it was not possible investigate protein expression in this system. Additionally, expression of mammalian protein in yeast cells may not provide the necessary post translational modifications necessary for the proteins to interact. For example, it has been documented that ZIC proteins are subject to the post translational modification SUMOylation (Chen et al. 2013; Bellchambers 2015). It is possible that the proteins require this modification (or others) for normal interaction or expression, and yet are not subject to the modifications in the yeast cells resulting in a false negative interaction. Lastly, the configuration of the protein fusion (i.e. the attachment of the GAL4 DNA binding or activation domain to the target protein) may physically block the interaction binding site of the target protein, thereby unintentionally inhibiting reporter gene activation. Indeed, to minimise this artefact, both proteins were made as both bait and prey proteins using the ProQuest vectors. There were observable differences between the two confirmations, with confirmation 1 (ZIC3 as the bait and TCF7L1 as the prey) having a slightly greater interaction (Figure 3.5), confirming the arrangement of the fusion proteins can influence the strength of detecting protein interactions. The ProQuest vectors, however, attach the GAL4DNA binding or activating domain to the N-terminal of the target proteins. Future studies could investigate whether attaching the GAL4 domains to the C- terminus of the target proteins influences the strength of the protein interaction or include additional positive control plasmids such as ZIC2 and TCF7L2. Overall, the ProQuest yeast two-hybrid method was not deemed appropriate to measure the protein interaction between ZIC3 and TCF7L1.

3.3.3 Data lends support to ZIC3 acting as a Wnt co-repressor.

Negative regulation of the Wnt signalling pathway is essential for normal embryo development. It is suggested that one role of the ZIC3 transcription factor is to function as a co-repressor of Wnt/β-catenin mediated transcription. I have successfully confirmed that ZIC3 and TCF7L1 are capable of direct protein-protein interaction supporting a molecular mechanism of how ZIC3 and TCF7L1 can function together to inhibit canonical Wnt signalling during specific stages of development (Figure 3.12). Given that these proteins are capable of interacting *in vitro*, it is now practical to investigate this interaction *in vivo*. Furthermore, *in vivo* analysis of a variant in which the ZIC3-TCF7L1 binding is lost would also provide evidence of the functional consequence of this interaction.
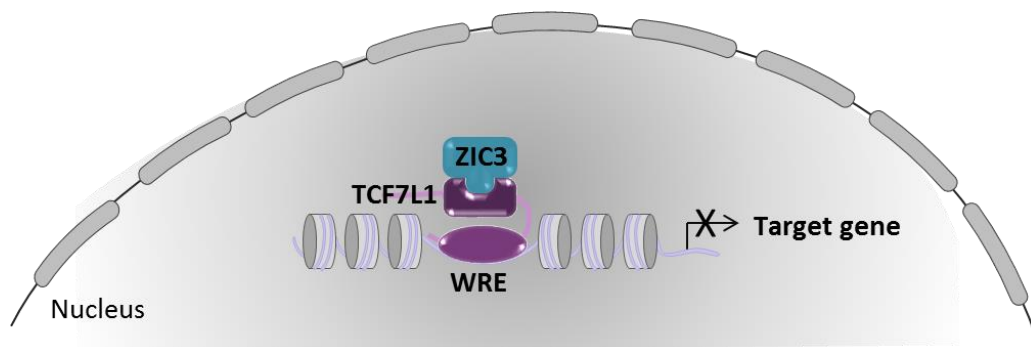
Figure 3.12: Simplified model of ZIC3-TCF7L1 dependent Wnt repression. ZIC3 can physically interact with a well characterised repressor of canonical Wnt signalling, TCF7L1. Both *Zic3* and *Tcf7l1* mutate to give similar phenotypes representing an elevated Wnt phenotype. As such, it is proposed ZIC3 directly interacts with TCF7L1 during gastrulation to prevent transcription of Wnt target genes.

Chapter 4: Do ZIC3 Heterotaxy-associated variant proteins have altered Wnt pathway function?

4.1 Introduction

ZIC3 is commonly thought to function as a classical transcription factor to influence transcription of target genes and can activate several reporters in cell-based assays (Chhin et al. 2007; Li et al. 2018). It is becoming apparent, however, that researchers must also consider the function of ZIC3 in the Wnt pathway. As discussed in Chapter 1 (section 1.5.4.2), two independent studies using both animal models (*Xenopus* and Zebrafish) and cell based reporter assays demonstrated ZIC proteins can function as canonical Wnt co-repressors (Pourebrahim et al. 2011; Fujimi et al. 2012). The Arkell laboratory has also demonstrated that all five human ZIC proteins (ZIC1-5) inhibit Wnt signalling in a HEK293T cell-based TOPflash assay (Chapter 1, Figure 1.7) and shown that mouse models of *Zic3* loss-of-function not only present with L-R axis defects, but also phenocopy characteristic elevated canonical Wnt phenotypes, including anterior truncation and axis duplication (Arkell Laboratory, unpublished). Furthermore, both *Zic3* and *Tcf7l1* exhibit overlapping expression domains in gastrulating mouse embryos, and I have demonstrated that these two proteins can physically interact in various cell-based assays (Chapter 3), suggesting ZIC3 transects the Wnt pathway via TCF7L1. Based on this evidence, if ZIC3 mediated repression of Wnt activity is required to prevent Heterotaxy in human probands, it would then be expected that ZIC3 Heterotaxy-associated protein variants have altered Wnt inhibition activity. Currently, however, the functional consequence of the reported human *ZIC3* variants in the Wnt pathway is not known.

4.1.1 Characterising genetic variants as pathogenic

Advances in DNA sequencing technologies have fuelled a vast expansion in the number of genetic variants being identified in human patients. In the case of *ZIC3*, more than 40 allelic variants have been associated with human Heterotaxy thus far. It is, however, difficult to determine *a priori* whether sequence variants found in probands are causative or are simply naturally occurring sequence variation present in the population (i.e. a polymorphism). Studies investigating previously reported disease causing variants (from other genes in the genome) established that many of the variants were either common polymorphisms, present in population control samples or did not have sufficient evidence to be classified as causative (Bell et al. 2011; Norton et al. 2012; Xue et al. 2012). This distinction is essential to functional genomic research as falsely assigning causation to a genetic variant can have negative consequences regarding downstream analysis, genomic research and medical treatments (Macarthur et al. 2014). One issue caused by the increase in variant identification is the inconsistent use of terms used to describe genetic variants. Macarthur et al. 2014 have outlined

definitions of commonly used terms (outlined in Table 4.1) and these will be used in this thesis. Of particular importance is the idea of pathogenicity regarding human variants. A variant is termed pathogenic only if there is strong evidence that it contributes mechanistically to the disease. This evidence is often a combination of family history, allelic frequency in the population, bioinformatics approaches and *in vitro* or *in vivo* functional studies demonstrating a damaging effect on the gene or gene product.

Computational (*in silico*) online prediction tools including Polyphen, SIFT, Align GVGD and Mutation Taster have been used to predict the effect on protein function of all previously identified human Heterotaxy-associated *ZIC3* missense mutations (Paulussen et al. 2016). These commonly used tools employ algorithms (which can differ between each program) to determine the effect of a particular amino acid change to overall protein function. All 16 *ZIC3* missense mutations so far assessed are predicted to be either damaging, deleterious or disease causing (Paulussen et al. 2016). Online tools such as these have an accuracy of 65-80% when predicting protein function (Thusberg et al. 2011) but cannot be used as the only evidence to determine whether a sequence variant is pathogenic (Richards et al. 2015). Instead they should be supported with rigorous experimental evidence to demonstrate causality. Currently there are gaps in our understanding of *ZIC3* variants associated with Heterotaxy, particularly with respect to functional studies. For example, the functional effects of only 3/42 known Heterotaxy-associated *ZIC3* variants have been investigated via an *in vivo* model organism (Paulussen et al. 2016).

Functional assays are widely used to assess protein function when determining whether a genetic variant is disease causing. These assays, however, vary in their supportive strength depending upon how well they represent the biological environment (Richards et al. 2015). The strongest support comes from animal models that directly mimic the variant. Such models (called avatars) are increasingly popular due to the relative ease of introducing precise genome alterations via CRISPR technology (Adli 2018; Gurumurthy and Lloyd 2019). It is, however, still impractical to produce and analyse all human variants via the avatar approach. Another method that lends strong support is the analysis of protein function from patient tissue. This often requires primary cells or the reprogramming of post-natal/adult cells. In some cases, CRISPR technology can be used to correct the mutation in the cell line and produce isogenic control cells. This method, however, also suffers the disadvantage of requiring a large amount of work for each mutant. Additionally, due to limitations regarding access to prospectively collected patient tissue, other methods generally must be adopted when studying rare congenital disorders. In regards to ZIC protein function, overexpression assays in easily transfectable and immortalised cell lines (eg HEK, NIH3T3 etc) are commonly used to overcome

Table 4.1: Definitions of terms used to describe sequence variants. *The following terms have been taken in their entirety from Macarthur et al 2014.*

Term	Definition
Pathogenic	Contributes mechanistically to disease, but is not necessarily fully penetrant (i.e., may not be sufficient in isolation to cause disease).
Implicated	Possesses evidence consistent with a pathogenic role, with a defined level of confidence.
Associated	Significantly enriched in disease cases compared to matched controls.
Damaging	alters the normal levels or biochemical function of a gene or gene product
Deleterious	Reduces the reproductive fitness of carriers, and would thus be targeted by purifying natural selection.

these problems. It must be noted, however, that these cell lines may not express all the potential co-factors present in individual tissues where ZIC3 functions *in vivo*.

4.1.2 Cell-based assays for assessing ZIC3 function

TOPflash assays are routinely used to measure TCF-β-catenin mediated transcription (Korinek et al. 1997). Of the Heterotaxy-associated *ZIC3* variants identified so far, only two mutant proteins (W255G and P217A) have been assayed for their Wnt repressive function through the use of a *Xenopus* TOPflash assay (Fujimi et al. 2012). ZIC3-W255G, when tested had reduced repressive function, whilst ZIC3-P217A repressed transcription as strongly as wildtype ZIC3. It is unknown whether any of the other *ZIC3* Heterotaxy-associated mutations can still inhibit Wnt signalling. To investigate this, a TOPflash reporter construct which consists of 3 tandem copies of a TCF binding motif (AGA TCA AAG GG) upstream of a β-globin minimal promoter and a luciferase CDS was created by the Arkell laboratory (Helen Bellchambers). This reporter can be transfected into HEK293T cells along with an expression construct for β-catenin and ZIC3. The exogenous β-catenin protein interacts with endogenous TCF proteins to drive luciferase expression from the reporter, which can then be measured. Wildtype ZIC3 has been shown to inhibit this β-catenin-mediated transcription (Chapter 1, Figure 1.7).

As this assay is optimised within the Arkell laboratory, it will be used to investigate whether *ZIC3* Heterotaxy associated variants affect ZIC3 dependent Wnt repression. Importantly, published studies have primarily focused on the transcription factor role of ZIC3 when investigating how mutation of *ZIC3* alters protein function. As the ZIC proteins are capable of both transcription factor and co-factor functions, it remains possible that the Heterotaxy phenotype in the human patients arises due to a loss of ZIC3's role as a classical transcription factor. Alternatively, the pathology associated with these variants could arise due to the loss of both of these functions or due to a separate unidentified mechanism. Therefore, both functions of ZIC3 (co-factor and transcription factor) must be investigated.

Transactivation assays are a useful tool when analysing the effect of a transcription factor on a specific DNA element due to their high sensitivity and ease of signal detection. These assays generally contain a DNA motif that can be bound by the protein of interest cloned upstream of a reporter protein coding sequence (CDS) such as green fluorescent protein (GFP), *lacZ* or luciferase (Nierode et al. 2016). Of the human *ZIC3* Heterotaxy-associated variants that have been studied in cell-based assays, many had reduced ability to initiate transcription (Ware et al. 2004; Chhin et al. 2007; D'Alessandro et al. 2013b; Cowan et al. 2014). However, experimental conditions (such as reporter and expression construct concentrations) were not always reported and cell types varied between experiments (Ware et al.

2004; D'Alessandro et al. 2013b; Cowan et al. 2014; Li et al. 2018). Moreover, many transactivation studies investigating ZIC3 function use either a reporter containing the herpes simplex thymidine kinase (TK) or the simian virus 40 (SV40) promoter, yet there is no evidence these are true ZIC responsive elements (ZREs) *in vivo*.

To accurately study ZIC transactivation ability, target DNA sequences must be known. Two independent studies have attempted to find such motifs and identified the same consensus sequence, utilising different experimental techniques. The first was a universal protein binding microarray (PBM), which was used to identify DNA sequences to which different transcription factors can bind in an *in vitro* study (Badis et al. 2009). In brief, 109 known mouse transcription factors were purified and tagged with a detectable epitope. In the case of the ZIC proteins investigated, only fragments containing the ZFD were purified. Computational analysis was conducted using an algorithm to determine the nucleotide sequence that best interacts with the protein and then generates a variety of 8-10 nucleotide sequences (motifs) to determine which sequence combination is most ideal for protein binding. This experiment identified a motif able to be bound by purified ZIC1, ZIC2 and ZIC3 ZFD (Badis et al. 2009). The motif, shown in Figure 4.1, is highly similar to that identified in an independent experiment with ZIC3 (Lim et al. 2010). These researchers performed a ChIP-chip experiment on mouse embryonic stem cells and found three sequences to be overrepresented in the presence of ZIC3, one of which showed high similarity to that identified in the PBM (Figure 4.1). Using this sequence, a highly reliable and robust ZIC reporter assay was developed in the Arkell laboratory (Ahmed 2015). The reporter construct contains a short stretch of DNA with six tandem copies of the consensus motif (with the same five bp spacer between each motif) and placed upstream of a β-globin minimal promoter and a luciferase CDS (the final construct is known as p[ZM2]6-β-gl-luc2; see Section 2.1.1).

4.1.3 ZIC3-Heterotaxy associated variants selected for functional analysis

Currently, there are 42 mutations (described in Chapter 1, section 1.3) associated with Heterotaxy in humans (Gebbia et al. 1997; Mégarbané et al. 2000; Ware et al. 2004; Chhin et al. 2007; Wessels et al. 2010; De Luca et al. 2010; Ma et al. 2012; D'Alessandro et al. 2013a, b; Cowan et al. 2014; Paulussen et al. 2016; El Malti et al. 2016; Li et al. 2018). The mutational spectrum (summarised in Figure 4.2), as well as current knowledge, all point to the importance of the ZFD, and so mutations within this domain were prioritised for functional analysis, namely nonsense mutations and missense mutations affecting key residues (such as cysteine or histidine) that make up individual zinc fingers.
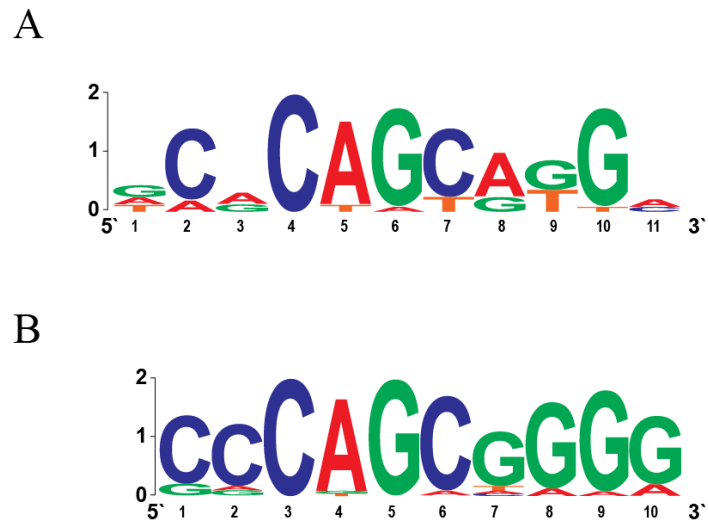
A

B

Figure 4.1 Putative ZIC3 DNA-binding sites. (A) Sequence identified by Badis et al. 2009 after performing a universal PBM. **(B)** Sequence identified by Lim *et al.,* (2010) after performing ChIP-chip on mouse embryonic stem cells.
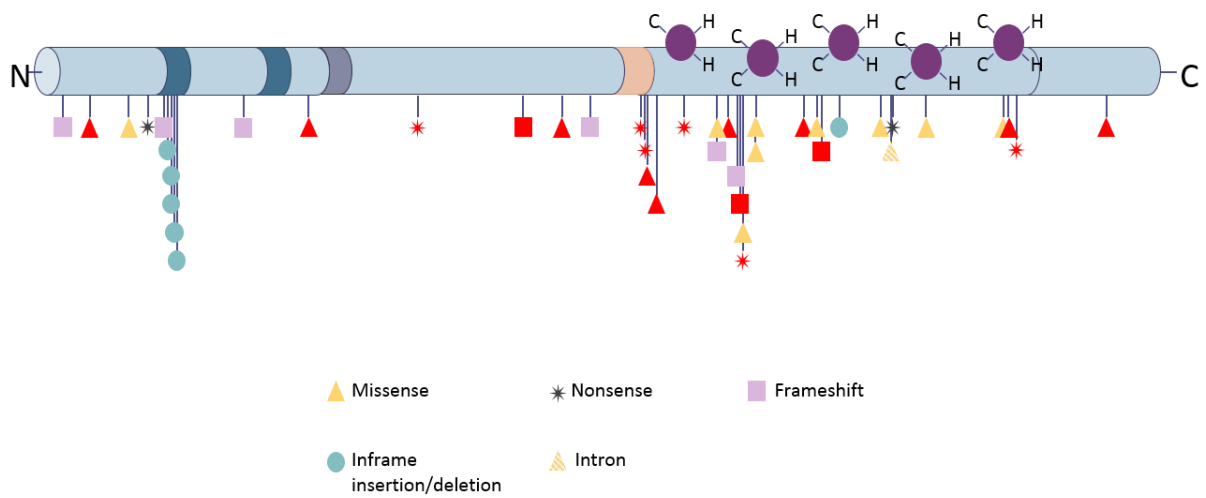
Figure 4.2: Mutation of ZIC3 is associated with human cases of Heterotaxy. Currently there are 42 known human mutations in ZIC3 that are associated with human Heterotaxy. These include 18 (42.9 %) missense, 9 (21.4%) frameshift, 8 (19%) nonsense, 6 in-frame expansions or deletions and 1 intron mutation (16.6%). Of these mutations, 22 (52%) occur within the highly conserved zinc finger domain of the ZIC3 protein. Experimental analysis is needed to determine if all 42 mutations are pathogenic. Seventeen variants (red) were selected for functional analysis and included missense and truncating variants both within and outside the zinc finger domain. C: Cysteine, H: Histidine, N': N-terminus, C': C-terminus.

Generally, it would be expected that nonsense and frameshift variants that produce a premature termination codon (PTC) would be subject to the cell surveillance mechanism known as nonsense mediated decay (NMD). The NMD pathway serves to remove mRNA transcripts containing PTCs before they are translated to avoid truncated proteins having damaging effects to the organism (Brogna and Wen 2009). The efficiency of NMD is however extremely variable (between tissues and developmental stages) and the factors that determine success not fully understood (Huang and Wilkinson 2012). One factor that is important in NMD is the splicing events that occur at each intron/exon boundary. NMD is typically triggered in transcripts containing nonsense mutations if the PTC is situated 50-55 nucleotides upstream of the last exon-exon junction, and therefore would presumably affect the nonsense mutations investigated here (Maquat 2005). There is evidence, however, of a *Zic3* PTC variant that evades NMD. The Arkell laboratory identified a *Zic3* nonsense mutation (truncated before the ZFD; 1283G >T, E249X) that produces a transcript that is not subjected to NMD during murine embryogenesis and instead generates a truncated ZIC3 protein (Ahmed et al. 2013). Therefore, it is also possible that some, if not all, of the human *ZIC3* PTC variants also evade NMD and produce detectable protein *in vivo*. As NMD relies on the presence of introns to determine whether the protein is degraded, human *ZIC3* variants were introduced into *ZIC3* cDNA (which lacks introns), to ensure the transcript is not subject to NMD. This allows for protein function to be assessed under the assumption the protein is made *in vivo*. Nonsense mutations that create an N-terminal PTC (<150 amino acids) were excluded from analysis due to predicted instability. For example, the ZIC3-S43X variant would generate a protein <50 amino acids in length; previous subcellular localisation assays could not detect the truncated protein after transfection of the ZIC3-variant plasmids into HeLa cells (Ware et al. 2004). Of the 42 known *ZIC3* Heterotaxy associated variants, 17 were selected for study (Figure 4.2, Table 4.2).

4.1.4 Chapter Aims

The overall aim of this Chapter is to investigate the functional consequences of 17 Heterotaxy-associated *ZIC3* variants to determine if the mutations affect the ability of ZIC3 to repress Wnt activity *in vitro*. Additionally, as ZIC3 can function as a classical transcription factor, it is possible this function is altered by the mutations. As such, the *ZIC3* variants will also be assessed for their ability to activate transcription at a ZRE. Specifically, this chapter aims to screen human variants to determine whether they alter:

i. ZIC3 dependent Wnt repression (measured by TOPflash assay)

ii. The interaction with TCF7L1 through a Bimolecular Fluorescence Complementation (BiFC) assay

iii. Transactivation at a ZRE using a reliable and robust ZIC responsive assay

Table 4.2: Characteristics of selected *ZIC3* Heterotaxy-associated mutations. *Clinical symptoms: ASD: atrial septal defect, CAVC: complete atrioventricular canal, CHD: congenital heart disease, DORV: double outlet right ventricle, d-TGA: dextro-transposition of the great arteries, DILV: double inlet left ventricle, HLHS: hypoplastic left heart syndrome, I-TGA: levo-transposition of the great arteries, LV: left ventricle, MV: mitral valve, PA: pulmonary atresia, PS: pulmonic stenosis, TA: tricuspid atresia, TAPVR: total anomalous pulmonary venous return, VSD: ventricular septal defect. In silico prediction: D: probably Damaging or B: Benign as tested by Polyphen, De: Deleterious or T: Tolerated as tested by SIFT, DC: Disease causing or P: polymorphism as tested by Mutation Taster*

Amino acid alteration	Inheritance	Clinical symptoms	In silico pathogenicity prediction
G17C	Isolated, familial	Isolated CHD/Heterotaxy, HLHS**,** PS, l-TGA**,** DILV, ASD, DORV, Heterotaxy, dextrocardia, d-TGA, TA	D/De/DC
S109C	Sporadic	Isolated CHD, d-TGA, PDA	D/De/P
E155X	Sporadic	Heterotaxy, Midline liver, dextrocardia, CAVC, PA, DORV, TAPVR	
Y199delfsX19	Sporadic	Heterotaxy, dextrocardia, VSD, ventriculomegaly, no stomach	
P217A	Sporadic, familial	Heterotaxy, Hypoplastic LV and MV, d-TGA, DORV, VSD, ventricular inversion, Heterotaxy, gut malrotation, polysplenia, midline liver	B/T/DC
Q249X	Familial	Single ventricle, PS, DORV, asplenia, horseshoe kidney, low-set ears, abnormal situs	
S252X	Familial	Unaffected carrier mother, Asplenia, CAVC, PA	
C253S	Familial	HLHS, right atrial isomerism, TGA, aplenia, malrotation webbed neck, abnormal liver lobation	D/De/DC
W255G	Familial	d-TGA, single ventricle, PA, polysplenia	D/De/DC
C268X	Familial	Situs ambiguous, sacral agenesis	
H286R	Unknown	Unknown	D/De/DC
E291GfsX53	Familial	Heterotaxy, Abdominal situs inversus, asplenia, CAVC, PA, d-TGA	
Q292X	Familial	Two male brothers with situs ambiguous and anal malformations	
H318N	Familial	Heterotaxy, VSD, L hip dysplasia	D/T/DC
E325Lfs*84	Familial	Two male deaths with Heterotaxy	
K408X	Familial	TGA	
A447G	Sporadic	Isolated CHD, PA, VSD, ASD	D/De/DC

4.2 Results

4.2.1 Generation of ZIC3 variant expression constructs

The 17 human Heterotaxy-associated mutations selected for analysis (Figure 4.2, Table 4.2) included 6 PTC mutations (4 of which were constructed previously in the Arkell lab), 8 missense mutations and 3 frameshift mutations. Each corresponding human mutation was introduced into the pENTR-ZIC3-wt vector backbone via site-directed mutagenesis. For TOPflash and transactivation assays, all ZIC3 expression constructs were then introduced into the pcDNA3.1-nV5-DEST (V5-Dest) expression plasmid via a Gateway recombination reaction (Life Technologies), which attaches a V5 epitope tag to the N-terminus of the ZIC3 protein. The V5 tag allows for easy protein detection when overexpressed in mammalian cells due to readily available antibodies for the V5 epitope. Each mutant ZIC3 cDNA was also transferred to the V1-ORF destination plasmid for BiFC protein-protein interaction analysis. Overall, 34 ZIC3 expression plasmids were generated for functional analysis.

4.2.2 ZIC3 variants produce detectable protein that localise to the nucleus

Alteration to DNA sequences (missense, nonsense, frameshift mutations, etc.) may alter protein stability or subcellular localisation. For example, there are documented cases for both ZIC2 and ZIC3 in which transfection of mutant expression plasmids into cells lines does not result in the production of detectable protein, or which the protein has impaired nuclear localisation (Ware et al. 2004; Brown et al. 2005). Moreover, all documented ZIC functions are nuclear; in such cases where protein is not detected in the nucleus, it would not be practical to perform further functional studies. Each ZIC3 variant expression construct used in this chapter was transfected into HEK293T cells and subcellular localisation and stability assessed using standard immunofluorescence techniques. Past members of the Arkell laboratory assisted with the analysis of a subset of the immunofluorescence images. For each of the human ZIC3 variant proteins investigated, all produced detectable protein that localised to the nucleus, although not always to the same degree as the wildtype protein (Figure 4.3). Mutant proteins E155X, Y199delfsX19, Q249X, S252X, C253S, W255G, C268X, H286R, E291GfsX53, Q292X, H318N, G325Lfs*84, K408X exhibited significantly reduced ($p<0.05$ one-way ANOVA) nuclear localisation compared to wildtype ZIC3 protein, however mean nuclear localisation was always above 65%. Due to this, it was decided that all variant proteins would be further analysed via functional studies.
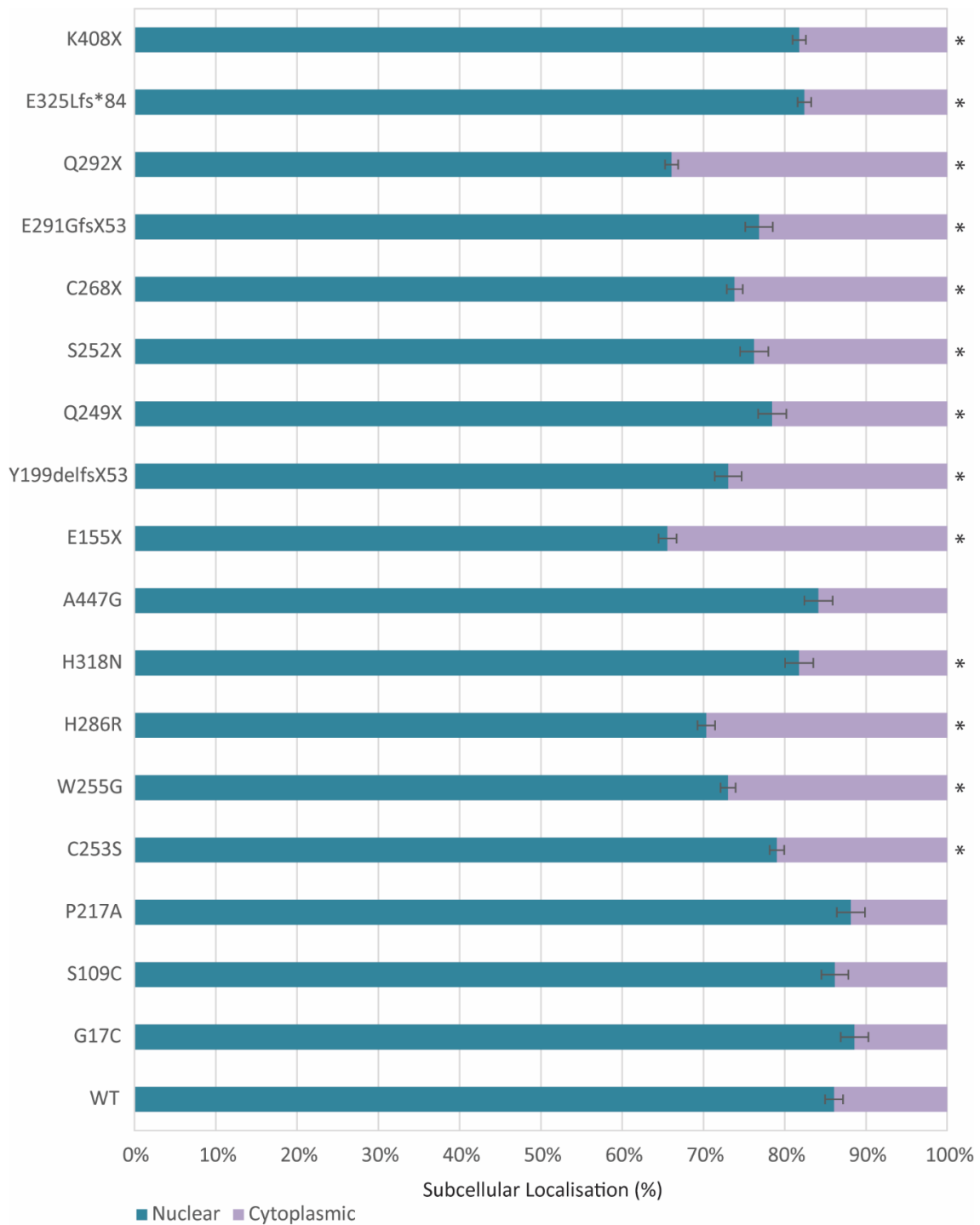
Figure 4.3: ZIC3 Heterotaxy-associated mutants localise to the nucleus. HEK293T cells were transfected with the V5-tagged ZIC3 expression constructs shown. 24 hours post-transfection, cells were prepared for immunofluorescence microscopy and stained with α-V5 antibody to detect the transfected V5-ZIC3 proteins, as well as α-LaminB1 to mark the nuclear membrane. Images of cells were captured with a fluorescent microscope. ImageJ software was used to determine the relative nuclear and cytoplasmic fluorescence intensity, which was converted to a percentage of total cell fluorescence. *: $p<0.05$; one-way ANOVA.

4.2.3 Human ZIC3 variants have altered Wnt repression ability

To test the ability of the ZIC3-variant proteins to repress β-catenin mediated transcription, a cell-based TOPflash reporter assay was used. HEK239T cells were co-transfected with V5-β-catenin, the TOP-β-gl-luc2 reporter construct and either V5-Dest (to detect background luciferase), V5-hZIC3-wt or a V5-hZIC3-variant expression construct. Cells were lysed and assessed for protein expression by SDS-PAGE and western blot, and luciferase activity was measured in a corresponding aliquot of cells 24 hours post-transfection. The data was expressed as relative luciferase activity (with respect to background luminescence). In each case, transfection of β-catenin in the absence of exogenous ZIC3 resulted in a ~25-30 fold increase in luciferase activity, whilst luciferase activity decreased to less than 10 fold with the addition of wildtype ZIC3, indicating that wildtype ZIC3 can repress transcription in this assay.

The ZIC3 variant proteins exhibited a range of repressive abilities. Variants affecting the ZIC3 ZFD, including both nonsense and frameshift mutations, had the greatest effect on Wnt inhibition. ZIC3 variants Q292X, K408X, E155X, E325LfsX84, Q249X, S252X and E291GfsX53 either lost all repressive ability or had severely reduced levels of repression (as indicated by high levels of luciferase expression) (p<0.05) (Figure 4.4 A-C). Comparatively, ZIC3 protein variants that alter critical amino acids within the first zinc finger (C253S, W255G, C268X) demonstrated decreased Wnt inhibition at varying degrees, indicated by increased luciferase values compared to wildtype protein. None of these proteins, however, completely lost the ability to repress Wnt signalling (Figure 4.4 A, C). Finally, missense mutations located outside the ZIC3 ZFD (G17S, S109C, P217A, A447G) were the only variant proteins, with the exception of H318N, that could repress Wnt signalling at a similar level to wildtype ZIC3 (p>0.05; Figure 4.4 D) suggesting these mutations will not result in an elevated Wnt phenotype *in vivo*. Overall, all ZIC3 variant proteins had varied ability to inhibit Wnt signalling, with missense mutations outside the ZFD retaining normal ZIC3 function.
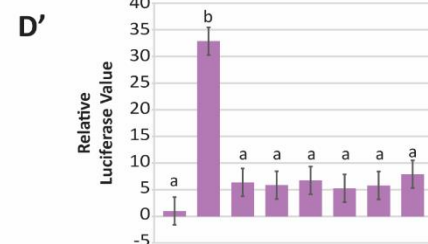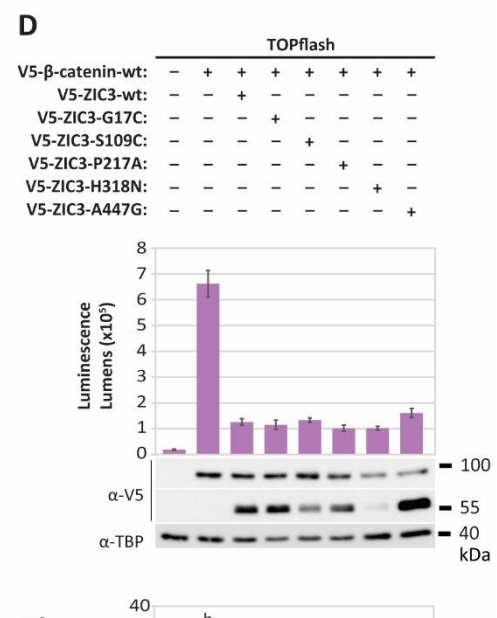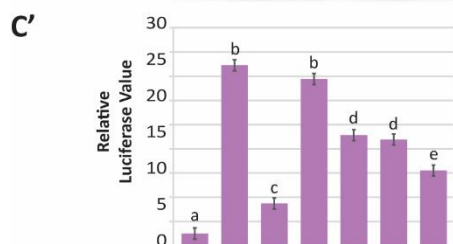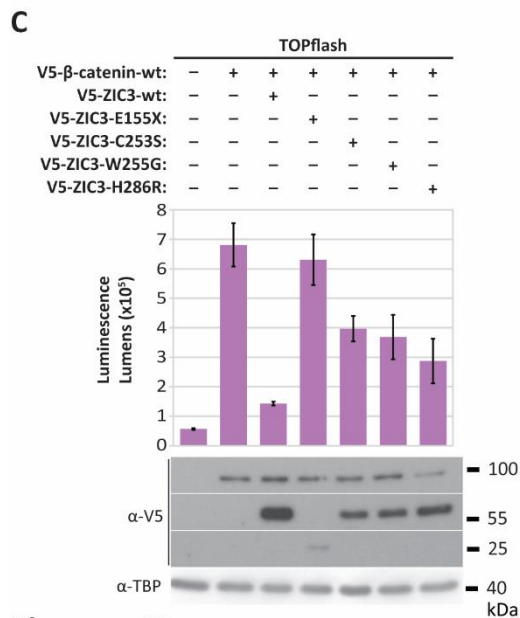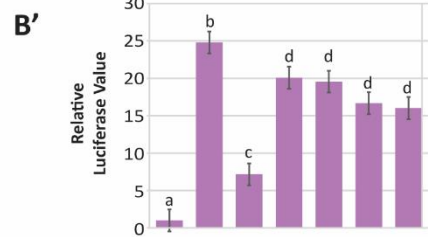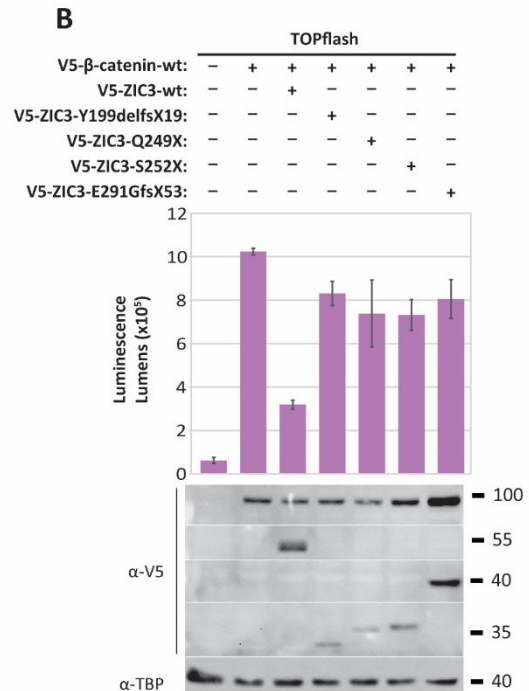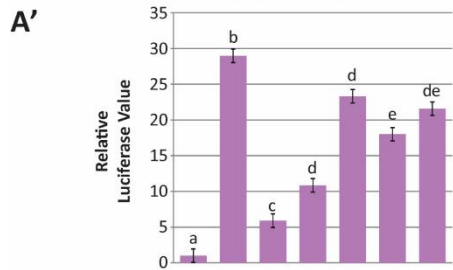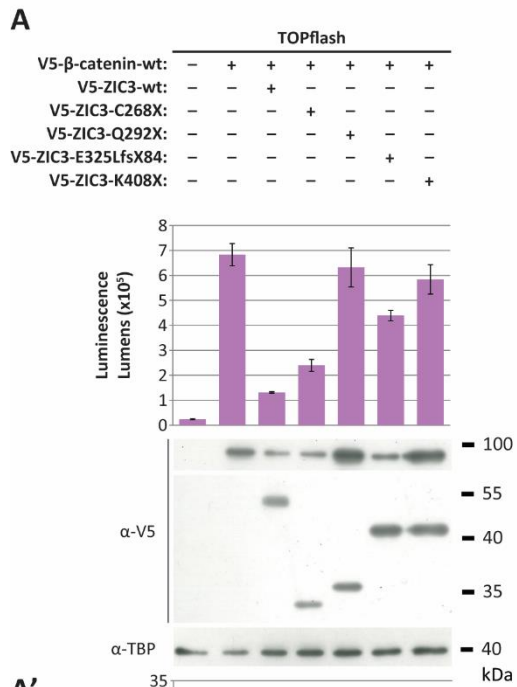
Figure 4.4: ZIC3 Heterotaxy-associated mutations have altered ability to repress Wnt signalling. HEK293T cells were transfected with a TOPflash reporter (TOP-β-gl-luc2), V5-tagged β-catenin-wt and a V5-tagged ZIC3 expression plasmid as shown. Luminescence was measured 24 hours post-transfection in each of three internal replicate samples. **(A-D)** One representative experiment with error bars representing the SD between the three internal replicates. A western blot was performed to confirm expression of transfected proteins with anti-TBP as a loading control. **(A'-D')** Pooled data from three independent repeats. Values depict the mean relative luciferase values. Error bars depict SEM. Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc.

4.2.4 Human ZIC3 variants have altered TCF7L1 interaction

The absent or relatively mild effect some ZIC3 missense mutant proteins exhibited on ZIC3 repressor activity in the TOPflash assay suggests that the ability of the variant proteins to interact with TCF7L1 may be intact. Conversely, missense mutations within the first zinc finger and several nonsense mutations displayed reduced repressor activity, suggesting either the ability of these proteins to interact with TCF proteins is affected, or the mutation affects the interaction with other repressor machinery involved in transcriptional repression with ZIC3.

To directly assess whether the ZIC3-variant proteins can still interact with TCF7L1, a BiFC assay was performed. For the BiFC assay, several ZIC3 variant expression constructs needed to be generated. Expression plasmid V1-hZIC3-wt was already available in the Arkell laboratory, whilst all 17 ZIC3 variant cDNAs were cloned into the V1-ORF expression plasmid via Gateway recombination technology (section 2.2.5). Variant ZIC3-H318N was not made until after the assay was conducted and was therefore excluded from downstream analysis.

To test the relative ability of the ZIC3-variant proteins to interact with TCF7L1 protein, HEK239T cells were co-transfected with V2-mTcf7l1 and either V1-hZIC3-wt or a V1-hZIC3-variant expression plasmid. Fluorescence intensity was measured 48 hours post-transfection and cells were lysed and assessed by SDS-PAGE and western blot (Figure 4.5). Wildtype ZIC3 resulted in approximately a 3-4 fold increase in fluorescence relative to background levels. All ZIC3 variants tested interacted with TCF7L1 either at similar levels to wildtype ZIC3, or displayed an increase in fluorescence intensity, suggesting an increased interaction between the two proteins. Unexpectedly, the truncation mutations resulted in significantly higher fluorescence compared to wildtype ZIC3. ZIC3-Q249X was the only mutant to have a significantly reduced, but not absent, interaction with TCF7L1 (Figure 4.5 B). Interestingly, Q249X and S252X differ by only 3 amino acids yet produce contrasting BiFC results: Q249X exhibits reduced fluorescence compared to wildtype ZIC3 whilst S252X produced increased fluorescence. This suggests that the ZIC3-TCF7L1 interaction site may encompass amino acids within this region.
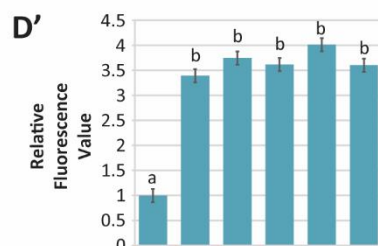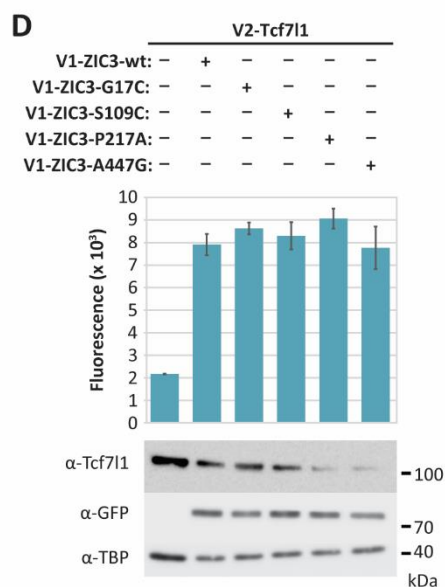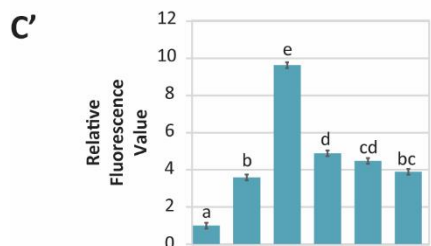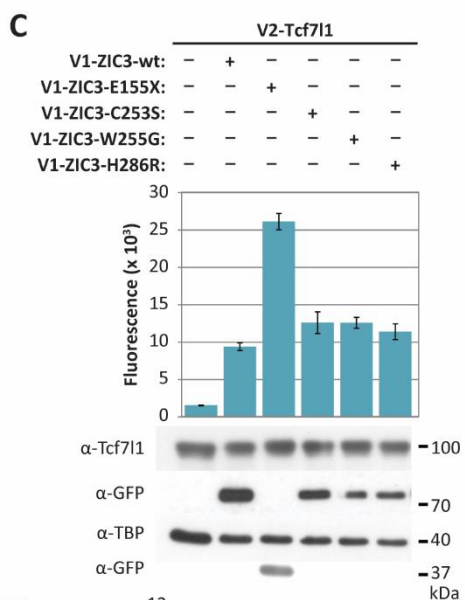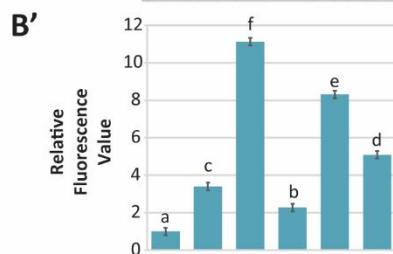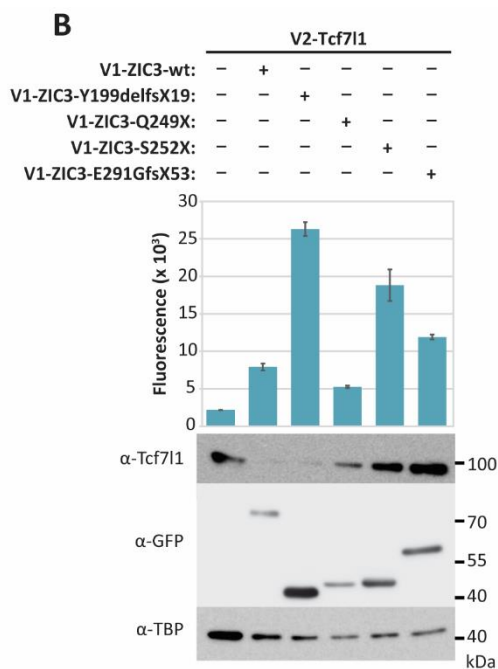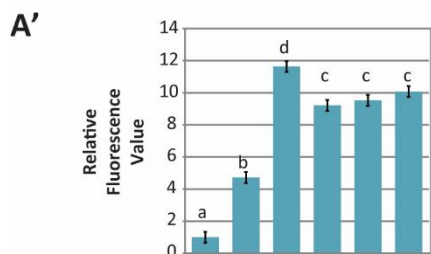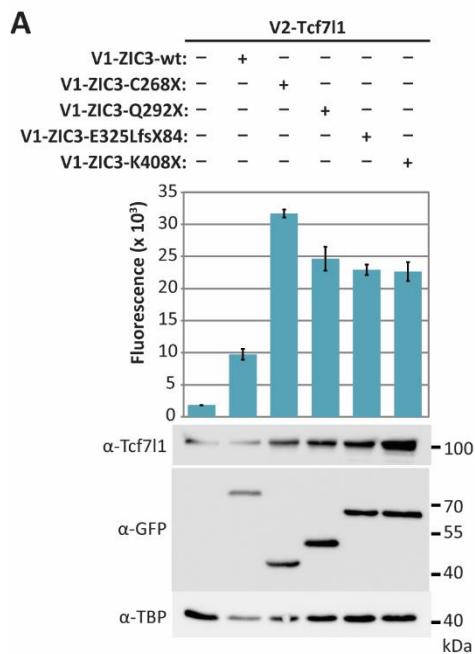
133

Figure 4.5: ZIC3 Heterotaxy-associated mutants have altered interaction with TCF7L1. HEK239T cells were co-transfected with V2-mTcf7l1 and either V1-hZIC3-wt or a V1-hZIC3-variant expression plasmid. The mean fluorescence intensity emitted from each of three internal replicate samples was detected 48 hours post-transfection. Three independent experiments were performed. **(A-D)** Mean fluorescence intensity of three internal replicates from one representative experiment. Error bars represent the SD between the three internal replicates. A western blot was performed to confirm expression of transfected proteins. The primary antibody α-Tcf7l1 was used to detect V2-Tcf7l1, α-GFP was used to detect V1-ZIC3 and α-TBP was used as a loading control. **(A'-D')** Pooled data from three independent repeats. Values depict the mean relative luciferase values. Error bars depict SEM. Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc.

4.2.5 Human ZIC3 variants have altered transactivation at a ZRE

Several of the human ZIC3 variant proteins were able to inhibit Wnt signalling at a level comparable to wildtype ZIC3. Similarly, all variant proteins were able to interact with TCF7L1 through a BiFC assay, suggesting the mutations do not disrupt this protein-protein interaction and that the Wnt repressive function of some of these mutant proteins is likely still intact. In addition to this co-factor (repressor) function of the ZIC proteins, another more common function associated with the ZICs is to act as a transcriptional activator. Studies have demonstrated that the transcription and co-factor function of the ZICs can be separated as shown by the ZIC2 mutation C270S (Brown et al. 2005; Pourebrahim et al. 2011). This mutation lies within the second cysteine of the fourth zinc finger of ZIC2 and can no longer bind DNA. Consequently, ZIC2 C270S protein cannot activate transcription of target genes, yet the same mutant protein is still able to repress Wnt signalling in a cell based TOPflash assay, indicating these two functions of the ZICs can, in some instances, be mutually exclusive (i.e. a mutation affecting DNA dependent transcriptional activation may not affect the repressive function though protein-protein interactions). It remains to be seen whether ZIC3s transcriptional activation ability is also impaired by the human Heterotaxy-associated mutations.

To assess the transcription factor function of the ZIC3 mutant proteins, a ZIC luciferase assay was utilised. HEK293T cells were co-transfected with the p(Z3M2)6-β-gl-luc2 reporter construct and either the V5-Dest vector (to detect background luciferase), V5-hZIC3-wt or a V5-hZIC3-variant expression construct. Luciferase activity was measured 24 hours post-transfection and cells were lysed and assessed by SDS-PAGE and western blot (Figure 4.6). Compared to wildtype ZIC3, the mutant proteins had variable abilities to transactivate the reporter. All truncated proteins failed to stimulate reporter transcription, confirming the necessity of the ZFD in DNA dependent transcription. A similar result was seen with each missense mutation within the first zinc finger (C253S, W255G and H286R). Missense mutations G17C and S109C were the only two mutations able to elicit transcription at the same level as the wildtype protein, whilst mutants P217A and A447G had mildly reduced transactivation ability (Figure 4.6 B, $p<0.05$).
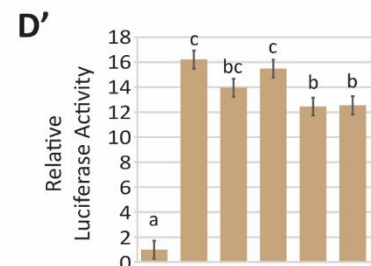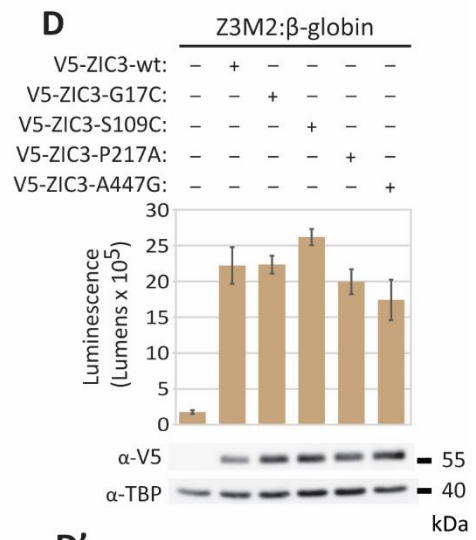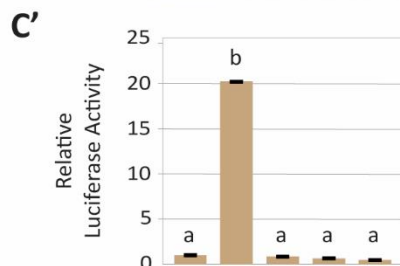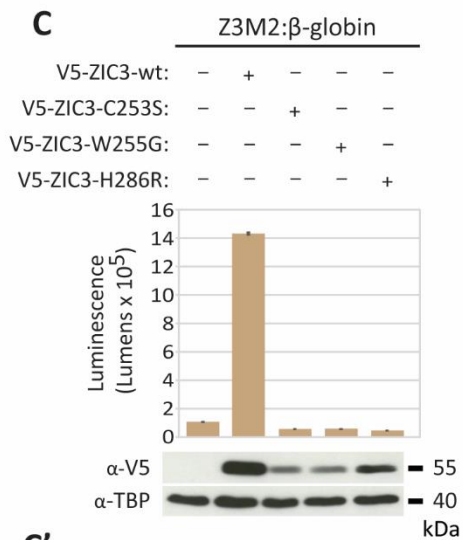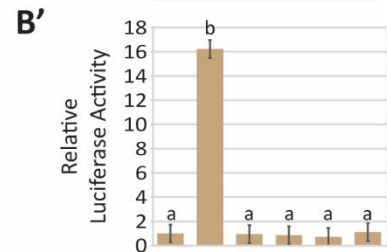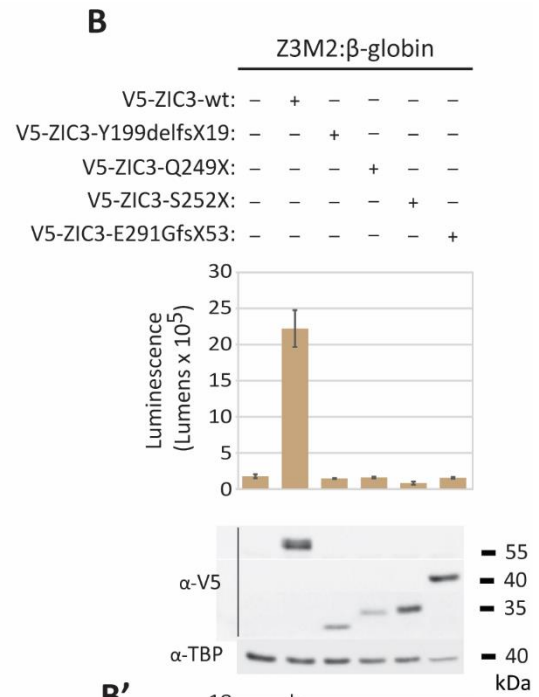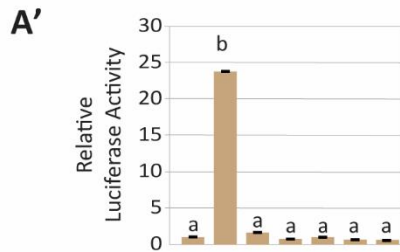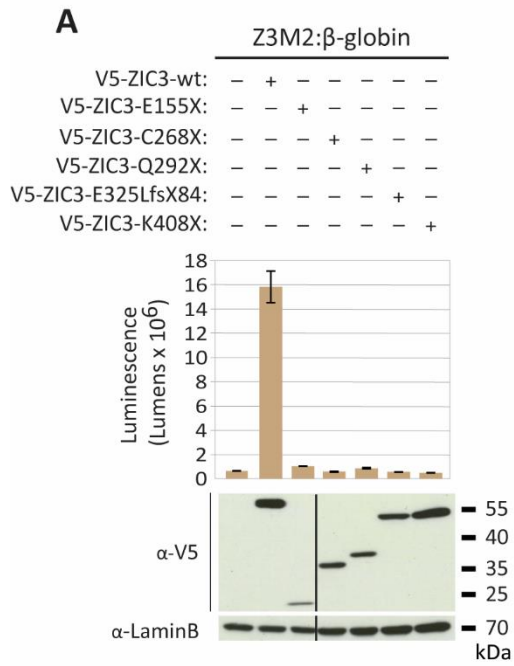
Figure 4.6: Heterotaxy-associated ZIC3 mutants have reduced ability to transactivate gene expression at a ZRE. HEK293T cells were transfected with the reporter construct p(Z3M2)6-β-gl-luc2 (containing a ZRE) and a V5-tagged ZIC3 expression plasmid as shown. Luminescence was measured 24 hours post-transfection in each of three internal replicate samples. **(A-D)** One representative experiment with error bars representing the SD between the three internal replicates. A western blot was performed to confirm expression of transfected proteins shown below. Anti V5 antibody was used to detect ZIC protein, while anti-TBP was used as a loading control. **(A'-D')** Pooled data from three independent repeats. Values depict the mean relative luciferase values. Error bars depict SEM. Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc. Assay from panel A and western blot from panel A and C performed by J. Ahmed, Arkell laboratory.

4.3 Discussion

Mutation of *ZIC3* is associated with Heterotaxy, a congenital abnormality that results in the abnormal arrangement of the internal thoraco-abdominal organs. There are currently 42 documented Heterotaxy-associated variants in human *ZIC3*. It is, however, unclear whether the reported variants are all pathogenic, and (if so) what the underlying molecular mechanism is. Given that ZIC3 has been shown to inhibit β-catenin mediated Wnt signalling in both *in vivo* and *in vitro* studies to prevent L-R axis defects (Fujimi et al. 2012; Cast et al. 2012; Ahmed et al. 2013) and that ZIC3 physically interacts with a known repressor of Wnt signalling, TCF7L1, when overexpressed in HEK293T cells (Chapter 3), it is possible that altered Wnt signalling underlies human cases of Heterotaxy. To test this hypothesis, 17 of the reported Heterotaxy-associated *ZIC3* variants were functionally assessed in cell based assays of Wnt activity and ZIC-TCF interaction. Additionally, to evaluate the alternative hypothesis (that *ZIC3*-associated Heterotaxy arises due to a loss of ZIC3 transactivation factor activity), each *ZIC3* variant was simultaneously assessed for the ability to transactivate gene expression at a ZRE. These analyses highlight the importance of the ZIC3 ZFD for both Wnt repression and transactivation. In most cases, mutation of *ZIC3* leads to severely reduced transactivation ability and reduced Wnt repression. Missense mutations within the ZFD and truncating ZIC3 variants had the greatest effect on ZIC3 function, while missense mutations outside the ZFD failed to substantially alter protein function, suggesting some identified variants may not be pathogenic. The data are consistent with altered Wnt signalling contributing to Heterotaxy, but do not rule out a critical role for ZIC3 as a classical transcription factor during establishment of the L-R embryonic axis.

4.3.1 Functional analysis corroborates and extends previous work

To date, only a subset of the 42 human *ZIC3* variants have been functionally assessed, primarily in cell based nuclear localisation and transactivation assays using overexpressed proteins. Analyses in these studies of the transcription factor function of ZIC3 alone correlates with ZIC3 predominately being considered a DNA binding transcriptional regulator and not a cofactor. As shown in Table 4.3, the results presented here largely align with previously published studies. For example, all of the proteins examined in this Chapter were able to accumulate in the nucleus at an appreciable level. Wildtype ZIC3 is 85% nuclear. Comparatively, 13 (75%) of the variant proteins tested have a significant decrease in nuclear localisation. However, in most cases this was a small decrease, with only two proteins (E155X; 65% and Q292X; 66%) falling below 70% nuclear localisation. In contrast to the results presented here, two of the proteins that alter the first two zinc fingers (W255G and C253S) failed to substantially accumulate in the nucleus when examined in previous publications using NIH3T3 cells (Chhin et al. 2007; Hatayama et al. 2008).

Table 4.3: Summary of ZIC3 functional analysis from published and current data. *In silico prediction: D: probably Damaging or B: Benign as tested by Polyphen, De: Deleterious or T: Tolerated as tested by SIFT, DC: Disease causing or P: polymorphism as tested by Mutation Taster.*

Amino acid alteration	In silico pathogenicity prediction	Nuclear localisation		Transactivation		Wnt inhibition (TOPflash)	
		Published results	Results from this thesis	Published results	Results from this thesis	Published results	Results from this thesis
G17C	D/De/DC	Normal (Cowan et al. 2014)	Normal (88%)	Normal (Cowan et al. 2014) / Normal (D'Alessandro et al. 2013b)	Normal		Normal
S109C	D/De/P	Small decrease (Cowan et al. 2014)	Normal (86%)	Normal (Cowan et al. 2014)	Normal		Normal
E155X		Large decrease (Cowan et al. 2014)	Decreased (65%)	Decreased (Cowan et al. 2014)	Ablated		Decreased
Y199delfsX19		Small decrease (Cowan et al. 2014)	Small decrease (73%)	Decreased (Cowan et al. 2014)	Ablated		Decreased
P217A	B/T/DC	Small decrease (Cowan et al. 2014) / Normal (Ware et al. 2004)	Normal (88%)	Normal (Cowan et al. 2014) / Increased (Ware et al. 2004) / Normal (D'Alessandro et al. 2013b)	Decreased	Normal (Fujimi et al. 2012)	Normal
Q249X		No protein detected (Ware et al. 2004)	Small decrease (78%)	Decreased (Ware et al. 2004)	Ablated		Decreased
S252X		Large decrease (Cowan et al. 2014)	Small decrease (76%)	Decreased (Cowan et al. 2014)	Ablated		Decreased

Amino acid alteration	In silico pathogenicity prediction	Nuclear localisation		Transactivation		Wnt inhibition (TOPflash)	
		Published results	Results from this thesis	Published results	Results from this thesis	Published results	Results from this thesis
C253S	D/De/DC	Large decrease (Chhin et al. 2007)	Small decrease (79%)	Decreased (Chhin et al. 2007)	Ablated		Decreased
		Small decrease (Ware et al. 2004)		Decreased (Ware et al. 2004)			
		Large decrease (Hatayama et al. 2008)		Decreased (D'Alessandro et al. 2013b)			
W255G	D/De/DC	Large decrease (Chhin et al. 2007)	Small decrease (73%)	Decreased (Chhin et al. 2007)	Ablated	Reduced Fujimi et al. 2012)	Decreased
		Large decrease (Hatayama et al. 2008)					
C268X		Large decrease (Ware et al. 2004)	Small decrease (74%)	Decreased Ware et al. 2004	Ablated		Decreased
H286R	D/De/DC	Large decrease (Chhin et al. 2007)	Small decrease (70%)	Decreased (Chhin et al. 2007)	Ablated		Decreased
		Large decrease (Ware et al. 2004)		Decreased (Ware et al. 2004)			
		Large decrease (Hatayama et al. 2008)					
E291GfsX53		Large decrease (Cowan et al. 2014)	Small decrease (77%)	Decreased (Cowan et al. 2014)	Ablated		Decreased
Q292X		Large decrease (Ware et al. 2004)	Decrease (66%)	Decreased (Ware et al. 2004)	Ablated		Decreased

Amino acid alteration	In silico pathogenicity prediction	Nuclear localisation		Transactivation		Wnt inhibition (TOPflash)	
		Published results	Results from this thesis	Published results	Results from this thesis	Published results	Results from this thesis
H318N	D/T/DC	Large decrease (Cowan et al. 2014)	Small decrease (82%)	Decreased (Cowan et al. 2014)	N/A		Normal
E325Kfs*84			Small decrease (82%)		Ablated		Decreased
K408X		Small decrease (Ware et al. 2004)	Small decrease (82%)	Decreased (Ware et al. 2004)	Ablated		Decreased
A447G	D/De/DC	Small decrease (Cowan et al. 2014)	Normal (84%)	Increased (Cowan et al. 2014)	Decreased		Normal

In another publication, however, Ware et al. 2004 found that the C253S variant protein did indeed accumulate in the nucleus to near wildtype levels, mirroring the results in this Chapter (Table 4.3). Given the differences in experimental protocols between each study (cell lines and level of overexpression), these proteins appear to have context dependent ability to accumulate within the nucleus. As the ZIC3 protein contains a cryptic NLS within the ZFD (Bedard et al. 2007; Hatayama et al. 2008), it may be expected that the PTC containing variants which produce proteins that lack the entire ZFD (such as E155X, Y199delfsX19, Q249X and S252X) would be unable to localise to the nucleus. Despite this, it was found that each of these variants were able to accumulate within the nucleus at appreciable levels (Figure 4.3). It is likely that (as for the murine *katun* variant protein [Ahmed et al. 2013]) these truncated proteins enter the nucleus via passive diffusion, since small proteins (less than ~40 kDa) are known to do so (Wei et al. 2003). Whether these proteins use passive transport could be assessed by attaching a tag (such as eGFP) that will artificially increase the size of the ZIC3 protein and prevent diffusion through the nuclear membrane.

The transactivation studies conducted here also confirm previous studies using either the herpes simplex thymidine kinase (TK) or the simian virus 40 (SV40) promoters (Koyabu et al. 2001a; Ware et al. 2004; Chhin et al. 2007; D'Alessandro et al. 2013a, b; Cowan et al. 2014). Although ZICs1-3 are able to stimulate transcription with either promoter (Mizugishi et al. 2001), neither promotor contains known ZIC DNA binding sequences nor are they considered *in vivo* ZIC protein targets. To test the ZIC3 variant proteins ability to transactivate a ZRE, the transactivation assays conducted here utilised a synthetic enhancer developed in the Arkell laboratory, containing a consensus DNA binding sequence for ZIC3. In all but one case, the transactivation ability of the ZIC3 variants aligned with previously results; variant A447G exhibited a small decrease in TA ability here, whereas in a previous study (Cowan et al. 2014) it exhibited increased transactivation function (Table 4.3 and Table 4.4).

Of the ZIC3 variants examined in this Chapter, only two had been previously examined for their ability to inhibit Wnt signalling (Table 4.3). Mutant proteins ZIC3-P217A and ZIC3-W255G were analysed by Fujimi et al. 2012 using a *Xenopus* TOPflash assay. The data presented here corroborates their finding that variant P217A is still able to repress reporter activation, while W255G has reduced repressor function. It should be noted, however, that the W255G variant failed to localise to the nucleus in the cell-based studies of Chhin et al. 2007 and Hatayama et al. 2008, and subcellular localisation was not assessed in the whole animal *Xenopus* TOPflash assay (Fujimi et al. 2012). In the cell-based assays presented here, the W255G protein did enter the nucleus but exhibited reduced inhibition of Wnt mediated transcription than the wildtype ZIC3 protein, suggesting that the decreased inhibition results from a change in the protein itself, rather than altered subcellular localisation. The tryptophan (W)

Table 4.4: Summary of functional properties of ZIC3 variants assessed in this Chapter

		Function assessed		
		Transcription factor	Co-factor	
		Transactivation	TOPflash	BiFC
ZIC3 variant	Empty vector	-	+++++	-
	wt	++++	+	++
	G17C	++++	+	++
	S109C	++++	+	++
	E155X	-	+++++	+++++
	Y199delfsX19	-	+++	+++++
	P217A	+++	+	++
	Q249X	-	+++	+
	S252X	-	+++	++++
	C253S	-	+++	+++
	W255G	-	+++	+++
	C268X	-	++	++++
	H286R	-	++	++
	E291Gfs53X	-	+++	+++
	Q292X	-	++++	++++
	H318N	n/a	+	n/a
	E325LfsX84	-	+++	++++
	K408X	-	++++	++++
	A447G	+++	+	++

[1] For Transactivation and BiFC, '-' denotes baseline activity and '+' to '+++++' indicates increasing activity, based on relative luminescence or fluorescence intensity within that assay.

[2] For TOPflash, '+' indicates an inhibitory effect compared to '+++++', which indicates no inhibition.

residue, affected in the W255G human variant, forms part of the $CWCH_2$ motif in the first zinc finger of the ZIC3 protein. Although the ZFD is made of five tandem zinc fingers, the first two zinc fingers of the ZIC proteins form a single structural unit and share a hydrophobic core (Hatayama and Aruga 2010). These zinc fingers don't follow the standard Cys-Xn-Cys-Xn-His-Xn-His (C_2H_2) pattern of zinc finger composition seen in zinc fingers 3-5. Instead, the first two zinc fingers follow the pattern Cys-X-Trp-Xn-Cys-Xn-His-Xn-His-Xn-Cys-X-Trp-Xn-Cys-Xn-His-Xn-His and hence termed the $CWCH_2$ motif. It has been suggested that these zinc fingers do not directly bind DNA, but facilitate DNA binding by the other zinc fingers and have a more important role in protein-protein interactions (Hatayama and Aruga 2010; Houtmeyers et al. 2013). In the experiments conducted here, the W255G variant entered the nucleus but was not able to stimulate any transcription from the ZRE reporter construct, could still bind TCF7L1 and lost some (but not all) Wnt inhibition ability (Table 4.3 and Table 4.4). Taken together, this variant exhibits similar properties to other ZFD missense mutants and the data does not confirm a unique role for zinc fingers 1 and 2 versus zinc fingers 3-5 of ZIC3.

4.3.2 Some reported ZIC3 variants do not appear pathogenic

All missense mutations outside the ZFD (ZIC3-G17C, -S109C, -P217A and -A447G) were able to inhibit Wnt signalling to the same degree as wildtype ZIC3 in the TOPflash assay. Similarly, the transactivation ability of variants G17C and S109C was indistinguishable from wildtype, whilst P217A and A447G showed a small but significant reduction in transactivation ability. The data suggests these four *ZIC3* variants are unlikely to be damaging mutations as they did not substantially alter ZIC protein function in the cell assays conducted.

ZIC3 variant G17C was identified in a screen for mutations in 10 genes associated with L-R defects (De Luca et al. 2010). Two patients from the same family were identified as heterozygous for the *ZIC3* G17C variant, however both patients also harboured a secondary mutation either affecting the *FOXH1* or *NKX2.5* genes. It was suggested the cardiac defects identified in these patients could be a result of either the second mutation alone or due to a compounded effect of both mutations. The same *ZIC3* variant was identified in two independent screens, and functional studies by both groups determined it did not alter transactivation compared to wildtype ZIC3 in overexpression reporter assays (D'Alessandro et al. 2013b; Cowan et al. 2014). I confirm and extend this data to show Wnt dependent inhibition is not altered, strongly suggesting this variant is neither damaging nor pathogenic.

ZIC3 variant P217A was identified by Ware et al. 2004 as a sporadic mutation in a male patient with a cardiac defect. The variant has since been identified in five other Heterotaxy patients via two different studies (D'Alessandro et al. 2013b; Cowan et al. 2014). Notably, this variant has appeared in various control groups, including that of D'Alessandro et al. 2013 and in the 1000 Genomes project with a frequency of 0.0082. Similar to the G17C variant, prediction software suggests a deleterious effect of

the mutation, however the functional analyisis presented here suggests the variant retains normal ZIC3 function. Moreover, functional studies by D'Alessandro et al. 2013 also demonstrated the mutation does not alter transactivation (in NIH3T3 cells), nor does it alter protein interaction of ZIC3 with GLI3 (Zhu et al. 2008).

Finally, variants S109C and A447G were identified by Cowan et al. 2014 in a screen of 440 unrelated indviduals with Heterotaxy. Published data shows that both mutants exhibit a mild reduction of nuclear localisation and A447G exhibits increased transactivation at a SV40 luciferase reporter (Table 4.3). Both mutations, however, displayed relatively minor functional deviation to wildtype ZIC3 in the assays described here and therefore there is only weak evidence to suggest these would be classified as damaging mutations.

4.3.3 All PTC variants appear pathogenic

It is generally predicted that transcripts containing a PTC (in a location that adheres to particular rules) will undergo NMD, thereby preventing accumulation of the truncated protein and ensuring a null allele. However, as previously stated, a murine *Zic3* nonsense mutation (*katun;* in which the PTC meets the rule for NMD) is known to evade NMD at gastrulation and instead is predicted to encode a stable, truncated protein which lacks both transcription factor and co-factor function (Ahmed et al. 2013). It is therefore possible that some of the human *ZIC3* nonsense alleles may also escape NMD. To understand whether the reported variants encode null alleles, each mutation was introduced into the *ZIC3* cDNA, a strategy which prevents targeting of the PTC containing transcript for NMD due to the absence of intron/exon junctions. Each of the PTC variants tested here lack most (C268X, Q292X, K408X, E325LeufsX84, E291GlyfsX53), if not all (Q249X, S252X), of the ZFD. Despite this, they encode stable and nuclear localised truncated proteins. Consistent with the lack of the ZFD, the transactivation ability of each protein is abolished. Based on results from the closely related ZIC2 protein (in which the ZFD and C-terminus are required for Wnt inhibition; Pourebrahim et al. 2011) it was predicted each truncated ZIC3 protein would also lose the ability to inhibit Wnt signalling in the TOPflash assay. It was, however, found that the majority of truncated ZIC3 proteins have reduced (but not ablated) ability to inhibit TCF-dependent transcription. The only *ZIC3* variant to lose all repressive function is ZIC3-E155X (Table 4.4) suggesting the first 155 amino acids of ZIC3 are not necessary for repression at WREs. Collectively, this analysis suggests that the *ZIC3* nonsense alleles are pathogenic, regardless of the extent of NMD. Additionally, the data suggest that the ZIC3 N-terminus contains regions that contribute to transcriptional repression at WREs. There are several conserved domains located in the N-terminal region of ZIC3 that have yet to be assigned a definitive function, including conserved regions such as the ZOC and ZFNC domains. Moreover, although located within the first 50 amino acids (and thus present in the E155X protein), low complexity regions such as the alanine repeat have

previously been associated with Heterotaxy in humans (Brown et al. 2005) and could possess TCF-dependent repressive function.

4.3.4 Many Heterotaxy-associated variants stabilised the ZIC-TCF interaction

In addition to the functional assays to examine transcription factor versus co-factor function, each variant protein was also assessed for interaction with TCF7L1. As can be seen in Table 4.4, a notable trend emerged from these experiments. Of the ten variant proteins that were substantially less able than wildtype ZIC3 to inhibit TCF-mediated transcription, nine exhibited increased relative fluorescence in a BiFC assay (indicating a stronger affinity for TCF7L1). This result was especially noteworthy because each of these variant proteins were either missing the ZFD or contained a missense mutation at a key ZFD residue, indicating the ZIC3 ZFD is dispensable for interaction. A previous study that mapped the ZIC2-TCF7L2 interacting domains showed the ZIC2 ZFD is necessary for this interaction (Pourebrahim et al. 2011). By analogy, it seemed likely that the highly homologous ZIC3 ZFD would also be essential for the ZIC3-TCF7L1 interaction. The increase in fluorescence seen in the BiFC assay conducted here could indicate several things. First, the ZIC3-TCF7L1 protein binding domain could potentially be located in the N-terminus of the ZIC3 protein, rather than the ZFD. Alternatively, the tertiary structure of the variant ZIC3 proteins could be altered in such a way that makes the TCF binding site on ZIC3 more accessible to the TCF7L1 compared to wildtype ZIC3 protein or allows the two Venus tags (V1 and V2) to interact with greater affinity.

The only ZIC3 variant protein that displayed reduced fluorescence/interaction in the BiFC assay was Q249X which truncates immediately prior to the ZFD. In the functional assays this variant has reduced ability to inhibit TCF-dependent transcription and ablated transactivation ability relative to wildtype ZIC3. Comparatively, mutant S252X truncates just 3 amino acids after Q249X, yet has increased ZIC3-TCF7L1 interaction relative to wildtype ZIC3. This could be interpreted to mean that the glutamine residue and/or the following two amino acids have a key role in TCF7L1 interaction. Indeed, the glutamine residue (Q), but not the serine (S), is located within the highly conserved ZIC3 ZFNC domain, potentially identifying a role for this domain in the ZIC3-TCF7L1 protein interaction. For example, the ZIC3 lysine residue K248, within the ZFNC, is targeted by the post-translational modification SUMOylation (Chen et al. 2013). In this process SUMO protein is covalently attached to the lysine amino acid and influences protein stability and subcellular localisation of the ZIC3 protein (Chen et al. 2013). Due to the proximity of Q249 to K248, it is possible that variants within this domain may alter the SUMOylation process. Whether SUMOylation or the ZFNC domain are involved in the ZIC3-TCF7L1 interaction will need to be further investigated, perhaps by altering the lysine residue to prevent SUMOylation or by deleting the ZFNC to investigate how this domain influences TCF7L1 interaction.

4.3.5 Additional molecules may be required for ZIC-TCF repression at WREs

Given that ZIC and TCF co-expression consistently leads to inhibition of transcription at WREs, that ZIC and TCF proteins can bind one another, and that they can be co-immunoprecipitated with DNA at WREs in cell-based assays (Pourebrahim et al. 2011), a model has emerged in which the ZIC-TCF binding is associated with transcriptional repression (and conversely a lack of binding with enhanced transcription at WREs). Assuming the BiFC data presented in this Chapter is an accurate representation of the ZIC3-TCF7L1 interaction, based on the molecular model proposed, it may be expected that a stronger ZIC3-TCF7L1 interaction results in stronger transcriptional repression. In most cases examined here, however, increased ZIC3-TCF7L1 interaction correlated with increased transcription (i.e. less repression) in the TOPflash assay (Table 4.4). One explanation consistent with this is that the ZIC3 protein variants in these assays are unable to interact with one or more additional proteins that are also critical for transcriptional repression.

Chromatin structure plays a large role in both transcriptional activation and repression, and when in a restrictive state prevents the access of RNA polymerase or other transcriptional regulators binding to DNA (Thiel et al. 2004). Published data suggests that ZIC proteins may be involved in chromatin modification. For example, in embryonic stem cells ZIC2 has been shown to interact with the Mbd3-NuRD complex which has nucleosome remodelling and histone deacetylase activities (Luo et al. 2015). Furthermore, knockdown of *ZIC2* was shown to affect expression of known Mbd3-NuRD targets. Whether these interactions are required for ZIC function *in vivo*, however, has not yet been examined. It is therefore possible that ZIC3 similarly interacts with NuRD complexes or other similar protein complexes to regulate nucleosome remodelling and gene expression. Additionally, TCF7L1 has been shown to recruit C-terminal binding protein (CtBP) and histone deacetylase 1 (HDAC1) proteins to repress transcription by stabilising chromatin in cell lines (Brannon et al. 1999; Eshelman et al. 2017). ZIC3 could perhaps provide a bridge between TCF proteins and these proteins. If the ZIC3 variant proteins retain TCF7L1 binding ability but are less able to recruit repressive machinery, increased expression at WREs would be observed.

4.3.6 Increased ZIC3-TCF7L1 interaction correlates with decreased transactivation

In every case in which the *ZIC3* variant proteins exhibited increased TCF7L1 interaction compared to wildtype ZIC3, the variant protein was unable to stimulate transcription in the transactivation assay (Table 4.4). It is possible that this stronger affinity for TCF7L1 protein prevents ZIC3 from acting upon ZREs to function as a classical transcription factor. Indeed, the transactivation assay experiments presented here were conducted in a Wnt unstimulated state (similar to the BiFC experiments), and a lack of transcriptional activation could reflect ZIC3 proteins being sequestered by endogenous TCF7L1 protein. To test this theory, the same transactivation assay can be conducted in a Wnt stimulated state

(which biases against ZIC-TCF interaction). This can be achieved by co-transfecting in a form of β-catenin that drives high levels of Wnt signalling (such as the stabilised β-catenin delta N89 variant). Alternatively, Wnt signalling can be stimulated by using Wnt conditioned media (e.g. by the addition of a Wnt activator such as Wnt3A or by addition of CHIR99021, an inhibitor of GSK3) to drive high, yet controlled levels of Wnt activity. If ZIC3 is in fact sequestered by TCF protein, the sequestration will presumably be alleviated in a high Wnt state and result in increased transactivation ability by these ZIC3 protein variants.

In summary, I have investigated the functional effects of 17 *ZIC3* Heterotaxy-associated variants in well-established molecular cell-based assays. Based on the variants investigated, four human variants located outside the ZIC3 ZFD do not alter protein function in the assays examined, implying these variants are not pathogenic. Of those that do exhibit pathogenic qualities, it was not possible to distinguish whether the pathology associated with these variants is due to a loss of transcription factor function or due to a loss of ZIC3 as a Wnt co-repressor as both functions were altered. ZIC3 is likely to interact with other co-repressors/chromatin complexes to mediate its activity, and the variants examined in this Chapter suggest these interactions are also affected.

Chapter 5: Can the transcription factor and co-factor functions of ZIC3 be separated?

5.1 Introduction

ZIC proteins are unusual in that they can influence transcription in both a DNA binding and protein binding manner, however, it is not well understood how the ZICs can alternate between these two molecular functions or when each role is necessary during embryonic development. Molecular analysis of human Heterotaxy associated variants in Chapter 4 could not definitively determine which of the two functions of ZIC3 (transcription factor or co-factor) is the underlying mechanism responsible for the ZIC3-associated Heterotaxy phenotype, as no mutant protein affected one function without altering the other. This may suggest that both functions of ZIC3 are needed to prevent Heterotaxy. In order to determine whether only one function underlies ZIC3-associated Heterotaxy, two alleles are needed, that each affect one, but not the other, function of ZIC3. Once such mutations are identified, a detailed comparison of the respective mouse strains could reveal when and where these distinct functions are required during embryonic development. Whether these functions of ZIC3 can be separated, however, remains to be elucidated.

5.1.1 Can these functions be separated?

Evidence that the transcription factor and co-factor functions of ZIC proteins can be isolated comes from a *Zic2* mouse allele. A variant of *Zic2*, that alters the second cysteine of the fourth zinc finger to a serine, can no longer bind DNA (Brown et al. 2005) and exhibits *in vivo* loss-of-function (Elms et al. 2003). This mutant is still able to repress β-catenin mediated transcription in a TOPflash assay and bind TCF7L2 when overexpressed in cells (Pourebrahim et al. 2011), providing the first evidence that these two distinct functions of the ZICs and the location of these domains within the gene are independent of each other. This indicates that the resulting *Zic2* mouse mutant phenotype produced is a consequence of the loss of DNA binding only and not due to a loss of a co-factor role (such as interaction with TCF proteins or a loss of Wnt inhibition). The same mutation is yet to be identified within the other ZIC family members, and the inverse mutation has not been identified in any ZIC protein, i.e. a mutation that results in a loss of Wnt inhibition but retains DNA binding/transactivation activity. Understanding the role of ZIC3 during development would be greatly heightened by identification of both of these mutations within the *ZIC3* gene.

5.1.2 A screen for mutations that separate function

Identification of these mutations will be accelerated by understanding the specific functional domains within ZIC3. Currently, however, there is limited knowledge surrounding where the functional domains within the ZIC proteins lie and often large regions of the protein are assigned overlapping

functions. For example, the ZIC ZFD has been implicated in both DNA binding and protein binding activities (Brown et al. 2005; Pourebrahim et al. 2011). Badis et al. 2009 conducted a protein binding microarray using the purified ZIC3 ZFD and identified a DNA motif bound by the ZFD. Additionally, mutations within the ZFD of ZIC proteins can result in a loss of DNA binding as demonstrated by EMSA assays and, consequently, alter transcriptional activation in cell based overexpression assays (Brown et al. 2005). Published data also suggests that the ZIC ZFD is necessary for protein-protein interactions, particularly in terms of TCF binding and Wnt transcriptional repression. ZIC2 deletion constructs lacking the ZFD and C-terminus could no longer bind TCF7L2 protein or repress TOPflash reporter activity (Pourebrahim et al. 2011). This study, however, used large deletion mutants of ZIC2 without confirming nuclear localisation of the truncated proteins (the site within the cell the ZIC-TCF interaction occurs). However, given the ZFD of ZIC2 and ZIC3 differ by only 14 amino acids, the ZFD is also predicted to be the site where ZIC3 binds TCF proteins, and necessary for Wnt repression.

One difficulty in interpreting these results is evidence suggesting there are 'subdomains' within each identified functional domain, or that multiple regions are necessary for protein function. For example, as mentioned previously, there is some evidence to suggest that not all five zinc fingers participate in DNA binding. The first two zinc fingers of the ZFD share a hydrophobic core and form a single structural unit termed the $CWCH_2$ motif. This motif is characterised by the presence of a tryptophan (W) residue between the two canonical cysteines of each zinc finger (Aruga et al. 2006) as well as an increased linker sequence between the two zinc fingers. Due to these features it is suggested that this motif may not contain canonical (i.e. DNA binding) zinc fingers but instead may facilitate DNA binding by zinc fingers 3-5, however further analysis is required to confirm these findings (Hatayama and Aruga 2010; Houtmeyers et al. 2013). Nevertheless, as a *Zic2* variant affecting the fourth zinc finger (described above) affects the transcription factor, but not co-factor, function of ZIC2, perhaps not all zinc fingers are required for each role and a similar mutation affecting the ZIC3 ZFD may also separate these functions. Similarly, as the ZFD was found to be necessary for TCF interaction, it would be expected that proteins lacking the ZFD would no longer interact with TCF proteins. Despite this, *ZIC3* human nonsense and frameshift mutations (investigated in Chapter 4) that truncate within or prior to the ZFD could interact with TCF7L1 in a BiFC assay and were still able to inhibit Wnt signalling in the cell based TOPflash assays (albeit at lower levels to wildtype), suggesting that there are N-terminal regions of the ZIC3 protein that are involved in repressing Wnt dependent transcription. Further domain mapping experiments are therefore necessary to determine if the ZIC3 Wnt repressor domain can be focused to a more isolated area within the ZFD, or to another region outside the ZFD of the ZIC3 protein.

In mapping these domains, it would be logical to assume that the domains required for each function of ZIC3 be localised to a conserved region. There are several evolutionary conserved regions in the ZIC proteins including the ZOC, ZFNC, SANC, SACC and also several low complexity regions that have not been assigned a functional role despite their strong conservation. It is possible that some of these domains contribute, at least in part, to the transcription factor or co-factor function of the ZIC proteins. Indeed, low complexity regions such as those found in the ZIC proteins are characteristic of intrinsic disorder, a term used to describe proteins or regions of proteins that lack a conventional, well-defined tertiary structure in their native state. Instead, disordered proteins, or regions of disorder, are unstructured but will fold into a more structured conformation upon binding to their target or in response to post-translational modifications (Mészáros et al. 2007; Dyson 2016). This flexibility in structure appears to confer multiple advantages including the ability to interact with numerous binding partners. Additionally, it enables interactions with high specificity, but low affinity, which facilitates the rapid disassembly of the protein complex (Dyson and Wright 2005; Dyson 2016). This structural polymorphism allows disordered proteins to be involved in a vast array of biological processes and protein functions, such as accessibility to post translational modifications, involvement in cell signalling, control of the cell cycle and transcriptional activation (Mészáros et al. 2007). Consequently, identification of disordered regions within ZIC3 may indicate regions important for one specific function.

5.1.3 Chapter Aims

The ZIC3 protein contains many conserved domains but not all have a well characterised function. This chapter will investigate putative functional domains of the ZIC3 protein via TOPflash, BiFC and transactivation assays. The goal of this chapter is to identify a mutation in *ZIC3* that ablates only one ZIC3 protein function (i.e. either the transcription factor function or the co-factor function), but not both. Specifically, this chapter will investigate whether:

   i)   Evolutionary conserved domains within ZIC3 are required for Wnt inhibition
   ii)   Individual zinc fingers within ZIC3 contribute differently to Wnt inhibition
   iii)   Evolutionary conserved domain within ZIC3 are required for transactivation
   iv)   Individual zinc fingers within ZIC3 contribute differently to transactivation

5.2 Results

5.2.1 ZIC proteins contain intrinsically disordered regions

As protein-protein interaction domains and transactivation domains can be associated with intrinsic disorder (Liu et al. 2006; Shammas 2017), the ZIC proteins were screened for disordered regions outside of the ZFD. Computer algorithms can predict disordered proteins or regions of disorder from sequence composition since disorder is characterised by a high concentration of charged and polar amino acids and a low prevalence of hydrophilic residues (Dyson and Wright 2005; Mészáros et al. 2018). There are several online bioinformatics prediction tools available to determine the tendency of a protein to be disordered. One of these, IUPred2A, was used to assess the level of disorder predicted to occupy the protein sequence of the human ZIC proteins (ZIC1-5) (Mészáros et al. 2018). The program provides intrinsic disorder predictions using both IUPred2 and ANCHOR2 algorithms. The output from the analysis is a graph showing the disorder tendency of each residue of the protein. Higher values indicate predicted disordered regions, while lower values suggest protein order.

Analysis of the human ZICs identifies intrinsically disordered regions in the residues adjacent to the ZFD, with only minimal disorder predicted within the ZFD itself for each of the five proteins (Figure 5.1). For example, in all human ZIC proteins, the C-terminus is predicted to be almost completely disordered as are multiple regions of the N-terminal regions. ZIC3 contains several predicted regions of disorder within the N-terminal region some of which overlap with conserved domains. For example, the ZIC3 ZOC, SANC, SACC domains and histidine repeat were predicted to be disordered regions by both algorithms, while the alanine repeat was predicted to be a disordered region by the ANCHOR2 algorithm but not IUPred2 (Figure 5.1).

In addition to exhibiting intrinsic disorder, domains directly involved in protein function, such as a transactivation domain or other protein binding domains are likely to be conserved. Given that annotated domains such as the ZOC, ZFNC, SANC, SACC and low complexity regions are conserved both across species and within the ZIC protein family, it is likely these domains contribute somewhat to protein function. Taken together, it is possible these conserved regions play a currently unrecognised role in ZIC protein function.

To investigate whether these conserved regions contribute to the function of ZIC3, each region was selectively deleted from wildtype ZIC3 (Figure 5.2). The N' terminal region of the ZIC proteins has not yet been assigned a particular function and truncation mutations lacking part or all of the ZFD (assessed in Chapter 4) did not lose all inhibitory function, suggesting there is some N-terminal regions of ZIC3 that play a role in Wnt inhibition. Overall, seven deletion constructs were generated to study

ZIC1

ZIC2

ZIC3

ZIC4

ZIC5
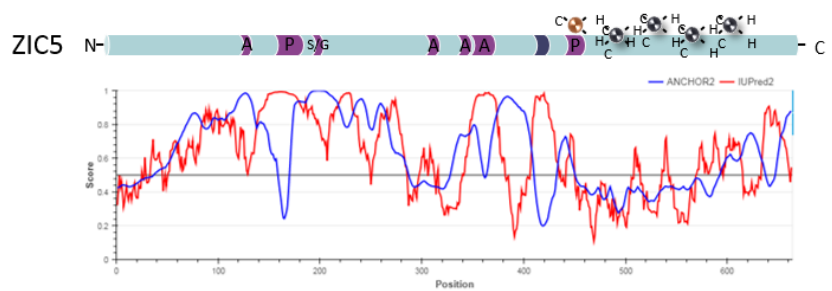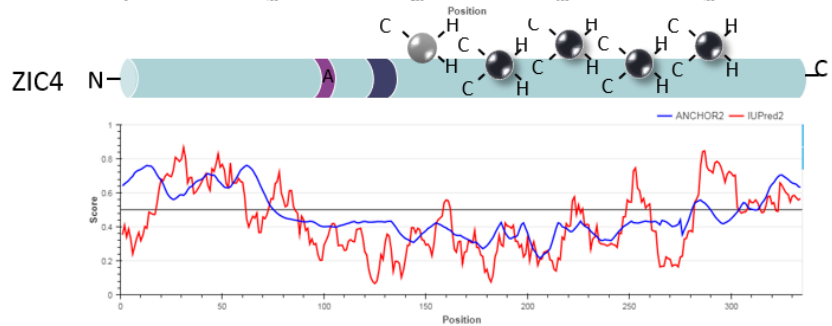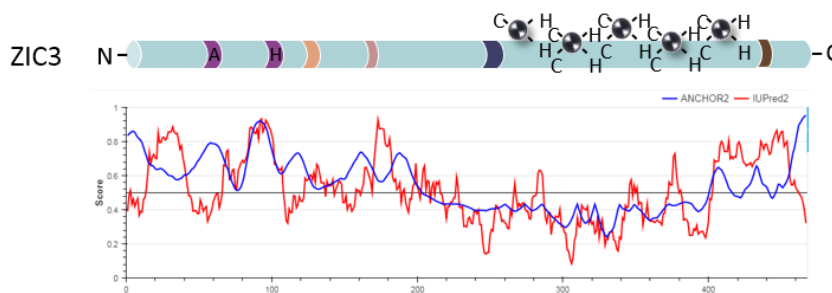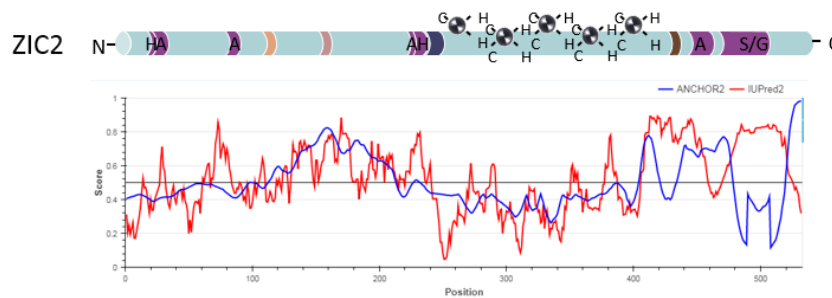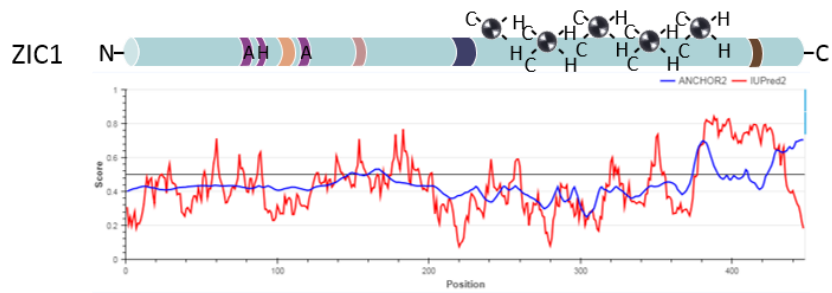
C₂H₂ zinc finger

Diverged C₂H₂ zinc finger

Diverged C₂H₂ zinc finger

Low complexity region

ZOC domain

SANC domain

ZFNC domain

SACC domain

Figure 5.1: ZIC proteins are predicted to have intrinsically disordered regions. IUPred2 was used to predict disorder within each of the five human ZIC proteins and displays outputs from IUPred and ANCHOR2 algorithms (red and blue output, respectively). A score between 0-0.5 (threshold line) indicates predicted order for each residue, whilst a score of 0.5-1 indicates a predicted disordered region of the protein. ZOC: Zic-opa conserved, ZFNC; zinc finger N-terminus conserved, SANC: Subclass A N terminally conserved, SACC: Subclass A C terminally conserved.
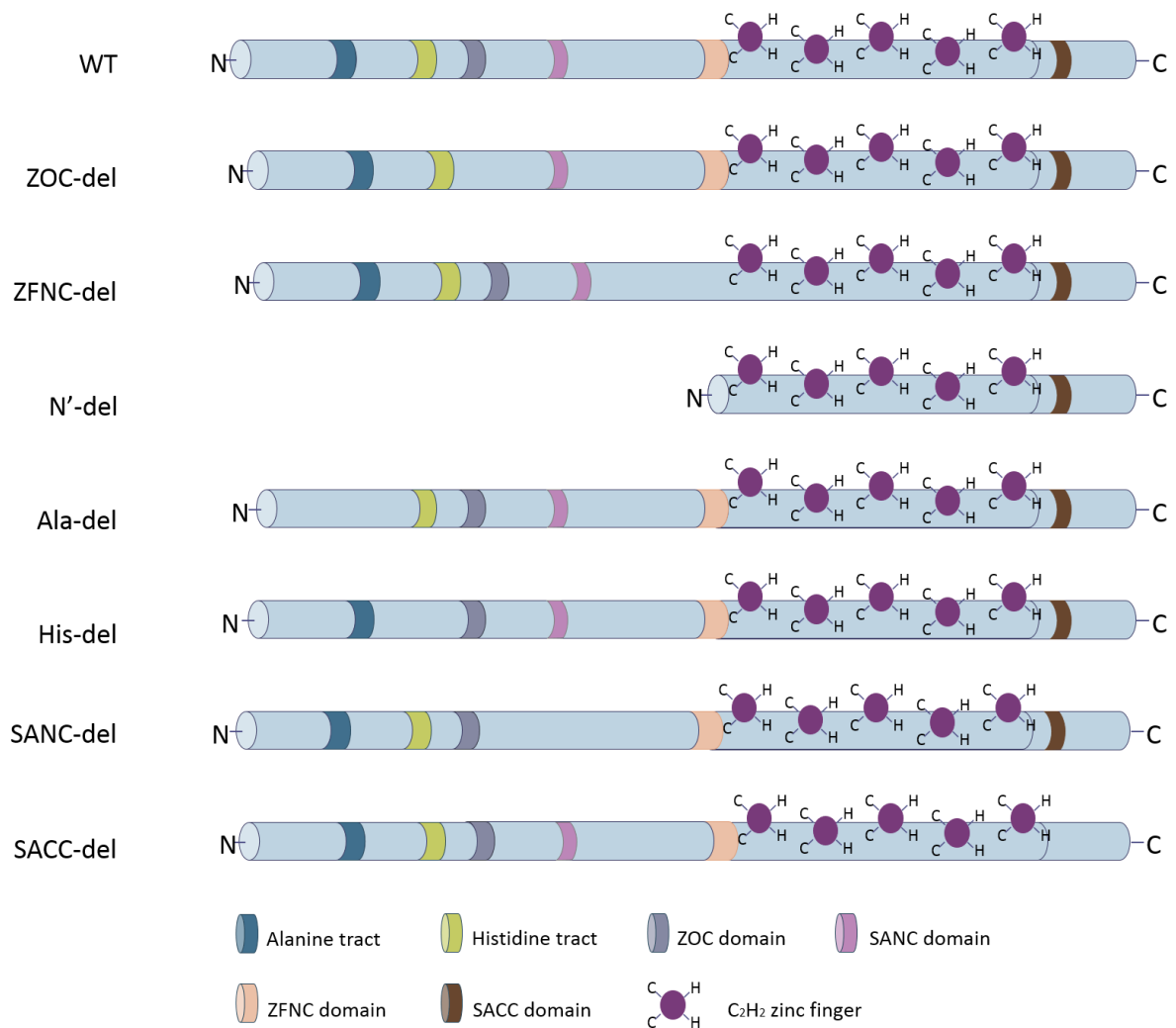
Figure 5.2: Schematic representation of the protein structure of ZIC3 deletion mutations. Seven evolutionary conserved domains within the ZIC3 protein were selectively deleted from the ZIC3 coding sequence via site directed mutagenesis to produce expression constructs lacking the specific domains. Additionally, the entire N-terminal domain prior to the ZIC3 zinc finger domain (ZFD) was deleted to produce a ZIC3 protein with the ZFD and C-terminus only (termed N'-del). ZOC: Zic-opa conserved, ZFNC; zinc finger N-terminus conserved, SANC: Subclass A N terminally conserved, SACC: Subclass A C terminally conserved.

the function of evolutionary conserved domains within the ZIC3 protein. The ZOC, ZFNC, SANC, SACC domains, as well as the alanine repeat, histidine repeat and the entire N 'domain prior to the ZFD was deleted from ZIC3 (producing a protein that contains only the ZFD and C-terminal region) via site-directed mutagenesis to generate seven expression constructs to study protein function.

Additionally, individual zinc fingers were targeted for mutation to determine whether each zinc finger contributes differently to ZIC3 function. The importance of the ZIC ZFD has been well documented for both DNA binding and protein interaction (Lim et al. 2007; Pourebrahim et al. 2011). Published data on the $CWCH_2$ motif suggests that the zinc fingers do not undertake the same role. To investigate whether individual zinc fingers contributes more or less to Wnt inhibition, and to localise the ZIC3-TCF7L1 interaction to a more specific region within the ZFD, each zinc finger was selectively targeted by a single amino acid substitution. A missense mutation was introduced into ZIC3 that altered the second cysteine of each zinc finger to a serine, generating five mutant constructs affecting the ZFD (C268S, C302S, C335S, C365S and C393S) (Figure 5.3). The cysteine residue is directly involved in complexing the zinc ion that makes each C_2H_2 zinc finger. Alteration of the cysteine residue results in the amino acid residue unable to chelate the zinc ion which ultimately destroys the zinc finger complex. It is predicted that this single amino acid substitution would disrupt each individual zinc finger, allowing for functional roles to be mapped to specific zinc fingers within this domain.

Overall, 12 new ZIC3 variant proteins were generated. Each of the variant constructs were transferred to both the V5-Dest expression vector (to attach a V5-epitope tag to the N-terminus of the protein) for use in immunofluorescence, transactivation and TOPflash assays as well as the V1-ORF expression vector (to attach the N-terminal half of the Venus fluorescent tag to the N-terminus of the ZIC3 protein) to be used in BiFC analysis, thereby generating 24 new ZIC3 expression constructs.
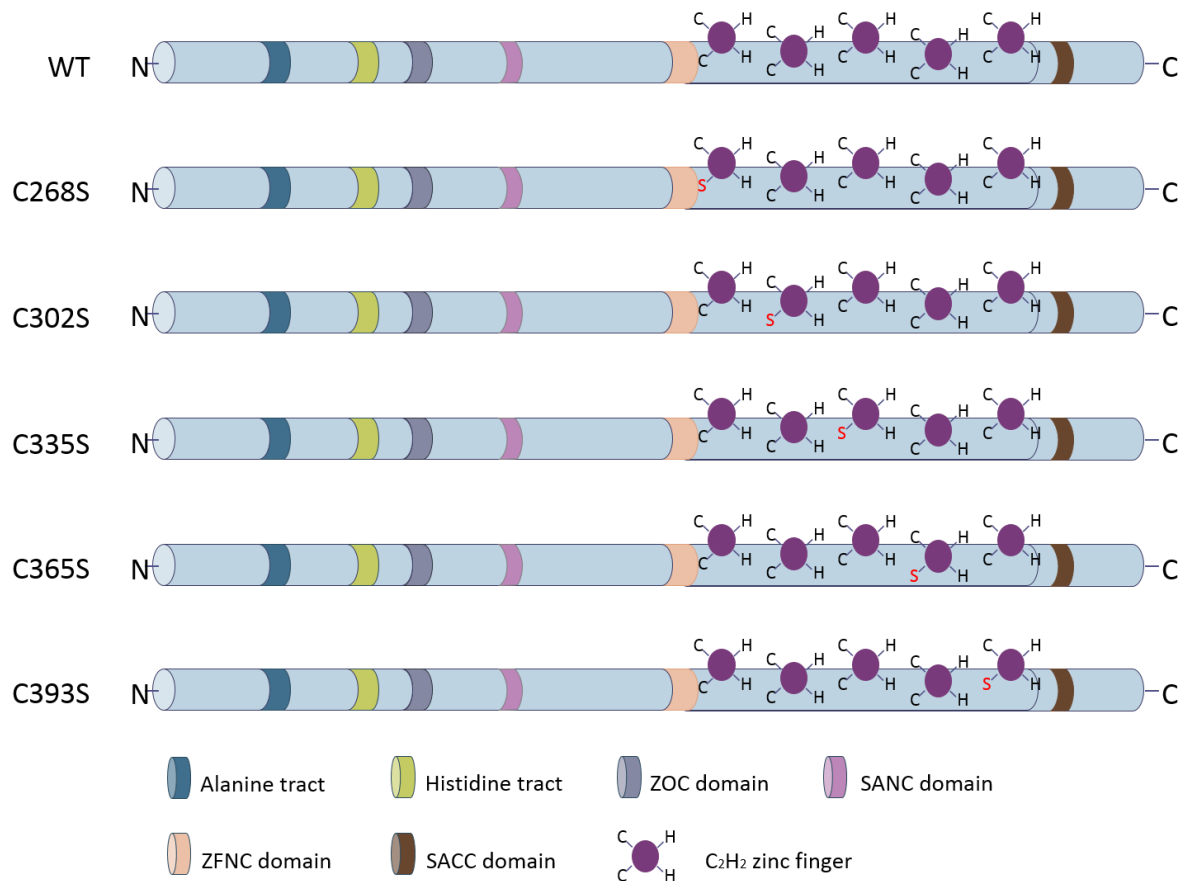
Figure 5.3: Schematic representation of the ZIC3 protein structure of targeted missense mutations. The second cysteine of each zinc finger was converted to serine through site directed mutagenesis. Position of each mutation (cysteine to serine) is outlined above and the amino acid change on each zinc finger is shown in red. ZOC: Zic-opa conserved, ZFNC; zinc finger N-terminus conserved, SANC: Subclass A N terminally conserved, SACC: Subclass A C terminally conserved, C: cysteine, H: histidine, S: serine

5.2.2 ZIC3 variants produce detectable protein that localise to the nucleus.

Each of the 12 V5-ZIC3 variant expression constructs were transfected into HEK293T cells and subcellular localisation of the protein was assessed using standard immunofluorescence techniques. Each of the ZIC3 deletion mutations investigated produced detectable protein that localised to the nucleus (Figure 5.4). The N-terminal deletion had a small but significant increase in nuclear localisation while the deletion of the ZFNC and missense mutations C302S, C335S, C365S had slightly reduced, but significant, nuclear localisation compared to wildtype ZIC3. In all cases however, nuclear protein was above 80% and thus were deemed appropriate to use in downstream functional studies.

5.2.3 Analysis of ZIC3 domains in Wnt repression and TCF7L1 protein interaction

5.2.3.1 Deletion of well characterised ZIC3 domains alter WNT repression ability, but still interact with TCF7L1

To determine which regions of the ZIC3 protein are required for Wnt inhibition, ZIC3 deletion proteins lacking specific conserved domains were assayed in the established TOPflash assay described in Chapter 4 (section 4.1.2). Additionally, a BiFC assay was used to determine whether the mutant proteins could still interact with TCF7L1. Deletion of the ZOC domain or the ZFNC domain did not significantly alter the repression ability of ZIC3 in the TOPflash assay, as these mutations inhibit transcription at the same level as wildtype ZIC3 (Figure 5.5). Similarly, the ZOC deletion could interact with TCF7L1 at a similar level to wildtype protein in the BiFC assay, but deletion of the ZFNC region reduced the ZIC3-TCF7L1 interaction. The reduced protein-protein interaction however does not produce a noticeable functional consequence in the TOPflash assay, as the ZFNC deletion did not significantly alter transcriptional inhibition compared to wildtype protein. The N-terminal deletion drove overstimulation at the TOPflash reporter as luciferase values were higher than when β-catenin was transfected alone, suggesting that despite the presence of the ZFD this mutant has lost all repression ability and instead adopts an activator role. Interestingly, even though this mutant protein lost its repressive function, it was still able to interact with TCF7L1 protein at stronger levels than wildtype ZIC3. This result suggests that the ZFD alone is not sufficient to inhibit β-catenin mediated transcription and that the N-terminus possesses regions necessary for ZIC3-mediated repression and/or correct protein conformation.
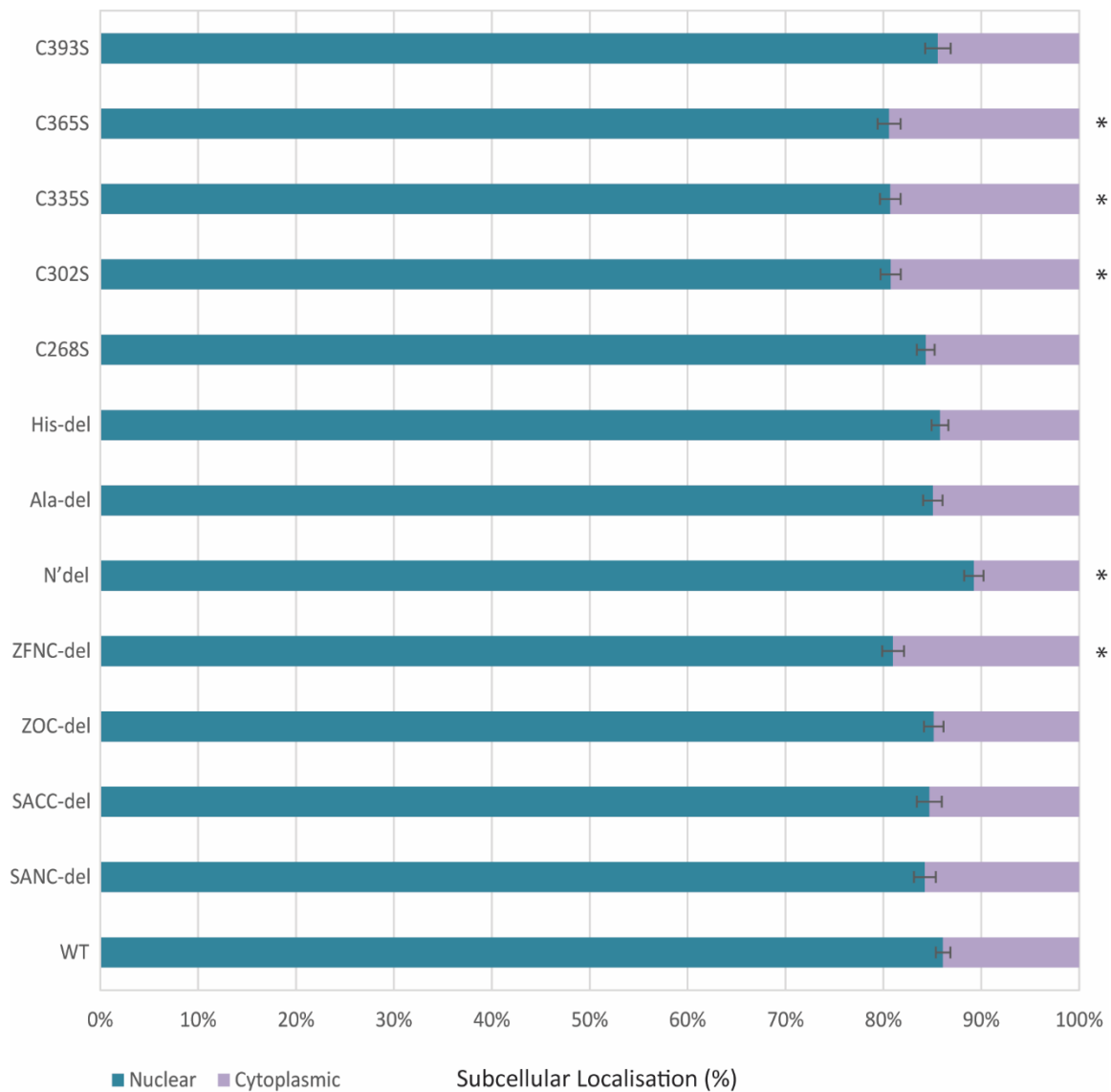
Figure 5.4: ZIC3 variant proteins localise to the nucleus. HEK293T cells were transfected with the V5-tagged ZIC3 expression constructs shown. 24 hours post-transfection, cells were prepared for immunofluorescence microscopy and stained with α-V5 antibody to detect the transfected V5-ZIC3 proteins as well as α-LaminB1 to mark the nuclear membrane. Images of cells were captured with a fluorescent microscope and ImageJ software was used to determine the relative nuclear and cytoplasmic fluorescence intensity. Error bars represent standard error *: $p < 0.05$ one-way ANOVA.
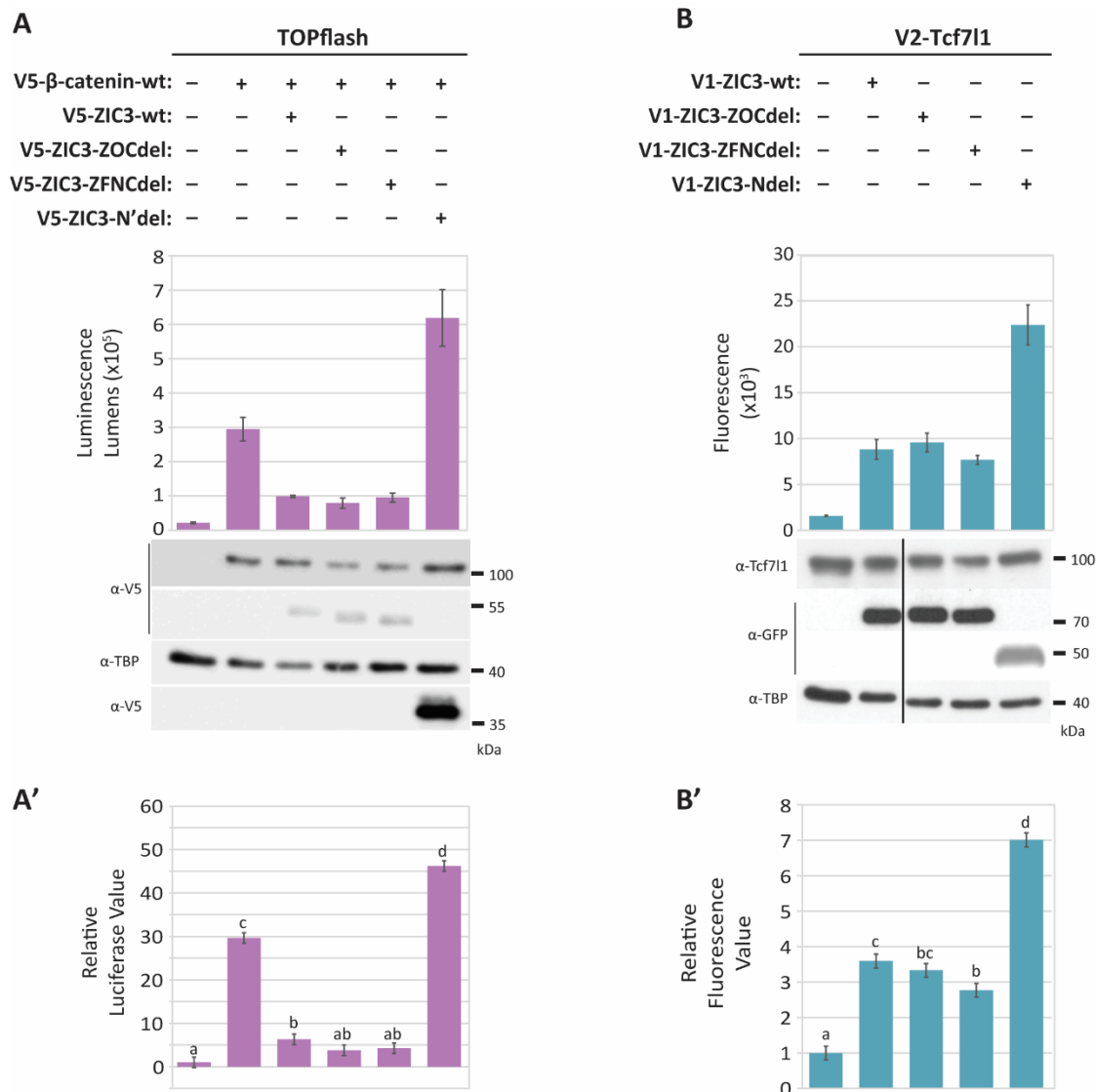
Figure 5.5: ZIC3 N-terminal regions have a role in Wnt inhibition. (A) HEK293T cells were transfected with a TOPflash reporter, V5-tagged β-catenin-wt and a V5-tagged ZIC3 expression construct as shown. Luminescence was measured 24 hours post-transfection in each of three internal replicate samples. **(B)** HEK239T cells were co-transfected with pV2-TCF7L1 and either pV1-ZIC3-wt or a pV1-ZIC3-variant expression plasmid. The mean fluorescence intensity emitted from each of three internal replicate samples was detected at 48 hours post-transfection. Three independent experiments were performed. **(A, B)** One representative experiment with error bars representing the SD between the three internal replicates. A western blot was performed to confirm expression of transfected proteins. For BiFC experiments α-GFP was used to detect V1-ZIC3 protein and α-TBP was used as a loading control for both experiments. **(A', B')** Pooled data from three independent repeats. Values depict the mean relative luciferase or fluorescent values. Error bars depict SEM. Different letters denote statistical differences of $p < 0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc.

5.2.3.2 Subclass A conserved domains in Wnt signalling

To investigate whether the subclass A conserved domains SANC and SACC also have a role in Wnt repression, ZIC3 constructs were generated that delete these two conserved domains from the protein. Following transfection in HEK293T cells for analysis in the TOPflash assay, both the SANC and SACC deletion mutants were able to inhibit Wnt signalling at the same level as wildtype ZIC3 suggesting these regions do not have a role in the repression of β-catenin-mediated transcription (Figure 5.6). When assessed in the BiFC assay, the SANC deletion was able to interact with TCF7L1 at a similar level to wildtype ZIC3. Conversely, the SACC deletion mutant was able to interact with TCF7L1 at a significantly stronger level than wildtype ZIC3. These results suggest that these two evolutionary conserved domains do not play a role in Wnt inhibition, however the SACC domain is required for normal protein-protein interactions.

5.2.3.3 Low complexity regions in Wnt signalling

Deletion of both the alanine and histidine repeat from ZIC3 did not significantly alter the ability to repress Wnt signalling compared to wildtype ZIC3 (Figure 5.7). Thereby suggesting the alanine and histidine repeat do not contribute to the repression of β-catenin-mediated transcription. Additionally, both the alanine and histidine deletions had a small but significant decrease in fluorescence in the BiFC assay (Figure 5.7), suggesting these mutants have decreased TCF7L1 interaction. Similar to the ZFNC deletion, the removal of these domains and consequent reduced TCF7L1 interaction does not impact the proteins ability to repress Wnt signalling in the TOPflash assay conditions used here.

5.2.3.4 Zinc finger mutants in Wnt signalling

When assessed in the cell based TOPflash assay, each of the five ZIC3 zinc finger mutant protein did not completely lose their Wnt repressive ability (Figure 5.8). Zinc finger variants C268S (ZF1), C302S (ZF2), C335S (ZF3), and C393S (ZF5) could inhibit at levels similar to wildtype ZIC3, whereas ZF4 (C365S) mutant protein had a small but significant decrease in repression ability compared to wildtype protein. Together, this would suggest that the ZFD does not play as big of a role in Wnt repression as previously thought and that, as expected, not all the zinc fingers have an equal role in Wnt inhibition. To confirm whether the relatively minor effect on Wnt repressive ability of the ZIC3 zinc finger variants is associated with normal TCF7L1 interaction, a BiFC assay was conducted. It was found that all zinc finger mutant proteins were able to interact with TCF7L1 at varying degrees (Figure 5.8). Only variant C302S, which affects the second zinc finger, interacts with TCF7L1 at the same level as wildtype ZIC3. ZIC3 variants C268S, C335S and C393S had a small but significant increase in TCF7L1 interaction, while C356S had a slight decrease in fluorescence, supporting the partial loss of repression in the TOPflash
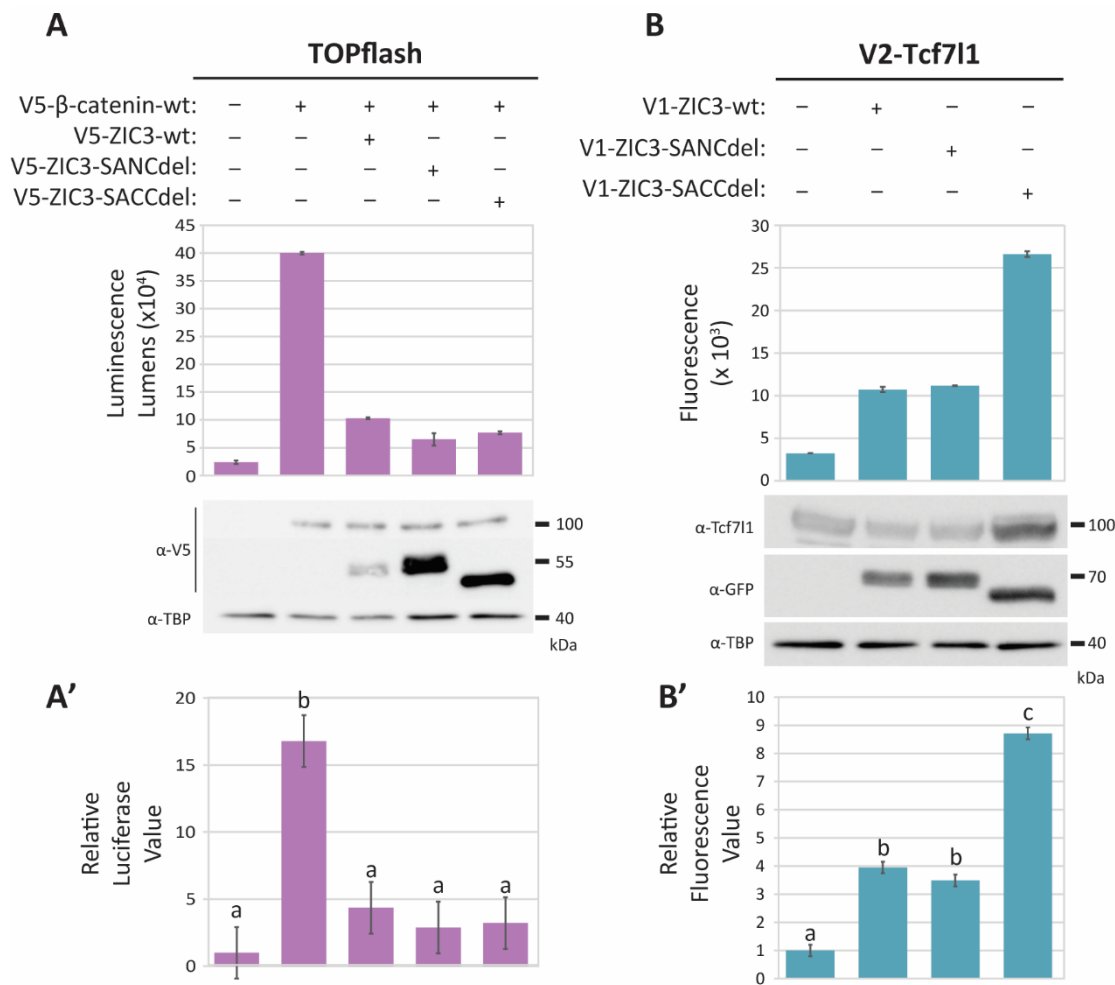
Figure 5.6: ZIC3 subclass A conserved domains do not play a role in Wnt inhibition. (A) HEK293T cells were transfected with a TOPflash reporter, V5-tagged β-catenin-wt and a V5-tagged ZIC3 expression plasmid shown. Luminescence was measured 24 hours post transfection in each of three internal replicate samples. **(B)** HEK239T cells were co-transfected with pV2-TCF7L1 and either pV1-ZIC3-wt or pV1-ZIC3-variant expression plasmid. The mean fluorescence intensity emitted from each of three internal replicate samples was detected at 48 hours post transfection. Three independent experiments were performed. **(A, B)** One representative experiment with error bars representing the SD between the three internal replicates. A western blot was performed to confirm expression of transfected proteins. For BiFC experiments α-GFP was used to detect V1-ZIC3 protein and α-TBP was used as a loading control for both experiments. **(A', B')** Pooled data from three independent repeats. Values depict the mean relative luciferase or fluorescent values. Error bars depict SEM. a Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc.
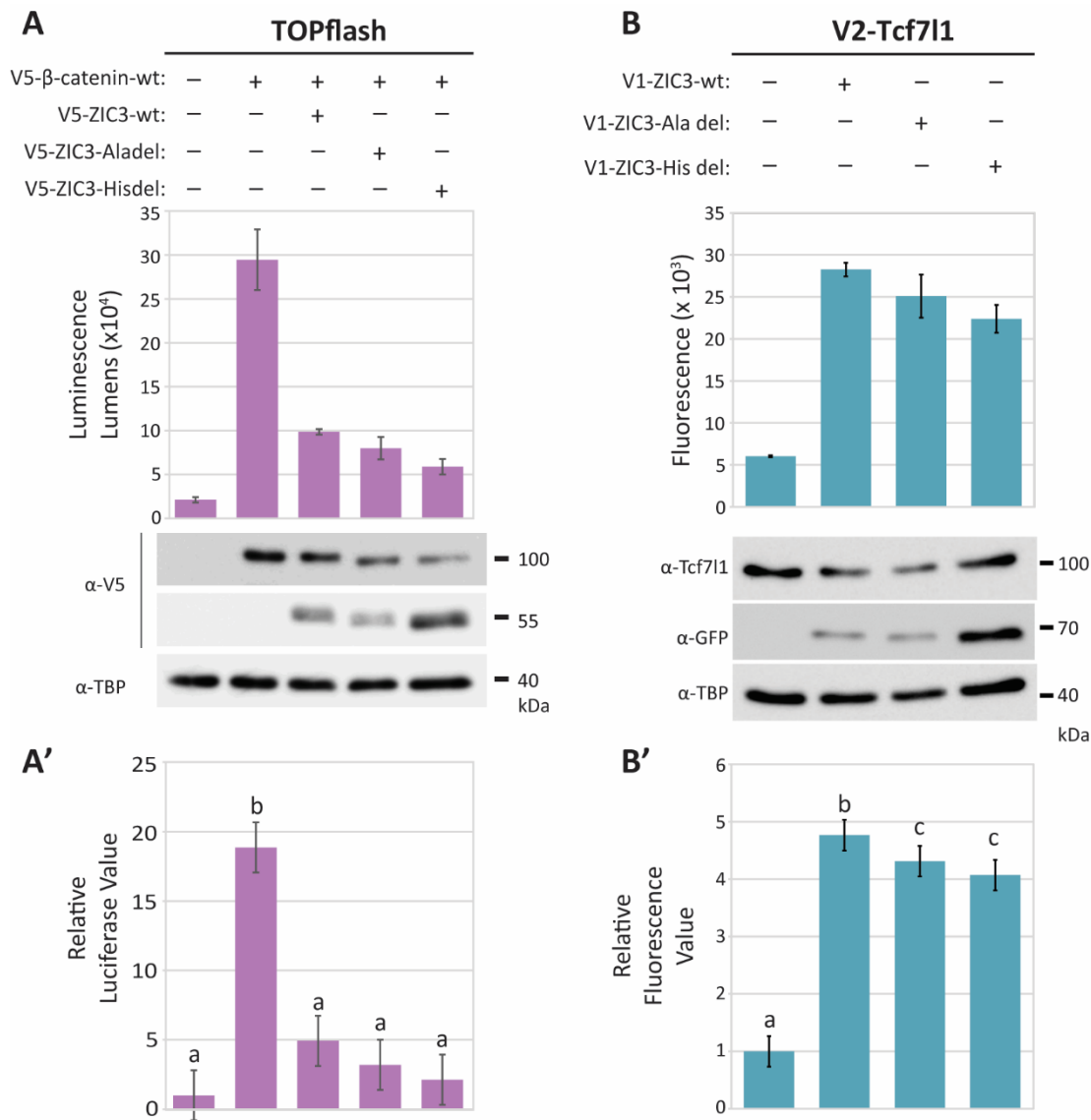
162

Figure 5.7: ZIC3 low complexity regions are required for normal protein interaction but are not needed for Wnt inhibition. (A) HEK293T cells were transfected with a TOPflash reporter, V5-tagged β-catenin-wt and a V5-tagged ZIC3 expression plasmid shown. Luminescence was measured 24 hours post-transfection in each of three internal replicate samples. **(B)** HEK239T cells were co-transfected with pV2-TCF7L1 and either pV1-ZIC3-wt or a pV1-ZIC3-variant expression plasmid. The mean fluorescence intensity emitted from each of three internal replicate samples was detected at 48 hours post transfection. Three independent experiments were performed. **(A, B)** One representative experiment with error bars representing the SD between the three internal replicates. A western blot was performed to confirm expression of transfected proteins. For BiFC experiments α-GFP was used to detect V1-ZIC3 protein and α-TBP was used as a loading control for both experiments. **(A', B')** Pooled data from three independent repeats. Values depict the mean relative luciferase or fluorescent values. Error bars depict SEM. Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc.
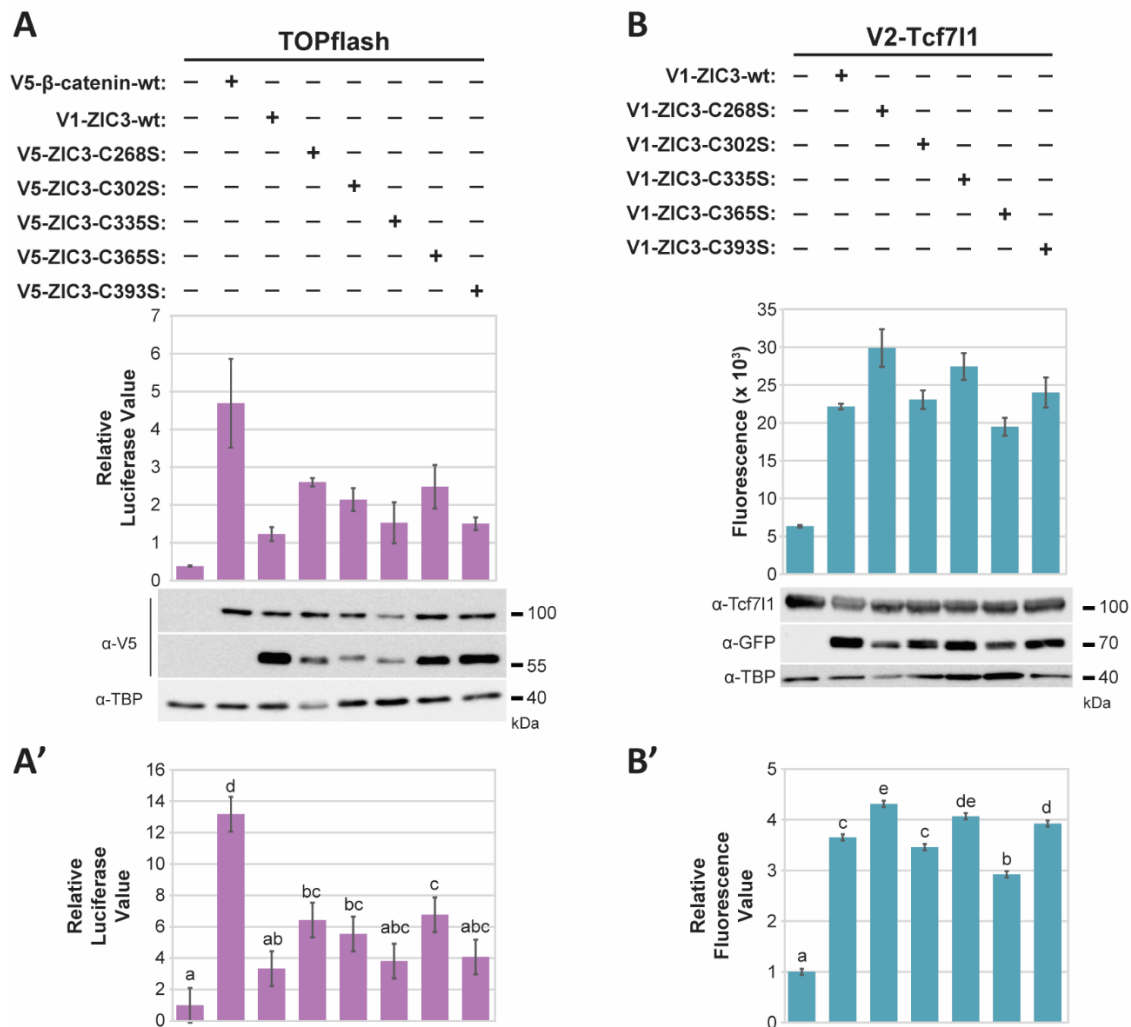
Figure 5.8: ZIC3 zinc finger mutant proteins can still repress Wnt signalling and interact with TCF7L1.
(A) HEK293T cells were transfected with a TOPflash reporter, V5-tagged β-catenin-wt and a V5-tagged ZIC3 expression plasmid shown. Luminescence was measured 24 hours post-transfection in each of three internal replicate samples. **(B)** HEK239T cells were co-transfected with pV2-TCF7L1 and either pV1-ZIC3-wt or a pV1-ZIC3-variant expression plasmid. The mean fluorescence intensity emitted from each of three internal replicate samples was detected at 48 hours post transfection. Three independent experiments were performed. **(A, B)** One representative experiment with error bars representing the SD between the three internal replicates. A western blot was performed to confirm expression of transfected proteins. For BiFC experiments α-GFP was used to detect V1-ZIC3 protein and α-TBP was used as a loading control for both experiments. **(A', B')** Pooled data from three independent repeats. Values depict the mean relative luciferase or fluorescent values. Error bars depict SEM. Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc.

assay. Together this indicated that this zinc finger has a larger role in TCF7L1 mediated Wnt repression than the other zinc fingers.

5.2.4 ZIC3 conserved domains are required for transactivation

In addition to acting as a co-factor to inhibit Wnt signalling, ZIC proteins also act as transcription factors to regulate gene expression. Isolation of these protein functions would allow great insight into when and which ZIC3 function is necessary during development. Transcription factors require a DNA binding domain but also a transactivation domain, a region of the protein that interacts with transcriptional machinery to regulate target gene transcription. The ZFD of ZIC proteins has been implicated in DNA binding, but the transactivation domain is yet to be identified within the protein. Transactivation assays are routinely used to determine which regions of the protein are necessary for or sufficient to drive gene expression. To determine which regions of the ZIC3 protein are required for transactivation, the 12 structural ZIC3 deletion mutant proteins were assessed in the ZIC transactivation assay described in Chapter 4 (section 4.1.3). This work was performed in conjunction with Jerry Ahmed from the Arkell laboratory.

HEK293T cells were transfected with a V5-tagged ZIC3 variant and the Z3M2:β-globin reporter construct to determine whether the transactivation ability is altered. Deletion of the ZOC or the ZFNC domain led to significantly increased stimulation of the Z3M2:β-globin reporter compared to wildtype ZIC3 (Figure 5.9), suggesting these domains contain repressive functions at this isolated DNA element. Conversely, when the entire N-terminal region prior to the ZFD was deleted, the protein could no longer stimulate a transcriptional response at the reporter (Figure 5.9). This suggests that the ZFD is not sufficient to elicit a transcriptional response in isolation at this reporter element and that there are N-terminal regions that are necessary for normal transcriptional activation.

The SANC and SACC deletions could still activate transcription at the Z3M2:β-globin reporter, however this was at a reduced level compared to wildtype ZIC3 (Figure 5.9). Both these regions are therefore implicated in transactivation. Removal of these domains, however, does not completely abolish transcriptional activation, suggesting the transactivation domain is not restricted to one region of the protein. Removal of the alanine repeat did not significantly alter the proteins ability to activate transcription, however deletion of the Histidine repeat drove overstimulation at the Z3M2:β-globin reporter suggesting the Histidine also has a repressive role (Figure 5.9). Missense mutations affecting the second cysteine of each zinc finger were also assessed for their transactivation abilities. Each zinc finger mutant was unable to elicit a transcriptional response at the reporter, confirming the importance of the ZFD in transactivation.
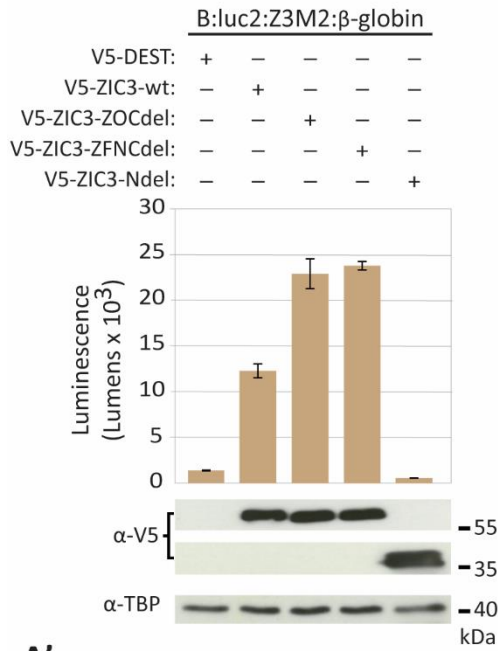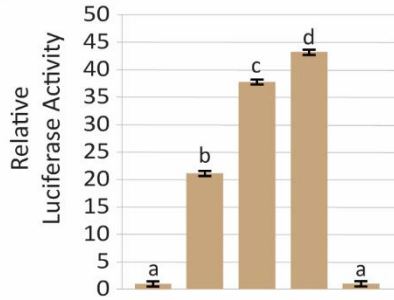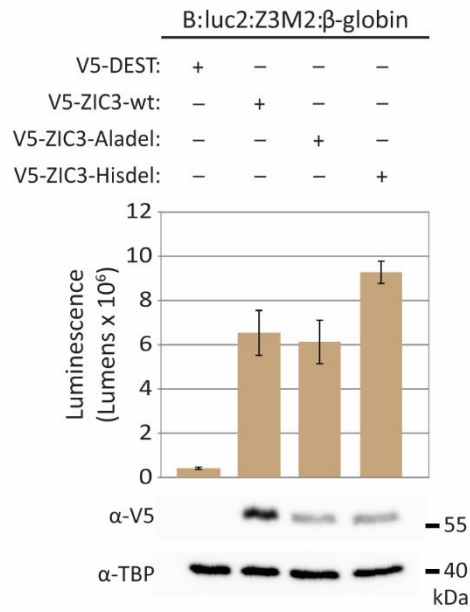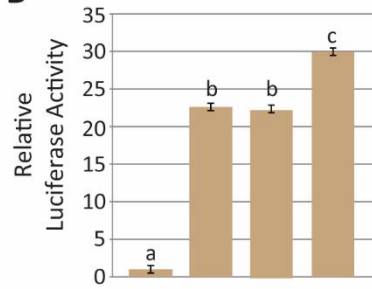
166
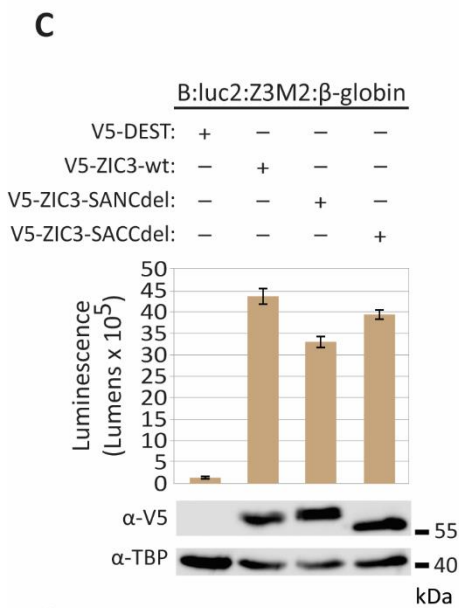
Figure 5.9: Evolutionary conserved ZIC3 mutants have reduced ability to transactivate gene expression at a ZRE. HEK293T cells were transfected with the reporter construct B:luc2:Z3M2- β-glo (containing a ZRE) and a V5-tagged ZIC3 expression plasmid shown. Luminescence was measured 24 hours post-transfection in each of three internal replicate samples. **(A-D)** One representative experiment with error bars representing the SD between the three internal replicates. A western blot was performed to confirm expression of tra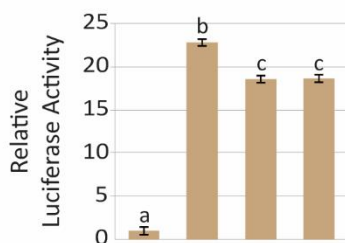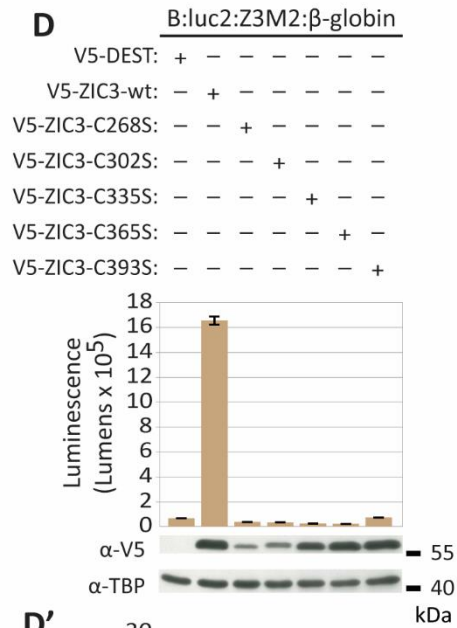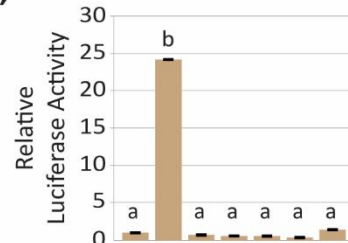nsfected proteins shown below and α-TBP was used as a loading control. **(A'-D')** Pooled data from three independent repeats. Values depict the mean relative luciferase values. Error bars depict SEM. a Different letters denote statistical differences of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post hoc test, where a is statistically significant to b etc. Transactivation experiments performed by (Panel A) or in conjunction with (Panel C, D) J. Ahmed, Arkell laboratory.

5.3 Discussion

The ZIC3 transcription factor is required at early embryogenesis to establish the L-R axis and prevent Heterotaxy in man and mouse. Recent advances in understanding ZIC3 molecular function demonstrate that ZIC3 can act as a classical transcription factor as well as a Wnt co-repressor (Ali et al. 2012; Houtmeyers et al. 2013). Consistent with this latter role, the murine *Zic3* null phenotype can be interpreted as a Wnt gain-of-function effect. The ZIC3 Heterotaxy-associated variant proteins (investigated in Chapter 4), however, appear to alter both transcription factor and co-factor functions. One way to clarify whether one or the other molecular function is critical to prevent Heterotaxy is to produce a murine allele that disrupts one but leaves the other intact. To further the development of such an allele, in this Chapter I undertook a directed search for mutations in *ZIC3* that may compromise one function but not the other. Several conserved domains were selectively deleted from the ZIC3 protein and point mutations were introduced into the ZIC3 ZFD. Molecular assays were used to assess the role of these domains in each function, including their ability to inhibit Wnt signalling in a cell based TOPflash assay, interact with TCF7L1 via a BiFC assay and activate transcription at a ZRE. From this functional analysis, I identified several point mutations within the ZIC3 ZFD that abolish transactivation activity but retain Wnt repressive function. These variants affect a key cysteine residue of each zinc finger and are comparable to a missense mutation affecting the fourth zinc finger of ZIC2 that similarly affects the transcription factor but not co-factor role of the protein (Brown et al. 2005). Overall, these mutations will be useful for determining whether the transactivation function of ZIC3 is necessary to prevent Heterotaxy by creating mouse models containing these variants and assessing the resulting phenotype.

5.3.1 Separating transcription factor and co-factor functions of ZIC3

Analysis of protein variants in this chapter showed that transcription factor and co-factor functions can indeed be separated within the ZIC3 protein. Several *ZIC3* variant proteins were found to abolish transactivation ability, yet retain Wnt repressive function and TCF7L1 interaction (Table 5.1). For example, transactivation function alone was disrupted when single missense mutations were introduced that affect key residues within the ZFD. Most notably, when the second cysteine (C) of ZF2 (C302) was substituted for a serine (S) residue, the ZIC3 protein was not able to transactivate at the ZRE, yet retained Wnt inhibition in the TOPflash assay and normal affinity for TCF7L1 in the BiFC assay. Therefore, the only function affected by this mutation was the ability to function as a transcription factor. Similarly, substituting the second cysteine of ZF1 (C268), ZF3 (C335) or ZF5 (C393) for a serine residue also abolished transactivation ability while retaining Wnt inhibition with very minor increased affinity for TCF7L1 protein (Table 5.1). Given the aim of understanding which of these functions

Table 5.1: Summary of functional properties of targeted *ZIC3* variants

		Function assessed		
		Transcription factor	**Cofactor**	
		Transactivation (Z3M2)	TOPflash	BiFC
ZIC3 variant	Empty vector	-	+++++	-
	wt	++++	+	++
	ZOCdel	++++++	+	++
	ZFNCdel	+++++	+	+
	N'del	-	++++++	++++
	SANCdel	+++	+	++
	SACCdel	+++	+	++++
	Aladel	++++	+	+
	Hisdel	+++++	+	+
	C268S (ZF1)	-	+	+++
	C302S (ZF2)	-	+	++
	C335S (ZF3)	-	+	+++
	C365S (ZF4)	-	++	+
	C393S (ZF5)	-	+	+++

[1] For Transactivation and BiFC '-' denotes baseline activity and '+' to '+++++' indicates increasing activity, based on relative luminescence or fluorescence intensity within that assay.

[2] For TOPflash '+' indicates an inhibitory effect compared to '+++++', which indicates no inhibition.

prevents Heterotaxy, it is now imperative that a mouse model be created that contains one of these *ZIC3* variants, preferably C302S (ZF2), as the only function altered was the transactivation function. Analysis of this mouse phenotype (eg identification of laterality defects in developing embryos) would allow insight into whether transactivation ability is required to prevent Heterotaxy as no current *Zic3* mouse allele exists that separates these functions.

In all cases, it is not known whether the amino acid deletions or alterations change the 3-dimentional structure of the protein and/or result in a loss of DNA binding. An equivalent mutation affecting the second cysteine of the fourth zinc finger of the ZIC2 protein was shown to ablate DNA binding (Brown et al. 2005). Due to the homology of the ZIC ZFDs, it is presumed that the cysteine-to-serine mutations would also affect DNA binding of ZIC3. Therefore the loss of reporter activation could be due to a loss of DNA binding and not due to alteration to the transactivation domain. This can be confirmed through DNA binding assays such as ChIP or EMSA experiments. These DNA binding experiments could similarly be used to assess all the other mutations to determine whether it is DNA binding or the transactivation domain that is affected by the mutation.

Ideally, the mouse studies proposed above would be augmented by the production of a mouse in which transactivation is intact but Wnt inhibition is lost. However, such a mutation was not found via this study. Similar to all of the *ZIC3* variants investigated in Chapter 4, loss of Wnt inhibition was always accompanied by loss of transactivation. All deletion mutations in this Chapter, with the exception of the N-terminal deletion, were able to inhibit β-catenin mediated Wnt signalling, albeit with different interaction levels with TCF7L1. In fact, in the majority of cases, any change to the TCF7L1 interaction (with either increased or reduced affinity) did not significantly influence the inhibition ability in the TOPflash assay (Table 5.1). The N-terminal deletion was the only *ZIC3* variant investigated in this Chapter to lose Wnt repressive function, implicating regions outside the ZFD for Wnt inhibition. Interestingly, this protein could still interact with TCF7L1 in the BiFC assay, suggesting that the interaction between ZIC3 and TCF7L1 is not sufficient for Wnt repression. Instead, it is possible that the N-terminal region of ZIC3 contains uncharacterised domains necessary for repression independent of TCF7L1 binding. Under this scenario, the N-terminal region would contain domains necessary for interaction with one or more unidentified repressor molecules or repressive chromatin complexes. Given that the N-terminally deleted protein also lost transactivation ability, it is possible that the transactivation domain structures are co-located in this region. If these domains (i.e. co-repressor interaction and transactivation domain) overlap, it is possible this influences the molecular role of ZIC3 (as a transcription factor or co-factor) in a given situation. For example, the co-expression of the repressive molecule(s) could bias against ZIC3 functioning as a classical transcription factor.

5.3.2 How does ZIC3 bind TCF7L1?

Published domain mapping experiments by Pourebrahim et al. 2011, determined that the ZIC2 ZFD is necessary for binding to TCF7L2 protein. Due to the high conservation of the ZIC ZFDs, it was expected that the ZFD of ZIC3 would be necessary for the interaction between TCF7L1. However, a *ZIC3* nonsense variant associated with human Heterotaxy (investigated in Chapter 4: ZIC3-S252X) truncates just prior to the ZIC3 ZFD and is still able to interact with TCF7L1 in the BiFC assay despite lacking the entire ZFD (Chapter 4, Table 4.4). In contrast, a ZIC2 deletion protein lacking the ZFD and C-terminus can no longer interact with TCF7L2 in a co-immunoprecipitation assay when overexpressed in HEK293T cells (Pourebrahim et al. 2011). This results suggests the ZIC3 ZFD is not necessary for this protein-protein interaction and is consistent with TCF7L1 binding occurring via the N-terminus of ZIC3. To investigate this further, an inverse protein was generated that lacks the N-terminal domain of ZIC3 (i.e. contains the ZFD and C-terminus only), however, it was found that this N-terminal deletion (ZIC3-N'del) could still interact with TCF7L1 in the BiFC assay (Table 5.1). While this might suggest that both the ZFD and N-terminal regions are sufficient for TCF7L1 interaction, it may be an artefact due to an extreme effect on the structure of one or both proteins.

Alanine tract expansions are associated with pathology in *ZIC* genes (Brown et al. 2005; Cowan et al. 2014), however it is not clear what the functional role of the domain is. Removal of the alanine or histidine repeats suggests these low complexity regions are needed for correct TCF7L1 interaction (Figure 5.7). Similarly, the deletion of the ZFNC domain reduced ZIC3-TCF7L1 interaction (Figure 5.6). Coupled with the fact that loss of all three of these domains not only results in a reduced affinity of ZIC3 for the TCF7L1 protein but also drives higher reporter gene activation at a ZRE (Table 5.1), it suggests these domains either contain some repressor properties (where loss of repressor domains lead to increased gene transactivation), but also provide further support to the model where reduced affinity for the TCF7L1 protein allows ZIC3 to function at an increased rate/greater capacity as a transcription factor. Overall, further domain mapping experiments are needed to isolate the ZIC3-TCF7L1 binding domain, potentially through smaller deletions of ZIC3 protein regions or more targeted and specific amino acid substitutions. Furthermore, mutations directly targeting TCF7L1 should be assessed to identify the domain within TCF7L1 that interacts with ZIC3.

5.3.3 These studies inform the search for the transactivation domain of ZIC3

The ZFD of the ZIC proteins has been well characterised as the DNA binding domain of the protein. It is not known, however, where the transactivation domain is within the ZIC proteins. Analysis of variants investigated in this Chapter affecting domains outside of the ZFD suggest these regions are necessary for normal transactivation. For example, removal of the SANC domain, a region conserved in Subclass A ZIC proteins (ZIC1-3), could function as well as wildtype protein in both TOPflash and

BiFC assays, while had significantly reduced transactivation ability at the ZRE, suggesting this domain is necessary for reporter activation and may form part of the transactivation domain for ZIC3. Similarly, removal of the SACC domain significantly reduced reporter activation which also implicates this region in transactivation. Future studies should create a double mutant construct that lacks both the SANC and SACC domains and put it through these established cell-based assays to determine whether the loss of both these domains compounds the loss of transactivation at this reporter. This will identify whether they play a co-operative role in ZIC3 transactivation. Alternatively, as the SANC domain appears to play a greater role in transactivation at the ZRE tested in this study, this region could be further analysed by introducing missense mutations into the SANC domain, or by rearranging the domain to effectively scramble the amino acid sequence. This will assess function without deleting several amino acids and, presumably, result in a less severe protein conformational change. The function of this domain could also be tested in isolation using an artificial system such as the GAL4-UAS fusion reporter system, a common method for identifying transcriptional regulatory domains (Weston and Bishop 1989; Pan and Pei 2003; Mizugishi et al. 2004). In this gain-of-function system, specific protein regions are fused to the DNA binding domain of the yeast transcription factor GAL4 and transfected into mammalian cells along with a luciferase reporter (Ma 2018). The GAL4-DNA binding domain interacts with the luciferase reporter and if the transfected fusion protein contains sufficient transcriptional regulatory domains, the reporter will be activated and gene expression quantified.

In summary, I have shown that mutations outside of the ZFD either reduce or increase transactivation ability of ZIC3, suggesting multiple regions combine to form the transactivation domain. Conversely, Wnt inhibition was relatively refractory to mutation and the experiments conducted here did not localise this function to a specific protein region. Importantly, these assays demonstrate the transcription factor and co-factor functions of ZIC3 can be separated and identifies mutations within the ZIC3 ZFD that affect transactivation but not Wnt repression. Therefore, a murine allele can be created to investigate whether the transactivation function of ZIC3 is required to prevent the Heterotaxy disorder, whilst further functional studies are required to identify a mutation that disrupts Wnt repression without affecting transactivation.

Chapter 6: Conclusion and Future Perspectives

The overall aim of this project was to investigate the molecular mechanism underlying *ZIC3*-associated Heterotaxy, with the hypothesis that ZIC3 interacts with TCF7L1 to repress canonical Wnt signalling during mammalian gastrulation to enable correct development of the L-R axis. At the onset of this project, there was evidence to suggest ZIC3 has a role in restricting Wnt target gene expression in the developing embryo, as loss-of-function alleles of *Zic3* resemble Wnt gain-of-function phenotypes in mouse studies. However, there was a lack of understanding regarding the molecular mechanism required for ZIC3 inhibition of this signalling pathway, and if (and how) elevated canonical Wnt signalling underlies human cases of Heterotaxy.

The data presented in this thesis significantly advances our understanding of ZIC3's molecular role in the Wnt pathway and identifies molecular functions affected by the ZIC3 Heterotaxy-associated protein variants. In Chapter 3, a variety of molecular assays were used to show that ZIC3 and TCF7L1 proteins are capable of direct protein-protein interaction when overexpressed in HEK293T cells. Coupled with the fact that these genes, when mutated, give rise to the same phenotype and that both genes have overlapping expression patterns during mammalian gastrulation, I present a model in which ZIC3 interacts with TCF7L1 during gastrulation to inhibit Wnt signalling and ultimately prevent L-R axis defects. Overall, the analysis conducted in this thesis provides a prime opportunity for further study of ZIC3 function in the Wnt pathway and establishes a new understanding of ZIC3-associated Heterotaxy phenotypes.

6.1 Future experiments to identify the ZIC3-TCF7L1 repressive complex

Combined analysis from Chapter 4 and 5 suggests that ZIC3 must form a repressive complex not only with TCF7L1, but with additional protein(s) or repressive chromatin complexes to inhibit the transcription of Wnt target genes (Figure 6.1). Whether this additional repressive machinery interacts solely with ZIC3, or with both ZIC3 and TCF7L1 (i.e. has both ZIC3 and TCF7L1 binding domains) remains unknown, however, the experiments presented here warrant the search for this protein complex. One way this complex can be identified is through a CRISPR mutagenesis screen. For example, a CRISPR guide library can be employed to target and knockout HEK293T genes (such as the commercially available Brunello library, which targets 19,114 genes in the human genome via 76,441 guide RNAs; Doench et al. 2016) under traceable conditions. These screens are well-established means for identifying currently unknown interactions, as long as good reporter systems are in place. In such an assay, the TOPflash system, described and used in this thesis (along with transiently transfected ZIC3 protein), can be used as the reporter to determine which genes/proteins are required for ZIC3
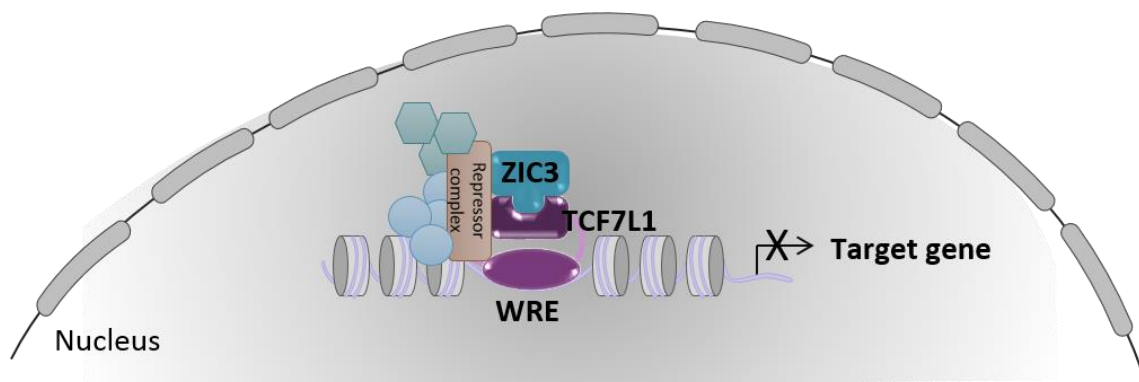
Figure 6.1 Final model for ZIC3 dependent repression of Wnt target genes. Molecular analysis strongly suggests that ZIC3 interacts with the TCF7L1 transcription factor to repress the transcription of Wnt target genes. Gene repression is likely accompanied by additional unknown repressor complexes that interact with the N-terminus of ZIC3.

dependent repression (with the aim to identify a mutation that directly disrupts the ZIC3 dependent inhibition). Once a target(s) is identified, a secondary screen can be performed to validate the gene(s) of interest and further protein binding assays to confirm an interaction between ZIC3 and/or TCF7L1.

6.2 Canonical Wnt inhibition is a key player in the prevention of human Heterotaxy

All ZIC3 Heterotaxy-associated protein variants investigated in this thesis displayed altered Wnt repression, yet the proteins also displayed absent or reduced transcription factor function. Consequently, it remains unknown which function is required to prevent Heterotaxy, and remains possible both functions are necessary to prevent the disorder. A key detail remains, however, in which transactivation is never solely affected in the Heterotaxy variants. Zinc finger variants investigated in Chapter 5, such as C302S, result in loss of transactivation and retain Wnt repression, demonstrating protein function can be separated. On the other hand, human variants assessed in Chapter 4 impair both functions and lead to Heterotaxy. Logically, one would have to assume that if the loss of transactivation alone was enough to cause Heterotaxy in humans, then *ZIC3* variants that affect transactivation but retain Wnt repression (such as C302S), would appear in the Heterotaxy cohort. However, this is not the case as all the human variants have altered Wnt inhibition in parallel with transactivation function. This ultimately makes ZIC3's role in the Wnt pathway the limiting factor and highly implicates it in Heterotaxy over the transcription factor role. Alternatively, Heterotaxy may only result as a combined effect when both functions are impaired. Although these molecular functions can be separated by mutating individual zinc fingers of the ZFD, a *ZIC3* variant has not yet been identified that alters Wnt repression whilst keeping transactivation intact and indeed, it is unknown whether such a variant will exist. If this mutation (or a variant of TCF7L1 that disrupts the ZIC3-TCF7L1 interaction) is identified, comparison of each allele individually as well as a double mutant (a cross with both alleles) mouse strain will identify whether the transactivation role, the cofactor role, or both functions are required for the correct development of the L-R axis.

6.3 Effect of node development (and impact of L-R axis formation) in *ZIC3* Heterotaxy variants
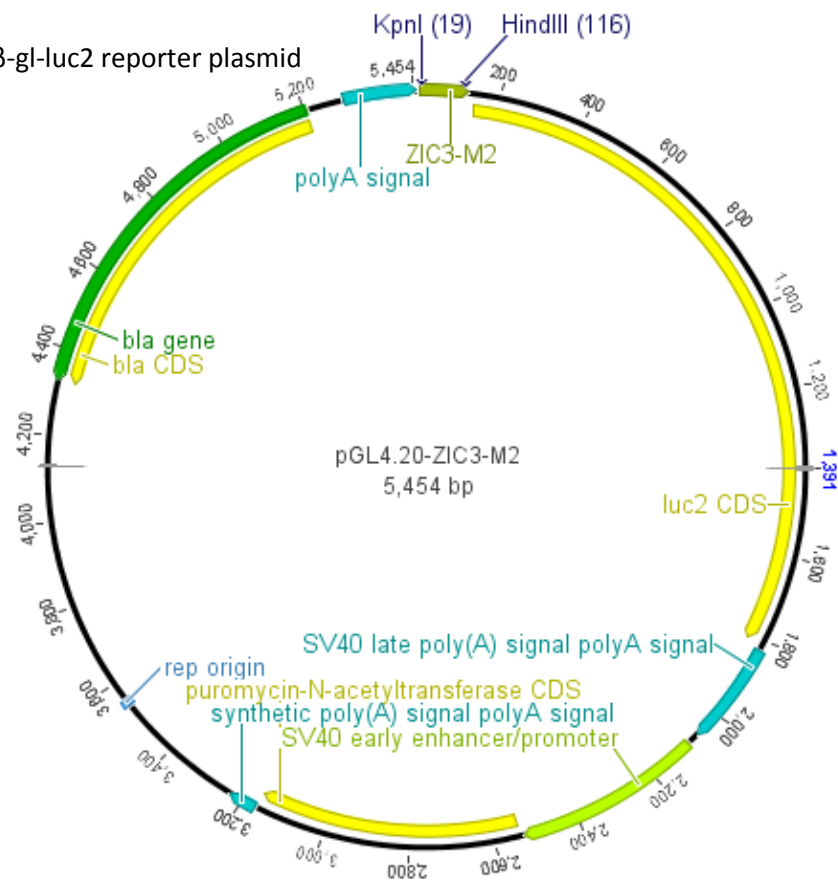
Despite *Zic3* being linked to Heterotaxy more than 20 years ago, the underlying cellular mechanisms relating to the correct development of cardiac structures and establishment of the L-R axis still remain unsolved. The work presented here, however, in combination with work from other members of the Arkell laboratory, extends our knowledge of ZIC3 function during gastrulation and supports a model for ZIC3's role in the correct establishment of the L-R axis via inhibition of Wnt signalling through the following steps: 1) ZIC3 and TCF7L1 interact to inhibit Wnt signalling during gastrulation. 2) This leads to the correct formation of the embryonic node (via an unresolved mechanism) followed by 3) the

correct establishment of the L-R axis. The Arkell laboratory's work using mouse models and cell-based assays, in conjunction with Xenopus gain-of-function studies suggesting that Zic3 regulates organiser and notochord formation by repressing β-catenin mediated Wnt signalling (Fujimi et al. 2012), leads us to propose a model underlying *ZIC3* Heterotaxy associated cases. Firstly, elevated canonical Wnt signalling results from a lack of ZIC3-TCF7L1 dependent repression during gastrulation. This elevated Wnt signalling results in downstream effects such as node abnormalities. Consequently, it is assumed that the dysmorphic node disrupts the leftward flow of node fluid required for L-R determination and subsequently, the L-R axis is not correctly established resulting in CHDs and Heterotaxy. Importantly the work presented here is one of the first to link elevated Wnt signalling to human cases of Heterotaxy and this model not only supports the role of ZIC3 in the prevention on Heterotaxy, but also implicates Wnt genes and Wnt regulators in the development of Heterotaxy and therefore provides new directions for screening patients and explaining unsolved cases.
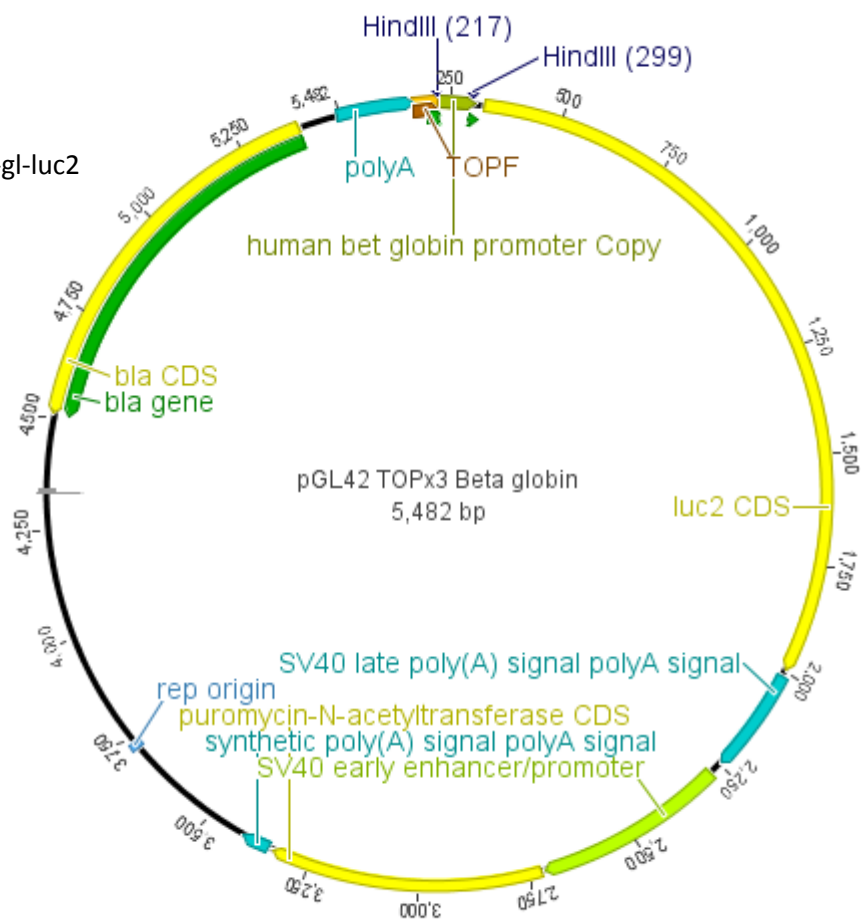
To complement my molecular analysis regarding ZIC3 protein function, this study provides support for several areas of future *in vivo* work regarding L-R defects and the Wnt pathway. For example, although there is evidence that *Zic3* loss-of-function alleles and Wnt gain-of-function alleles produce abnormal shaped nodes during gastrulation, it is unknown what molecular events result in the irregular node shape, nor is it known if these dysmorphic nodes are able to support a leftward flow of extracellular fluid and ultimately cause L-R defects. Future analysis could investigate whether the *ZIC3* Heterotaxy variants that alter ZIC3 function also result in abnormal node development and if so, determine whether this node dysmorphology results in defective fluid flow across the node causing the Heterotaxy phenotype.
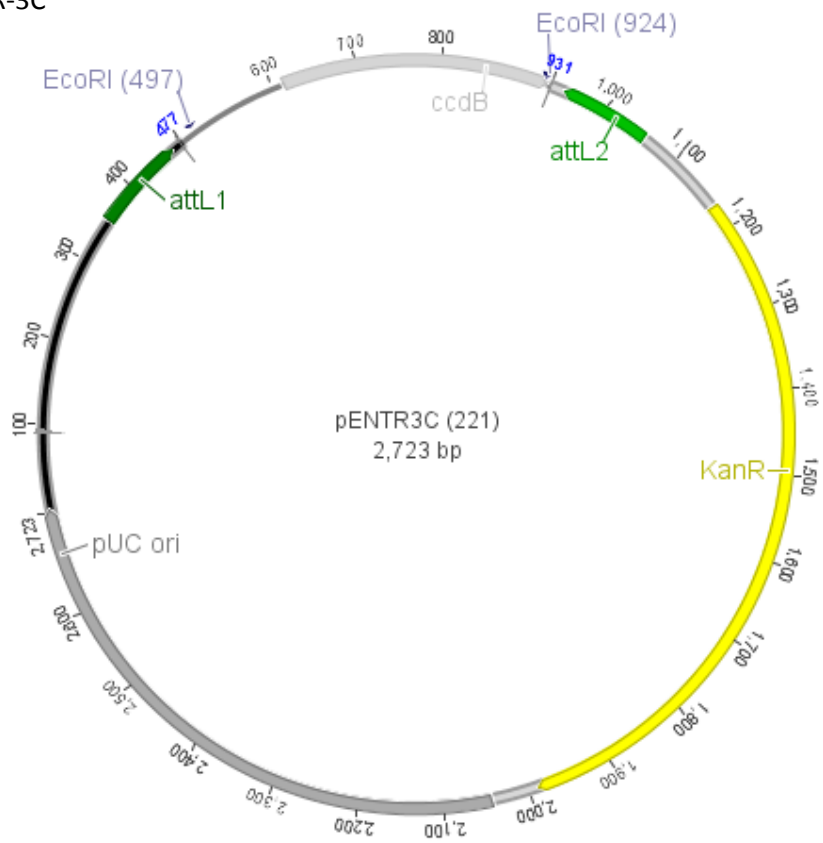
Appendix A: Plasmid Maps
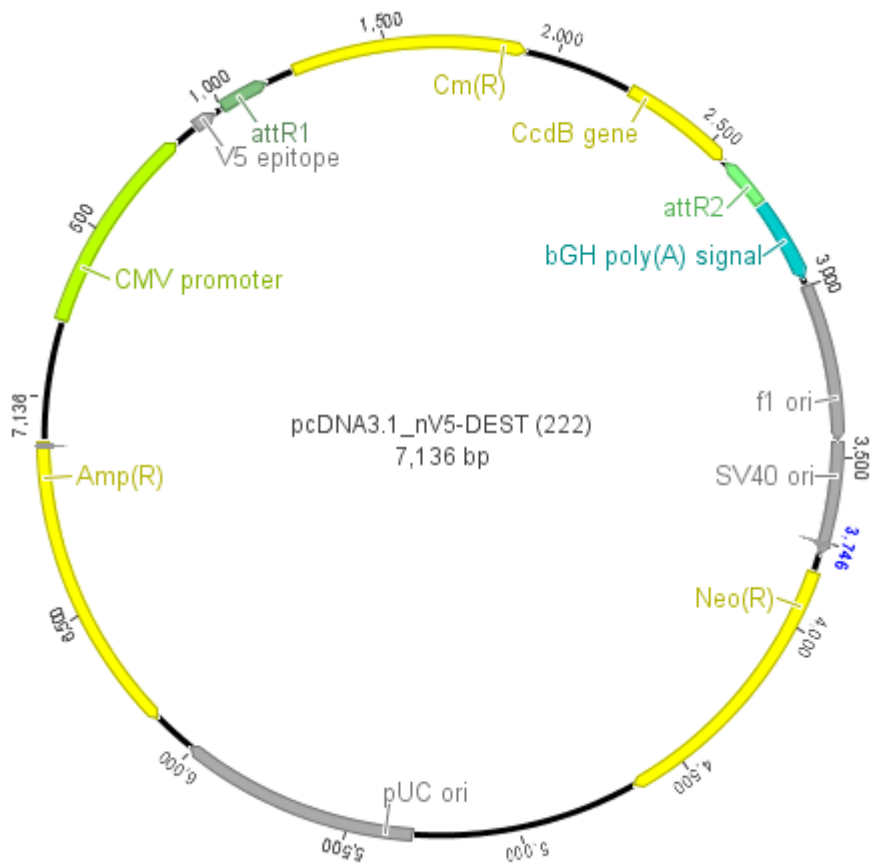
1. P(ZM2)6-β-gl-luc2 reporter plasmid

2. pTOP-β-gl-luc2

3. pENTR-3C

4. V5-Dest

5. V1-ORF

pDEST-V1-ORF (466)
7,215 bp

- CmR
- lac UV5 promoter
- attR1
- V1 CDS
- SacI (928)
- SacI (819)
- T7
- CMV enhancer&promoter
- AmpR promoter
- AmpR
- Ori
- SV40 Poly (A) signal
- BleoR
- SV40 promoter
- F1 Ori
- bGH Poly (A) signal
- attR2
- ccdB

6. ORF-V1

ORF_V1 -1
5,438 bp

- ClaI (2,650)
- attR2
- V1
- XbaI (3,164)
- AmpR
- CMV promoter
- attR1

7. V2-ORF

8. V2-Tcf7l1

9. Myc-Tcf7l1

pcDNA3 Myc_Tcf7l1 (pBM58)
7,432 bp

10. pENT3C-Tcf7l1

pENT3C Tcf7l1
4,217 bp

11. V5-β-catenin

pcDNA3.1-hCTNNB1-V5 (471) b-cat WT
7,831 bp

T7
CMV promoter
CMV enhancer
AmpR
ori
NeoR/KanR
SV40 promoter
bGH poly (A) signal
V5 tag
B-catenin CDS (Human)

12. pΔN89- β-catenin

pcDNA3.1-hCTNNB1-V5 (471) b-cat N89
7,831 bp

T7
CMV promoter
CMV enhancer
AmpR
ori
NeoR/KanR
SV40 promoter
bGH poly (A) signal
V5 tag
B-catenin CDS (Human)

13. V1-ZIC3-wt

14. V5-ZIC3-wt

Appendix B: *ZIC* expression is low in HEK cell lines

Genevestigator was used to investigate the level of *ZIC* gene expression in cell lines. ZIC1, ZIC3, ZIC4 and ZIC5 have low-medium expression in HEK cell lines, whilst ZIC2 expression is in the high range.

References

Adli M (2018) The CRISPR tool kit for genome editing and beyond. Nat. Commun.

Ahmed J (2015) Mapping Domains for ZIC3 Molecular Function.

Ahmed JN, Ali RG, Warr N, et al (2013) A murine Zic3 transcript with a premature termination codon
    evades nonsense-mediated decay during axis formation. Dis Model Mech 6:755–767. doi:
    10.1242/dmm.011668\rdmm.011668 [pii]

Ali RG, Bellchambers HM, Arkell RM (2012) Zinc fingers of the cerebellum (Zic): Transcription factors
    and co-factors. Int. J. Biochem. Cell Biol. 44:2065–2068.

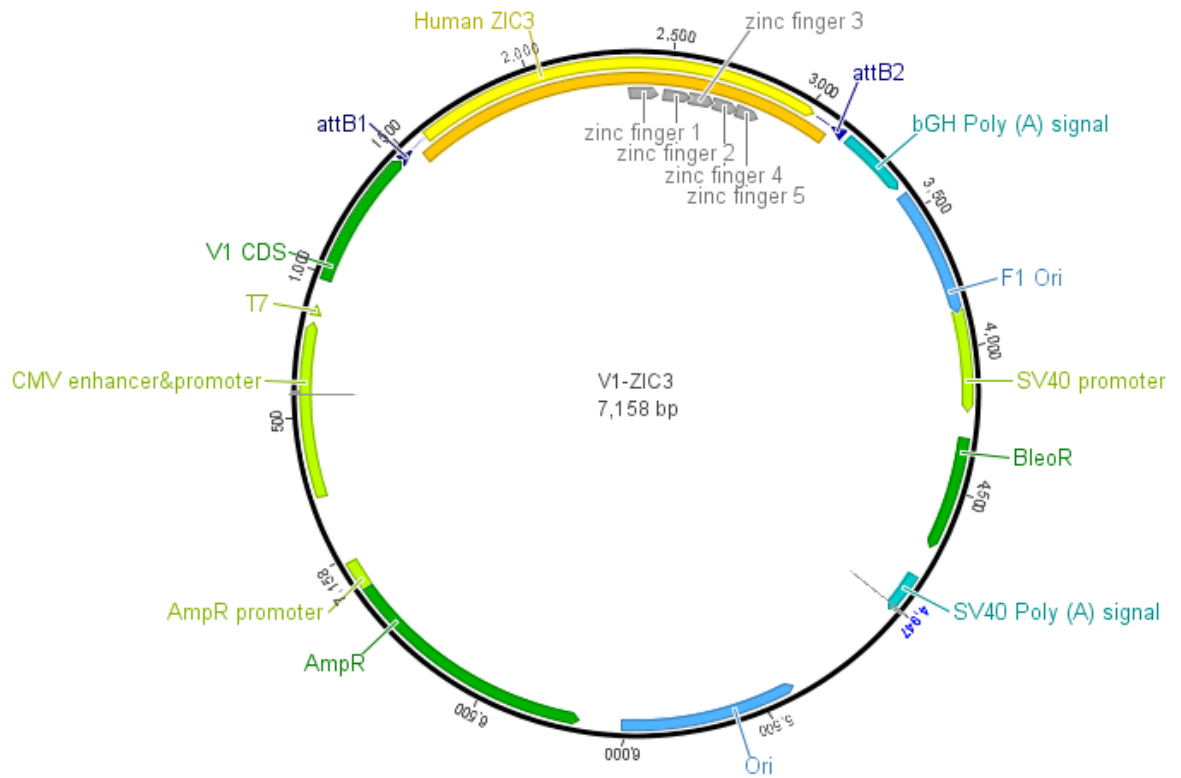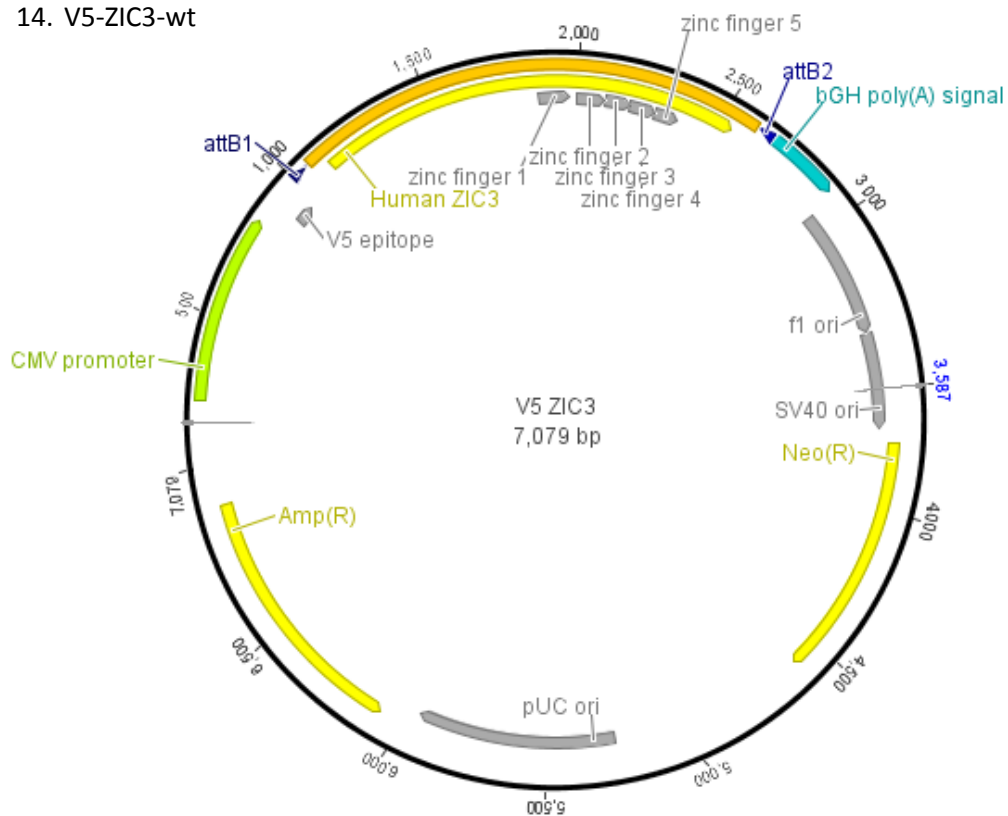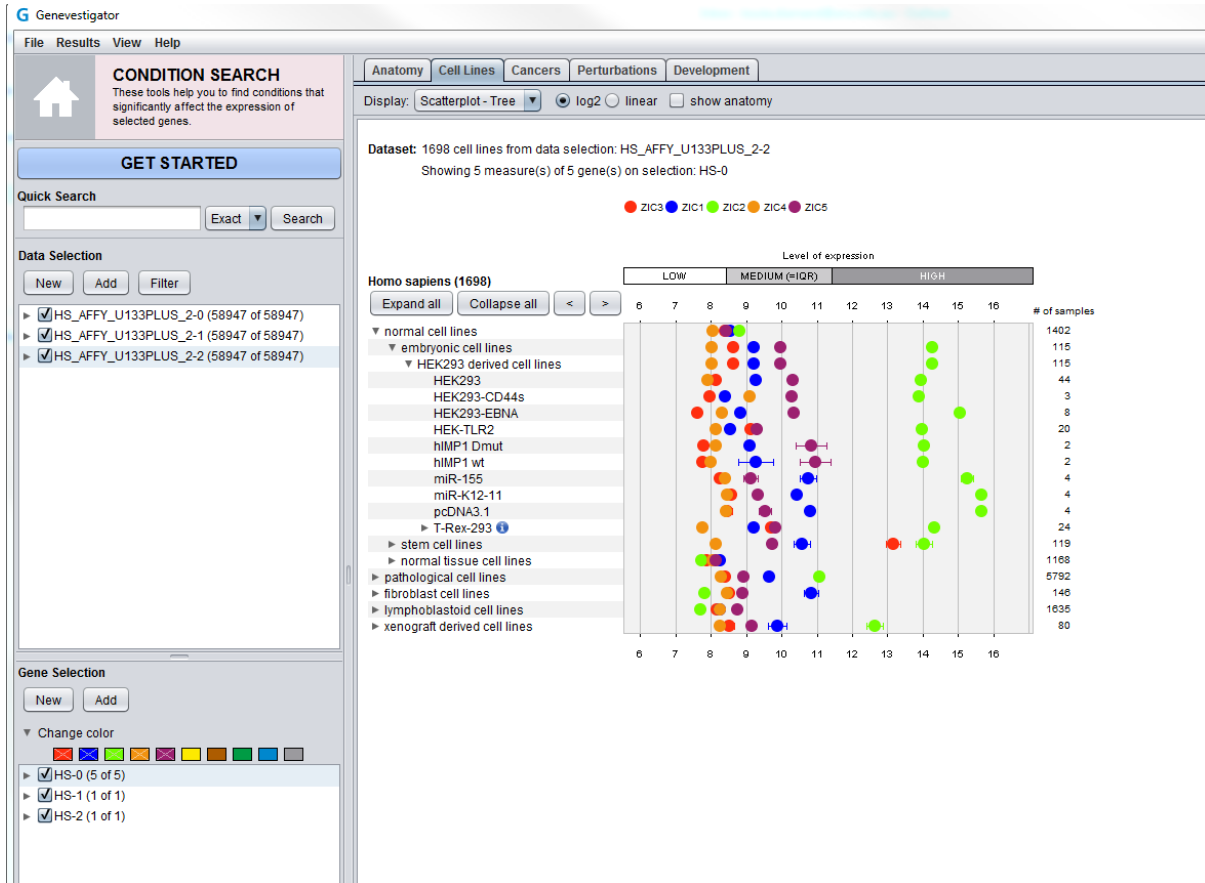Alzahrani A (2017) Identification and Characterisation of Genes which are Critical for Murine
    Gastrulation.

Arce L, Yokoyama NN, Waterman ML (2006) Diversity of LEF/TCF action in development and disease.
    Oncogene 25:7492–7504. doi: 10.1038/sj.onc.1210056

Arkell RM, Fossat N, Tam PPLP (2013) Wnt signalling in mouse gastrulation and anterior
    development: new players in the pathway and signal output.

Aruga J (2004) The role of Zic genes in neural development. Mol Cell Neurosci 26:205–221. doi:
    10.1016/j.mcn.2004.01.004

Aruga J, Kamiya A, Takahashi H, et al (2006) A wide-range phylogenetic analysis of Zic proteins:
    Implications for correlations between protein structure conservation and body plan complexity.
    Genomics 87:783–792. doi: 10.1016/j.ygeno.2006.02.011

Aruga J, Minowa O, Yaginuma H, et al (1998) Mouse Zic1 Is Involved in Cerebellar Development.

Aruga J, Mizugishi K, Koseki H, et al (1999) Zic1 regulates the patterning of vertebral arches in
    cooperation with Gli3. Mech Dev 89:141–50.

Aruga J, Nagai T, Tokuyama T, et al (1996a) The Mouse Zic Gene Family: Homologues of the
    Drosophila pair-rule gene odd- paired. J Biol Chem 271:1043–1047.

Aruga J, Tohmonda T, Homma S, Mikoshiba K (2002) Zic1 Promotes the Expansion of Dorsal Neural
    Progenitors in Spinal Cord by Inhibiting Neuronal Differentiation. Dev Biol 244:329–341. doi:
    10.1006/dbio.2002.0598

Aruga J, Yokota N, Hashimoto M, et al (1994) A novel zinc finger protein, zic, is involved in neurogenesis, especially in the cell lineage of cerebellar granule cells. J Neurochem 63:1880–1890. doi: 10.1046/j.1471-4159.1994.63051880.x

Aruga J, Yozu a, Hayashizaki Y, et al (1996b) Identification and characterization Zic gene family of Zic4 , a new member of the mouse. Gene 172:291–294.

Badis G, Berger MF, Philippakis AA, et al (2009) Diversity and complexity in DNA recognition by transcription factors. Science 324:1720–3. doi: 10.1126/science.1162327

Barratt K (2018) New aspects of ZIC2-associated Holoprosencephaly.

Barratt KS, Glanville-Jones HC, Arkell RM (2014) The Zic2 gene directs the formation and function of node cilia to control cardiac situs. Genesis 52:626–35. doi: 10.1002/dvg.22767

Bedard JEJ, Purnell JD, Ware SM (2007) Nuclear import and export signals are essential for proper cellular trafficking and function of ZIC3. Hum Mol Genet 16:187–198. doi: 10.1093/hmg/ddl461

Behrens J, Jerchow BA, Würtele M, et al (1998) Functional interaction of an axin homolog, conductin, with β-catenin, APC, and GSK3β. Science (80-). doi: 10.1126/science.280.5363.596

Bell CJ, Dinwiddie DL, Miller NA, et al (2011) Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med. doi: 10.1126/scitranslmed.3001756

Bellchambers H (2015) Wnt responsive SUMOylation regulates ZIC protein activity to promote murine neural crest specification.

Bellchambers HM, Ware SM (2018) ZIC3 in heterotaxy. In: Advances in Experimental Medicine and Biology.

Belmont JW, Mohapatra B, Towbin JA, Ware SM (2004) Molecular genetics of heterotaxy syndromes. Curr Opin Cardiol 19:216–20.

Benedyk MJ, Mullen JR, DiNardo S (1994) odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in Drosophila embryos. Genes Dev 8:105–17.

Blank MC, Grinberg I, Aryee E, et al (2011) Multiple developmental programs are altered by loss of Zic1 and Zic4 to cause Dandy-Walker malformation cerebellar pathogenesis. Development 138:1207–16. doi: 10.1242/dev.054114

Bogani D, Warr N, Elms P, et al (2004) New semidominant mutations that affect mouse development. Genesis 40:109–117. doi: 10.1002/gene.20071

Bonnafe E, Touka M, AitLounis A, et al (2004) The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. Mol Cell Biol 24:4417–27.

Brannon M, Brown JD, Bates R, et al (1999) XCtBP is a XTcf-3 co-repressor with roles throughout Xenopus development. Development 126:3159–70.

Brantjes H (2001) All Tcf HMG box transcription factors interact with Groucho-related co-repressors. Nucleic Acids Res. doi: 10.1093/nar/29.7.1410

Brayer KJ, Segal DJ (2008) Keep Your Fingers Off My DNA: Protein–Protein Interactions Mediated by C2H2 Zinc Finger Domains. Cell Biochem Biophys 50:111–131. doi: 10.1007/s12013-008-9008-5

Brogna S, Wen J (2009) Nonsense-mediated mRNA decay (NMD) mechanisms. Nat. Struct. Mol. Biol.

Brown L, Paraso M, Arkell R, Brown S (2005) In vitro analysis of partial loss-of-function ZIC2 mutations in holoprosencephaly: alanine tract expansion modulates DNA binding and transactivation. doi: 10.1093/hmg/ddi037

Brown LY, Odent S, David V, et al (2001) Holoprosencephaly due to mutations in ZIC2: alanine tract expansion mutations may be caused by parental somatic recombination. Hum Mol Genet 10:791–796.

Brückner A, Polge C, Lentze N, et al (2009) Yeast Two-Hybrid, a Powerful Tool for Systems Biology. Int J Mol Sci 10:2763–2788. doi: 10.3390/ijms10062763

Cadigan KM, Peifer M (2009) Wnt signaling from development to disease: insights from model systems. Cold Spring Harb. Perspect. Biol.

Carrel T, Purandare SM, Harrison W, et al (2000) The X-linked mouse mutation Bent tail is associated with a deletion of the Zic3 locus. Hum Mol Genet 9:1937–1942.

Carvalho LR, Brinkmeier ML, Castinetti F, et al (2010) Corepressors TLE1 and TLE3 Interact with HESX1 and PROP1. Mol Endocrinol. doi: 10.1210/me.2008-0359

Cast AE, Gao C, Amack JD, Ware SM (2012) An essential and highly conserved role for Zic3 in left–right patterning, gastrulation and convergent extension morphogenesis. Dev Biol 364:22–31.

doi: 10.1016/J.YDBIO.2012.01.011

Catana A, Apostu AP (2017) The determination factors of left-right asymmetry disorders- a short review. Clujul Med 90:139–146. doi: 10.15386/cjmed-701

Cavallo RA, Cox RT, Moline MM, et al (1998) Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature 395:604–608. doi: 10.1038/26982

Chen L, Ma Y, Qian L, Wang J (2013) Sumoylation regulates nuclear localization and function of zinc finger transcription factor ZIC3. Biochim Biophys Acta - Mol Cell Res 1833:2725–2733. doi: 10.1016/j.bbamcr.2013.07.009

Chhin B, Hatayama M, Bozon D, et al (2007) Elucidation of penetrance variability of a *ZIC3* mutation in a family with complex heart defects and functional analysis of *ZIC3* mutations in the first zinc finger domain. Hum Mutat 28:563–570. doi: 10.1002/humu.20480

Chinnadurai G (2002) CtBP, an unconventional transcriptional torepressor in development and oncogenesis. Mol. Cell

Cimbora DM, Sakonju S (1995) Drosophila midgut morphogenesis requires the function of the segmentation gene odd-paired. Dev Biol. doi: 10.1006/dbio.1995.1171

Conlon FL, Lyons KM, Takaesu N, et al (1994) A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development 120:1919–28.

Costello I, Nowotschin S, Sun X, et al (2015) Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. Genes Dev 29:2108–22. doi: 10.1101/gad.268979.115

Cowan J, Tariq M, Ware SM (2014) Genetic and functional analyses of ZIC3 variants in congenital heart disease. Hum Mutat 35:66–75. doi: 10.1002/humu.22457

Craig NL, Nash HA (1984) E. coli integration host factor binds to specific sites in DNA. Cell. doi: 10.1016/0092-8674(84)90478-1

Croce JC, McClay DR (2008) Evolution of the Wnt pathways. Methods Mol Biol 469:3–18. doi: 10.1007/978-1-60327-469-2_1

D'Alessandro LCA, Casey B, Siu VM (2013a) Situs Inversus Totalis and a Novel *ZIC3* Mutation in a

Family with X-linked Heterotaxy. Congenit Heart Dis 8:E36–E40. doi: 10.1111/j.1747-0803.2011.00602.x

D'Alessandro LCA, Latney BC, Paluru PC, Goldmuntz E (2013b) The phenotypic spectrum of ZIC3 mutations includes isolated d-transposition of the great arteries and double outlet right ventricle. Am J Med Genet A 161A:792–802. doi: 10.1002/ajmg.a.35849

De Luca A, Sarkozy A, Consoli F, et al (2010) Familial transposition of the great arteries caused by multiple mutations in laterality genes. Heart 96:673–7. doi: 10.1136/hrt.2009.181685

Delling M, Indzhykulian AA, Liu X, et al (2016) Primary cilia are not calcium-responsive mechanosensors. Nature. doi: 10.1038/nature17426

Diamand KEM, Barratt KS, Arkell RM (2018) Overview of Rodent Zic Genes. In: Advances in experimental medicine and biology. pp 179–207

Doench JG, Fusi N, Sullender M, et al (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. doi: 10.1038/nbt.3437

Dyson HJ (2016) Making Sense of Intrinsically Disordered Proteins. doi: 10.1016/j.bpj.2016.01.030

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208. doi: 10.1038/nrm1589

El Malti R, Liu H, Doray B, et al (2016) A systematic variant screening in familial cases of congenital heart defects demonstrates the usefulness of molecular genetics in this field. Eur J Hum Genet 24:228–36. doi: 10.1038/ejhg.2015.105

Elms P, Scurry A, Davies J, et al (2004) Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. Gene Expr Patterns 4:505–511. doi: 10.1016/j.modgep.2004.03.003

Elms P, Siggers P, Napper D, et al (2003) Zic2 is required for neural crest formation and hindbrain patterning during mouse development. Dev Biol 264:391–406.

Eshelman MA, Shah M, Raup-Konsavage WM, et al (2017) TCF7L1 recruits CtBP and HDAC1 to repress DICKKOPF4 gene expression in human colorectal cancer cells. Biochem Biophys Res Commun 487:716–722. doi: 10.1016/j.bbrc.2017.04.123

Field S, Riley K-L, Grimes DT, et al (2011) Pkd1l1 establishes left-right asymmetry and physically

interacts with Pkd2. Development. doi: 10.1242/dev.058149

Fields S, Song OK (1989) A novel genetic system to detect protein-protein interactions. Nature. doi: 10.1038/340245a0

Fossat N, Jones V, Khoo P-L, et al (2011) Stringent requirement of a proper level of canonical WNT signalling activity for head formation in mouse embryo. Development 138:667–76. doi: 10.1242/dev.052803

Frank CL, Liu F, Wijayatunge R, et al (2015) Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum. Nat Neurosci 1–13. doi: 10.1038/nn.3995

Fredriksson S, Gullberg M, Jarvius J, et al (2002) Protein detection using proximity-dependent DNA ligation assays. Nat Biotechnol. doi: 10.1038/nbt0502-473

Fujimi TJ, Hatayama M, Aruga J (2012) Xenopus Zic3 controls notochord and organizer development through suppression of the Wnt/β-catenin signaling pathway. doi: 10.1016/j.ydbio.2011.10.026

Funayama N, Fagotto F, McCrea P, Gumbiner BM (1995) Embryonic axis induction by the armadillo repeat domain of β-catenin: Evidence for intracellular signaling. J Cell Biol. doi: 10.1083/jcb.128.5.959

Furushima K, Murata T, Kiyonari H, Aizawa S (2005) Characterization ofOpr deficiency in mouse brain: Subtle defects in dorsomedial telencephalon and medioventral forebrain. Dev Dyn 232:1056–1061. doi: 10.1002/dvdy.20253

Furushima K, Murata T, Matsuo I, Aizawa S (2000) A new murine zinc finger gene, Opr. Mech Dev 98:161–4.

Garber ED (1952) "Bent-Tail," A Dominant, Sex-Linked Mutation in the mouse.

Gebbia M, Ferrero GB, Pilia G, et al (1997) X-linked situs abnormalities result from mutations in ZIC3. Nat Genet 17:305–308. doi: 10.1038/ng1197-305

Gó Mez-Orte E, Sá Enz-Narciso B, Moreno S, Cabello J (2013) Multiple functions of the noncanonical Wnt pathway. Trends Genet 29:545–553. doi: 10.1016/j.tig.2013.06.003

Grinberg I, Northrup H, Ardinger H, et al (2004) Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. Nat Genet 36:1053–5. doi: 10.1038/ng1420

GrüNeberg H (1963) The pathology of development: a study of inherited skeletal disorders in animals.

Gurumurthy CB, Lloyd KCK (2019) Generating mouse models for biomedical research: technological advances. Dis Model Mech. doi: 10.1242/dmm.029462

Haegel H, Larue L, Ohsugi M, et al (1995) Lack of beta-catenin affects mouse development at gastrulation. Development 121:3529–37.

Hamada F, Bienz M (2004) The APC tumor suppressor binds to C-terminal binding protein to divert nuclear β-catenin from TCF. Dev Cell. doi: 10.1016/j.devcel.2004.08.022

Hatayama M, Aruga J (2010) Characterization of the tandem CWCH2 sequence motif: a hallmark of inter-zinc finger interactions.

Hatayama M, Tomizawa T, Sakai-Kato K, et al (2008) Functional and structural basis of the nuclear localization signal in the ZIC3 zinc finger domain. Hum Mol Genet 17:3459–3473. doi: 10.1093/hmg/ddn239

He TC, Sparks AB, Rago C, et al (1998) Identification of c-MYC as a target of the APC pathway. Science 281:1509–12.

Herrera E, Brown L, Aruga J, et al (2003) Zic2 patterns binocular vision by specifying the uncrossed retinal projection. Cell 114:545–57.

Herrmann C, Horn G, Spaargaren M, Wittinghofer A (1996) Differential interaction of the Ras family GTP-binding proteins H-Ras, Rap1A, and R-Ras with the putative effector molecules Raf kinase and Ral-Guanine nucleotide exchange factor. J Biol Chem. doi: 10.1074/jbc.271.12.6794

Hicks GR, Raikhel N V. (2003) Protein Import into the Nucleus: An Integrated View. Annu Rev Cell Dev Biol. doi: 10.1146/annurev.cb.11.110195.001103

Hikasa H, Sokol SY (2013) Wnt signaling in vertebrate axis specification. Cold Spring Harb Perspect Biol 5:a007955. doi: 10.1101/cshperspect.a007955

Hinde E, Cardarelli F, Digman MA, Gratton E (2010) In vivo pair correlation analysis of EGFP intranuclear diffusion reveals DNA-dependent molecular flow. Proc Natl Acad Sci. doi: 10.1073/pnas.1006731107

Hinde E, Pandžić E, Yang Z, et al (2016) Quantifying the dynamics of the oligomeric transcription factor STAT3 by pair correlation of molecular brightness. Nat Commun 7:11047. doi: 10.1038/ncomms11047

Hirokawa N, Tanaka Y, Okada Y, Takeda S (2006) Nodal Flow and the Generation of Left-Right Asymmetry. Cell 125:33–45. doi: 10.1016/J.CELL.2006.03.002

Houtmeyers R, Gainkam OT, Glanville-Jones HA, et al (2016) Zic2 mutation causes holoprosencephaly via disruption of NODAL signalling. Hum Mol Genet 25:3946–3959. doi: 10.1093/hmg/ddw235

Houtmeyers R, Souopgui J, Tejpar S, Arkell R (2013) The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. Cell Mol Life Sci 70:3791–811. doi: 10.1007/s00018-013-1285-5

Hrckulak D, Kolar M, Strnad H, Korinek V (2016) TCF/LEF Transcription Factors: An Update from the Internet Resources. Cancers (Basel). doi: 10.3390/cancers8070070

Hruz T, Laule O, Szabo G, et al (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinformatics. doi: 10.1155/2008/420747

Huang L, Wilkinson MF (2012) Regulation of nonsense-mediated mRNA decay. Wiley Interdiscip. Rev. RNA

Huelsken J, Vogel R, Brinkmann V, et al (2000) Requirement for β-catenin in anterior-posterior axis formation in mice. J Cell Biol. doi: 10.1083/jcb.148.3.567

Inoue T, Hatayama M, Tohmonda T, et al (2004) Mouse Zic5 deficiency results in neural tube defects and hypoplasia of cephalic neural crest derivatives. Dev Biol 270:146–162. doi: 10.1016/j.ydbio.2004.02.017

Inoue T, Ota M, Mikoshiba K, Aruga J (2007) Zic2 and Zic3 synergistically control neurulation and segmentation of paraxial mesoderm in mouse embryo. Dev Biol 306:669–684. doi: 10.1016/j.ydbio.2007.04.003

Invitrogen (2005) ProQuest ™ Two-Hybrid System A sensitive method for detecting protein-protein interactions.

Iuchi S (2001) Three classes of C2H2 zinc finger proteins. Cell. Mol. Life Sci.

Jans DA, Xiao CY, Lam MHC (2000) Nuclear targeting signal recognition: A key control point in nuclear transport? BioEssays

Jiang Z, Zhu L, Hu L, et al (2013) Zic3 is required in the extra-cardiac perinodal region of the lateral plate mesoderm for left-right patterning and heart development. Hum Mol Genet 22:879–89. doi: 10.1093/hmg/dds494

Kelly OG (2004) The Wnt co-receptors Lrp5 and Lrp6 are essential for gastrulation in mice. Development. doi: 10.1242/dev.01137

Kerppola TK (2009) Visualization of molecular interactions using bimolecular fluorescence complementation analysis: Characteristics of protein fragment complementationw. doi: 10.1039/b909638h

Klootwijk R, Franke B, van der Zee CE, et al (2000) A deletion encompassing Zic3 in bent tail, a mouse model for X-linked neural tube defects. Hum Mol Genet 9:1615–22.

Kodama Y, Hu C-D (2012) Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives. Biotechniques. doi: 10.2144/000113943

Komiya Y, Habas R (2008) Wnt signal transduction pathways. Organogenesis 4:68–75.

Korinek V, Barker N, Moerer P, et al (1998) Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet. doi: 10.1038/1270

Korinek V, Barker N, Morin PJ, et al (1997) Constitutive transcriptional activation by a β-catenin-Tcf complex in APC(-/-) colon carcinoma. Science (80-). doi: 10.1126/science.275.5307.1784

Koyabu Y, Nakata K, Mizugishi K, et al (2001a) Physical and functional interactions between Zic and Gli proteins. J Biol Chem 276:6889–92. doi: 10.1074/jbc.C000773200

Koyabu Y, Nakata K, Mizugishi K, et al (2001b) Physical and functional interactions between Zic and Gli proteins. doi: 10.1074/jbc.C000773200

Lacks S, Greenberg B (1975) A deoxyribonuclease of Diplococcus pneumoniae specific for methylated DNA.

Layden MJ, Meyer NP, Pang K, et al (2010) Expression and phylogenetic analysis of the zic gene family in the evolution and development of metazoans. Evodevo 1:12. doi: 10.1186/2041-9139-

1-12

Lee JD, Anderson K V. (2008) Morphogenesis of the node and notochord: The cellular basis for the establishment and maintenance of left-right asymmetry in the mouse. Dev. Dyn.

Leuchowius K-J, Weibrecht I, Söderberg O (2011) In Situ Proximity Ligation Assay for Microscopy and Flow Cytometry. Curr Protoc Cytom 56:9.36.1-9.36.15. doi: 10.1002/0471142956.cy0936s56

Li S, Liu S, Chen W, et al (2018) A novel ZIC3 gene mutation identified in patients with heterotaxy and congenital heart disease. Sci Rep 8:12386. doi: 10.1038/s41598-018-30204-3

Lim LS, Hong FH, Kunarso G, Stanton LW (2010) The pluripotency regulator Zic3 is a direct activator of the Nanog promoter in ESCs. Stem Cells 28:1961–1969. doi: 10.1002/stem.527

Lim LS, Loh Y-H, Zhang W, et al (2007) Zic3 is required for maintenance of pluripotency in embryonic stem cells. Mol Biol Cell 18:1348–58. doi: 10.1091/mbc.E06-07-0624

Lin AE, Ticho BS, Houde K, et al (2000) Heterotaxy: Associated conditions and hospital-based prevalence in newborns.

Liu C, Li Y, Semenov M, et al (2002) Control of β-catenin phosphorylation/degradation by a dual-kinase mechanism. Cell. doi: 10.1016/S0092-8674(02)00685-2

Liu J, Perumal NB, Oldfield CJ, et al (2006) Intrinsic disorder in transcription factors. Biochemistry. doi: 10.1021/bi0602718

Lu F-I, Thisse C, Thisse B (2011) Identification and mechanism of regulation of the zebrafish dorsal determinant. Proc Natl Acad Sci. doi: 10.1073/pnas.1106801108

Luo Z, Gao X, Lin C, et al (2015) Zic2 is an enhancer-binding factor required for embryonic stem cell specification. Mol Cell 57:685–694. doi: 10.1016/j.molcel.2015.01.007

Ma J (2018) Promoter Fusions to Study Gene Expression. In: eLS.

Ma L, Selamet Tierney ES, Lee T, et al (2012) Mutations in ZIC3 and ACVR2B are a common cause of heterotaxy and associated cardiovascular anomalies. Cardiol Young 22:194–201. doi: 10.1017/S1047951111001181

Macarthur DG, Manolio TA, Dimmock DP, et al (2014) Guidelines for investigating causality of sequence variants in human disease. Nature 508:469–476. doi: 10.1038/nature13127

MacDonald BT (2004) Hypomorphic expression of Dkk1 in the doubleridge mouse: dose dependence and compensatory interactions with Lrp6. Development. doi: 10.1242/dev.01126

Maquat LE (2005) Nonsense-mediated mRNA decay in mammals. J Cell Sci. doi: 10.1242/jcs.01701

McGrath J, Somlo S, Makova S, et al (2003) Two Populations of Node Monocilia Initiate Left-Right Asymmetry in the Mouse. Cell 114:61–73. doi: 10.1016/S0092-8674(03)00511-7

Mégarbané A, Salem N, Stephan E, et al (2000) X-linked transposition of the great arteries and incomplete penetrance among males with a nonsense mutation in ZIC3. Eur J Hum Genet 8:704–708.

Mehla J, Caufield JH, Uetz P (2015) The yeast two-hybrid system: a tool for mapping protein-protein interactions. Cold Spring Harb Protoc 2015:425–30. doi: 10.1101/pdb.top083345

Merrill BJ, Pasolli HA, Polak L, et al (2004) Tcf3: a transcriptional regulator of axis induction in the early embryo. Development 131:263–74. doi: 10.1242/dev.00935

Mészáros B, Erdös G, Dosztányi Z (2018) IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. doi: 10.1093/nar/gky384

Mészáros B, Tompa P, Simon I, Dosztányi Z (2007) Molecular Principles of the Interactions of Disordered Proteins. J Mol Biol. doi: 10.1016/j.jmb.2007.07.004

Miller KE, Kim Y, Huh W-K, Park H-O (2015) Bimolecular Fluorescence Complementation (BiFC) Analysis: Advances and Recent Applications for Genome-Wide Interaction Studies. J Mol Biol 427:2039–2055. doi: 10.1016/j.jmb.2015.03.005

Mizugishi K, Aruga J, Nakata K, Mikoshiba K (2001) Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI proteins. J Biol Chem 276:2180–2188. doi: 10.1074/jbc.M004430200

Mizugishi K, Hatayama M, Tohmonda T, et al (2004) Myogenic repressor I-mfa interferes with the function of Zic family proteins. Biochem Biophys Res Commun 320:233–240. doi: 10.1016/j.bbrc.2004.05.158

Morkel M, Huelsken J, Wakamiya M, et al (2003) -Catenin regulates Cripto- and Wnt3-dependent gene expression programs in mouse axis and mesoderm formation. Development 130:6283–

6294. doi: 10.1242/dev.00859

Mosimann C, Hausmann G, Basler K (2009) β-Catenin hits chromatin: regulation of Wnt target gene activation. Nat Rev Mol Cell Biol 10:276–286. doi: 10.1038/nrm2654

Mukhopadhyay M, Shtrom S, Rodriguez-Esteban C, et al (2001) Dickkopf1 Is Required for Embryonic Head Induction and Limb Morphogenesis in the Mouse. Dev Cell 1:423–434. doi: 10.1016/S1534-5807(01)00041-7

Munemitsu S, Albert I, Rubinfeld B, Polakis P (1996) Deletion of an amino-terminal sequence beta-catenin in vivo and promotes hyperphosporylation of the adenomatous polyposis coli tumor suppressor protein. Mol Cell Biol. doi: 10.1128/MCB.16.8.4088

Murcia NS, Richards WG, Yoder BK, et al (2000) The Oak Ridge Polycystic Kidney (orpk) disease gene is required for left-right axis determination. Development 127:2347–55.

Nagai T, Aruga J, Minowa O, et al (2000) Zic2 regulates the kinetics of neurulation. Proc Natl Acad Sci U S A 97:1618–23.

Nagai T, Aruga J, Takada S, et al (1997) The expression of the mouse Zic1, Zic2, and Zic3 gene suggests an essential role for Zic genes in body pattern formation. Dev Biol 182:299–313. doi: 10.1006/dbio.1996.8449

Nagai T, Ibata K, Park ES, et al (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat Biotechnol. doi: 10.1038/nbt0102-87

Nakamura T, Saito D, Kawasumi A, et al (2012) Fluid flow and interlinked feedback loops establish left-right asymmetric decay of Cerl2 mRNA. Nat Commun. doi: 10.1038/ncomms2319

Nauli SM, Alenghat FJ, Luo Y, et al (2003) Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. Nat Genet. doi: 10.1038/ng1076

Nierode G, Kwon PS, Dordick JS, Kwon S-J (2016) Cell-Based Assay Design for High-Content Screening of Drug Candidates. J Microbiol Biotechnol 26:213–25. doi: 10.4014/jmb.1508.08007

Nonaka S, Shiratori H, Saijoh Y, Hamada H (2002) Determination of left-right patterning of the mouse embryo by artificial nodal flow. Nature. doi: 10.1038/nature00849

Nonaka S, Tanaka Y, Okada Y, et al (1998) Randomization of left-right asymmetry due to loss of nodal

cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. Cell 95:829–37.

Nonaka S, Yoshiba S, Watanabe D, et al (2005) De novo formation of left-right asymmetry by posterior tilt of nodal cilia. PLoS Biol 3:e268. doi: 10.1371/journal.pbio.0030268

Norris DP (2012) Cilia, calcium and the basis of left-right asymmetry. BMC Biol. doi: 10.1186/1741-7007-10-102

Norton N, Robertson PD, Rieder MJ, et al (2012) Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. Circ Cardiovasc Genet. doi: 10.1161/CIRCGENETICS.111.961805

Norum JH, Bergström Å, Andersson AB, et al (2015) A conditional transgenic mouse line for targeted expression of the stem cell marker LGR5. Dev Biol. doi: 10.1016/j.ydbio.2015.05.002

Nüsslein-Volhard C, Wieschaus E (1980) Mutations affecting segment number and polarity in Drosophila. Nature 287:795–801. doi: 10.1038/287795a0

Ogura H, Aruga J, Mikoshiba K (2001) Behavioral abnormalities of Zic1 and Zic2 mutant mice: Implications as models for human neurological disorders. Behav Genet 31:317–324. doi: 10.1023/A:1012235510600

Okada Y, Takeda S, Tanaka Y, et al (2005) Mechanism of Nodal Flow: A Conserved Symmetry Breaking Event in Left-Right Axis Determination. Cell 121:633–644. doi: 10.1016/j.cell.2005.04.008

Pan GJ, Pei DQ (2003) Identification of two distinct transactivation domains in the pluripotency sustaining factor nanog. Cell Res. doi: 10.1038/sj.cr.7290193

Paulussen ADC, Steyls A, Vanoevelen J, et al (2016) Rare novel variants in the ZIC3 gene cause X-linked heterotaxy. Eur J Hum Genet 1–9. doi: 10.1038/ejhg.2016.91

Pennekamp P, Karcher C, Fischer A, et al (2002) The ion channel polycystin-2 is required for left-right axis determination in mice. Curr Biol. doi: 10.1016/S0960-9822(02)00869-2

Perantoni AO (2005) Inactivation of FGF8 in early mesoderm reveals an essential role in kidney development. Development. doi: 10.1242/dev.01945

Pérez-Pinera P, Menéndez-González M, Vega JA (2006) Deletion of DNA sequences of using a polymerase chain reaction based approach. Electron J Biotechnol. doi: 10.2225/vol9-issue5-fulltext-10

Pöpperl H, Schmidt C, Wilson V, et al (1997) Misexpression of Cwnt8C in the mouse induces an ectopic embryonic axis and causes a truncation of the anterior neuroectoderm. Development. doi: 7588062

Pourebrahim R, Houtmeyers R, Ghogomu S, et al (2011) Transcription factor Zic2 inhibits Wnt/Beta-catenin protein signaling. J Biol Chem 286:37732–37740. doi: 10.1074/jbc.M111.242826

Presman DM, Ball DA, Paakinaho V, et al (2017) Quantifying transcription factor binding dynamics at the single-molecule level in live cells. Methods 123:76–88. doi: 10.1016/j.ymeth.2017.03.014

Purandare SM (2002) Zic3 null mice model X-linked heterotaxy.

Purandare SM, Ware SM, Kwan KM, et al (2002) A complex syndrome of left-right axis, central nervous system and axial skeleton defects in Zic3 mutant mice. Development 129:2293–2302.

Quinn ME, Haaning A, Ware SM (2012) Preaxial polydactyly caused by Gli3 haploinsufficiency is rescued by Zic3 loss of function in mice. Hum Mol Genet 21:1888–96. doi: 10.1093/hmg/dds002

Rao TP, Kühl M (2010) An Updated Overview on Wnt Signaling Pathways. Circ Res 106:1798–1806. doi: 10.1161/CIRCRESAHA.110.219840

Richards S, Aziz N, Bale S, et al (2015) Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. doi: 10.1038/gim.2015.30

Roose J, Molenaar M, Peterson J, et al (1998) The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. Nature 395:608–612. doi: 10.1038/26989

Serebriiskii I, Khazak V, Golemis EA (1999) A two-hybrid dual bait system to discriminate specificity of protein interactions. J Biol Chem. doi: 10.1074/jbc.274.24.17080

Shammas SL (2017) Mechanistic roles of protein disorder within transcription. Curr. Opin. Struct. Biol.

Shen MM (2007) Nodal signaling: developmental roles and regulation. Development 134:1023–1034. doi: 10.1242/dev.000166

Söderberg O, Gullberg M, Jarvius M, et al (2006) Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat Methods 3:995–1000. doi: 10.1038/nmeth947

Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. doi: 10.1038/nrg3207

Sutherland MJ, Wang S, Quinn ME, et al (2013) Zic3 is required in the migrating primitive streak for node morphogenesis and left-right patterning. Hum Mol Genet 22:1913–23. doi: 10.1093/hmg/ddt001

Sutherland MJ, Ware SM (2009) Disorders of left-right asymmetry: Heterotaxy and situs inversus. John Wiley & Sons, Ltd

Tanaka Y, Okada Y, Hirokawa N (2005) FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left-right determination. Nature. doi: 10.1038/nature03494

Tetsu O, McCormick F (1999) β-catenin regulates expression of cyclin D1 in colon carcinoma cells. Nature. doi: 10.1038/18884

Thiel G, Lietz M, Hohl M (2004) How mammalian transcriptional repressors work. Eur. J. Biochem.

Thusberg J, Olatubosun A, Vihinen M (2011) Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat. doi: 10.1002/humu.21445

Usami N, Sekido Y, Maeda O, et al (2003) β-catenin inhibits cell growth of a malignant mesothelioma cell line, NCI-H28, with a 3p21.3 homozygous deletion. Oncogene. doi: 10.1038/sj.onc.1206533

Valenta T, Lukas J, Korinek V (2003) HMG box transcription factor TCF-4's interaction with CtBP1 controls the expression of the Wnt target Axin2/Conductin in human embryonic kidney cells. Nucleic Acids Res. doi: 10.1093/nar/gkg346

Van Genderen C, Okamura RM, Farinas I, et al (1994) Development of several organs that require inductive epithelial- mesenchymal interactions is impaired in LEF-1-deficient mice. Genes Dev. doi: 10.1101/gad.8.22.2691

Verbeek S, Izon D, Hofhuis F, et al (1995) An HMG-box-containing T-cell factor required for thymocyte differentiation. Nature. doi: 10.1038/374070a0

Ware SM (2006) DNA Mutation Analysis in Heterotaxy. In: Congenital Heart Disease. Humana Press, New Jersey, pp 247–256

Ware SM, Harutyunyan KG, Belmont JW (2006a) Heart defects in X-linked heterotaxy: Evidence for a genetic interaction of Zic3 with the Nodal signaling pathway. Dev Dyn 235:1631–1637. doi: 10.1002/dvdy.20719

Ware SM, Harutyunyan KG, Belmont JW (2006b) Zic3 is critical for early embryonic patterning during gastrulation. Dev Dyn 235:776–785. doi: 10.1002/dvdy.20668

Ware SM, Peng J, Zhu L, et al (2004) Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. Am J Hum Genet 74:93–105. doi: 10.1086/380998

Warr N, Powles-Glover N, Chappell A, et al (2008) Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. Hum Mol Genet 17:2986–96. doi: 10.1093/hmg/ddn197

Wei X, Henke VG, Strübing C, et al (2003) Real-Time Imaging of Nuclear Permeation by EGFP in Single Intact Cells. Biophys J 84:1317–1327. doi: 10.1016/S0006-3495(03)74947-9

Weibrecht I, Leuchowius K-J, Clausson C-M, et al (2014) Expert Review of Proteomics Proximity ligation assays: a recent addition to the proteomics toolbox. doi: 10.1586/epr.10.10

Wessels MW, Kuchinka B, Heydanus R, et al (2010) Polyalanine expansion in the ZIC3 gene leading to X-linked heterotaxy with VACTERL association: a new polyalanine disorder? J Med Genet 47:351–5. doi: 10.1136/jmg.2008.060913

Weston K, Bishop JM (1989) Transcriptional activation by the v-myb oncogene and its cellular progenitor, c-myb. Cell. doi: 10.1016/0092-8674(89)90405-4

Wingender E, Schoeps T, Dönitz J (2013) TFClass: an expandable hierarchical classification of human transcription factors. Nucleic Acids Res 41:D165–D170. doi: 10.1093/nar/gks1123

Wingender E, Schoeps T, Haubrock M, et al (2018) TFClass: expanding the classification of human transcription factors to their mammalian orthologs. Nucleic Acids Res 46:D343–D347. doi:

10.1093/nar/gkx987

Wu C-I, Hoffman JA, Shy BR, et al (2012) Function of Wnt/ -catenin in counteracting Tcf3 repression through the Tcf3- -catenin interaction. Development. doi: 10.1242/dev.076067

Xu W, Kimelman D (2007) Mechanistic insights from structural studies of beta-catenin and its binding partners. J Cell Sci. doi: 10.1242/jcs.013771

Xue Y, Chen Y, Ayub Q, et al (2012) Deleterious- and disease-allele prevalence in healthy individuals: Insights from current predictions, mutation databases, and population-scale resequencing. Am J Hum Genet. doi: 10.1016/j.ajhg.2012.10.015

Yoder BK, Hou X, Guay-Woodford LM (2002) The polycystic kidney disease proteins, polycystin-1, polycystin-2, polaris, and cystin, are co-localized in renal cilia. J Am Soc Nephrol. doi: 10.1097/01.ASN.0000029587.47950.25

Zeng L, Fagotto F, Zhang T, et al (1997) The mouse Fused locus encodes Axin, an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation. Cell 90:181–92.

Zhan T, Rindtorff N, Boutros M (2017) Wnt signaling in cancer. Oncogene 36:1461–1473. doi: 10.1038/onc.2016.304

Zhang Y, Niswander L Zic2 is required for enteric nervous system development and neurite outgrowth: a mouse model of enteric hyperplasia and dysplasia.

Zhu L, Belmont JW, Ware SM (2006) Genetics of human heterotaxias. Eur J Hum Genet 14:17–25. doi: 10.1038/sj.ejhg.5201506

Zhu L, Zhou G, Poole S, Belmont JW (2008) Characterization of the interactions of human ZIC3 mutants with GLI3. Hum Mutat. doi: 10.1002/humu.20606