Australian National University

RESEARCH SCHOOL OF BIOLOGY

Mapping Domains for ZIC3 Molecular Function

Jehangir Nazir Ahmed

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

November 2015

The results presented in this thesis are, except where otherwise acknowledged, my own original work. It has not been submitted, in whole or in part, for any other degree.

Jehangir Nazir Ahmed

“Read in the name of your Lord who created”

“Created man, out of a (mere) clot of congealed blood”

“Read, and your Lord is the most Generous”

“Who taught by the pen”

“Taught man that which he knew not.”

(Holy Qur’an 96:1-5)

Acknowledgments

I am thankful to the Almighty for making me capable of pursuing a Doctor of Philosophy (PhD). The PhD journey was made possible due to the support and guidance of several people and I take this opportunity to recognize their contributions.

First and foremost I am grateful to my supervisor, Ruth Arkell, for the opportunity to conduct research under her tutelage and providing guidance and ingenious advice. I am indebted to her for funding my PhD and the assistance and feedback she provided for preparation of seminars and writing of this thesis. I have indeed learnt several skills while working under her supervision, but if there was one thing that I had to take away, it will be her insistence to, “make smart choices.”

Members of the Arkell lab (past and present) have contributed immensely to my development as a research scientist. Most notably, I am grateful to Radiya Ali for paving the way for my research career, and Nicole Thomsen for her helpful and dependable presence in the lab. My gratitude is extended to other lab members (Alaa, Koula, Kristen, and Hannah) for their support and assistance during my time in the lab, particularly Koula who time and again proved to be a gem.

I owe everything to the two people who form the biological basis of my existence. My parents have ensured provision of everything I needed to reach here. Moreover, the numerous lessons they have taught me has made my life a pleasurable and gratifying experience. I am grateful to my father for teaching me the value of hard work and my mother for instilling steadfastness in me. My siblings, I thank you for your constant encouragement. This PhD would not have been possible without my mentor, Manzoor, and his daughter (Maha): I thank you for inspiring me and pushing me to press harder.

Lastly I thank my darling wife, Natasha, for being a pillar of support. I consider myself lucky to have her as my partner, owing to her loving, supportive, encouraging and patient nature that has made this journey significantly easier. Her unwavering belief in my ability has allowed me to complete this journey and be the person I am today. I thank you with my heart and soul.

Abstract

The zinc finger of the cerebellum (*Zic*) genes encode a family of transcriptional regulators critical for early embryogenesis. All ZIC proteins contain a zinc finger domain (ZFD) and other evolutionary conserved regions. They are pleiotropic in nature since they can influence gene expression directly by acting as transcription factors due to their ability to bind target DNA sequences, or indirectly as co-factors by interacting with protein partners. Little is known, however, about the structural components that allow ZIC proteins to perform these functions. Among ZIC family members the protein structure of ZIC3 is relatively well characterised, yet details regarding its transactivation domain remain unknown. During embryonic development ZIC3 is involved in maintaining pluripotency of embryonic stem cells, formation of the left-right (L-R) axis and arrangement of visceral organs. Mutations in *Zic3* in humans and animal models cause congenital L-R axis defects. The work presented in this thesis maps structural domains required for ZIC3 molecular function. Characterisation of a novel allele of murine Zic3 revealed that removal of the ZFD and C-terminus renders the mutant protein functionally null and incapable of dominantly interfering with the function of other ZIC proteins. To further assess the transcription factor function of ZIC3, a new cell-based transactivation assay system using target ZIC3-DNA binding sequences was designed. This assay was used to identify regions within the ZFD and C-terminus vital for transactivation via ZIC3. In addition other evolutionary conserved domains were implicated in transactivation. This study provides a reliable and robust platform to investigate the transcription factor function of ZIC proteins and their variants.

Table of Contents

List of Figures

List of Tables

Abbreviations

ANOVA: analysis of variance

bp: base pairs

CD: circular dichroism

cDNA: complimentary deoxyribonucleic acid

CDS: coding deoxyribonucleic acid sequence

ChIP: Chromatin immunoprecipitation

CMV: *cytomegalovirus*

DBD: DNA-binding domain

DMEM: Dulbecco's Modified Eagle Medium

DNA: deoxyribonucleic acid

dpc: days post coitum

EDTA: Ethylenediaminetetraacetic acid

EGFP: enhanced green fluorescent protein

ESC: embryonic stem cell

HRP: horse radish peroxidase

HRMA: high resolution melt analysis

Ka: katun

kDa: kilo Daltons

L-R: left-right

LPM: lateral plate mesoderm

LSD: least significant difference

MCS: multiple cloning site

mRBP: messenger ribonucleic acid binding protein

mRNA: messenger ribonucleic acid

MW: molecular weight

NLS: nuclear localisation signal

NPC: nuclear pore complex

NTC: no template control

opa: *Drosophila* odd-paired

PBS: phosphate buffered saline

PBM: protein binding microarray

PCR: polymerase chain reaction

PTC: premature termination codon

PTM: post-translational modification

PVDF: polyvinylidene difluoride

qPCR: quantitative polymerase chain reaction

RT: reverse transcriptase

RNA: ribonucleic acid

RLA: relative luciferase activity

SD: standard deviation

SDS: Sodium dodecyl sulphate

SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SEM: standard error of means

SV40: simian virus 40

TBP: TATA-binding protein

TBS: Tris buffered saline

TCF: T-cell factor

Tris: Tris(hydroxymethyl)aminomethane

tRNA: transfer ribonucleic acid

TK: thymidine kinase

UAS: upstream activating sequence

WMISH: whole mount in situ hybridisation

ZF: zinc finger

ZFD: zinc finger domain

ZFNC: zinc finger N-flanking conserved

ZIC: zinc finger of the cerebellum

ZOC: Zic-Opa conserved

ZRE: ZIC-responsive element

CHAPTER 1

INTRODUCTION

1.1 TRANSCRIPTION FACTORS: KEY MEDIATORS OF BIOLOGICAL FUNCTION

An adult organism develops from a single cell, the fertilized zygote. This diploid cell has the potential to divide and grow into a multicellular organism that contains a multitude of differentiated cell types, each with its own specialized function, but with the same genomic DNA. The method by which cells with the same DNA become specialised and assemble into tissues and organs is the central problem of developmental biology. At the core of this phenomenon is the fact that specific regions in the genome (genes) code for proteins that influence cell identity. Hence by using different parts of the genome to express varying combinations (or an assortment) of proteins, cells become specialised for different functions. This differential gene expression is therefore fundamental to all cellular differentiation and tissue formation.

Differential gene expression is achieved by the combinatorial function of various classes of proteins (Kadonaga, 1998), one of which is called transcription factors (Latchman, 1990). Transcription factors are modular proteins which contain distinct DNA-binding and transcriptional regulatory domains. Typically they bind specific DNA sequences in the promoter region (upstream of the coding region of the gene) and/or distal regulatory elements through their DNA-binding domains, and interact with the basal transcriptional machinery through their transcriptional regulatory domains to promote or repress gene expression (Figure 1.1) (Mitchell *et al.*, 1989). Additionally these proteins may also have a protein-protein interaction domain that allows them to form dimers or bind with cofactor molecules (co-activator or co-repressor) required for their function (Latchman, 1997). Therefore transcription factors either act directly to activate/repress transcription, or indirectly by interacting with co-factors molecules which then allows them to interact with the transcriptional machinery. The activator or repressor functions of transcription factors are dictated by the regulatory element they are bound to, structure of the surrounding chromatin, and the type of molecules available in the environment (Boyle *et al.*, 2010). Naturally, given their role as gene-regulators, transcription factors are a

common target of mutations that result in human congenital defects and diseases (Jimenez-Sanchez *et al.*, 2001).

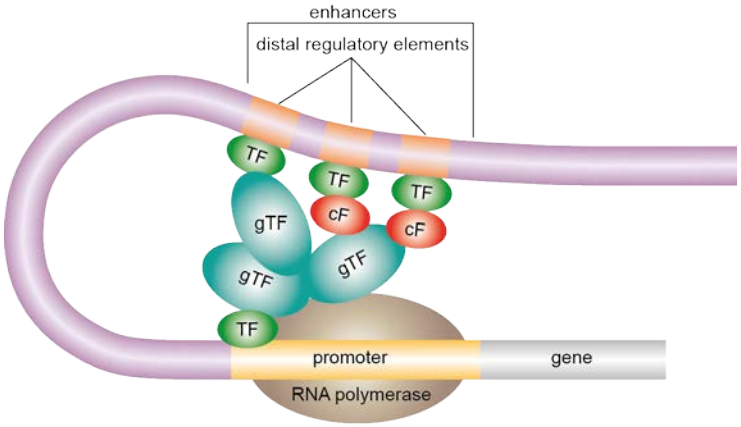


Figure 1.1: Transcription regulation via transcription factors. Transcription factors (TF) influence gene expression by binding to specific DNA sequences (promoters or enhancers). Depending on the target sequence, transcription factors interact directly or indirectly by binding to co-factor molecules (cF) with the basal transcriptional machinery [general transcription factors (gTF) and RNA polymerase] for initiation of transcription.

More than 1700 transcription factors have been identified in humans and most of these are conserved in other taxonomic groups such as yeast, plants, *Drosophila*, *C. elegans*, *Xenopus* and mice (Brivanlou *et al.*, 2002, Vaquerizas *et al.*, 2009). Transcription factors are classified based on sequence similarities and tertiary structure of their DNA-binding domains (Stegmaier *et al.*, 2004). These proteins drive the development of whole organisms, maintain physiological homeostasis and are involved in repair and regeneration (Latchman, 1997).

1.1.1 Regulation of Transcription Factor Steady State Levels and Activity

Transcription factors actively influence the rate of gene expression and it is vital their steady state levels and activity be strictly regulated. The first step in the synthesis of all proteins is conversion of the DNA code to messenger RNA (mRNA). The mRNA molecule is then processed for transport to the cytoplasm where the protein synthesizing machinery is located. The protein synthesizing machinery translates mRNA into a polypeptide chain, which may require further modification to form a mature protein. The mature protein in turn can be held in an inactive state, shuttled to and from its site of activity, modified and/or targeted for destruction. The regulation of protein steady state levels and activity can therefore occur at many different stages.

1.1.1.1 Transcription: mRNA synthesis

mRNA synthesis occurs in the nucleus, where genes are converted into mRNA via the transcriptional machinery. The main enzyme involved in transcription is RNA polymerase, which identifies specific DNA binding sites (called promoters), unwinds the double helix structure of DNA, catalyzes the formation of the mRNA molecule based on the DNA template and recognizes a termination sequence that specifies where the transcript should end (Berg *et al.*, 2002). In addition, RNA polymerase can interact with activator or repressor proteins (such as hormones, growth factors or transcription factors) to modulate the rate at which transcription takes place. In eukaryotes when DNA is transcribed into nascent mRNA, the mRNA molecule undergoes several processing steps in the nucleus before transport to the cytoplasm for translation. These include 5'-end capping, splicing, 3'-end cleavage and polyadenylation. The modifications at the 5`- and 3`-ends stabilize the nascent mRNA, protect it from degradation via exonucleases and participate in protein synthesis (Day *et al.*, 1998). Splicing involves removal of non-coding sequences (introns) from between coding sequences (exons) to form the mature mRNA transcript. If a gene consists of more than one exon, these can be differentially included or

excluded from the mature mRNA transcript via alternate splicing (Nilsen *et al.*, 2010). Hence a single gene can produce multiple mRNA transcripts. Following the formation of the mature mRNA transcript it is bound by several mRNA binding proteins (mRBPs), which interact with other proteins in the nuclear membrane to enable cytoplasmic transport.

1.1.1.2 RNA Degradation

In addition to preparation of mRNA for translation, other mechanisms can regulate mRNA stability, inevitably effecting protein steady state levels. These include:

- **miRNA mediated regulation:** miRNAs are single-stranded RNA molecules of made up of ~22 nucleotides (Bartel, 2004). In animals, these molecules have been found to have important functions in regulation of signaling pathways, apoptosis, metabolism, cardiogenesis and brain development (Kloosterman *et al.*, 2006). miRNAs typically regulate protein synthesis via mRNA degradation or translational inhibition. In order to target an mRNA molecule for degradation, miRNAs form imperfect hybrids with the 3'UTR of the target mRNA and employ exonucleases for degradation of the transcript. For translational inhibition, miRNAs bind to mRBPs that attach themselves onto the target mRNA and obstruct the binding of the translational machinery, hence inhibiting protein synthesis (Pillai *et al.*, 2007).
- **Nonsense Mediated Decay (NMD):** NMD is an evolutionary conserved quality-control process in yeast, *Drosophila*, *C. elegans*, mammals and plants, which selectively degrades mRNAs harbouring premature termination codons (PTCs) (Chang *et al.*, 2007). PTCs can arise as a result of mutations (nonsense or frame-shift) or errors during RNA processing; hence NMD protects cells from deleterious truncated proteins that may perturb the normal cellular homeostasis (Brogna *et al.*, 2009). A critical feature of NMD is that factors involved in this process are able to distinguish between PTCs and canonical stop codons. How exactly this is achieved is not completely understood, however it appears dependent upon the position of the PTC; a transcript will be committed to NMD if a PTC is located 50-55 bases upstream of the 3'-most exon-exon junction (Nagy *et al.*, 1998). Thus canonical stop codons

bypass NMD as most are located within the last exon. In addition PTCs in the last exon and genes lacking introns escape NMD (Maquat *et al.*, 2001).

1.1.1.3 Translation: Protein Synthesis

Once a mature mRNA transcript reaches the cytoplasm, it is recognized by the protein synthesizing machinery. Major components of this machinery are ribosomes (site of protein synthesis) and transfer RNA (tRNA; forms a link between nucleic acids and amino acids). Each tRNA molecule carries a particular amino acid, which is coded for by a combination of three specific nucleotides present on the mRNA transcript (Lodish *et al.*, 2007). The first amino acid of all proteins is always methionine. Hence translation is initiated when a tRNA molecule carrying methionine binds on to the corresponding nucleotides on the mRNA template. Once the amino acid is incorporated, the tRNA molecule carrying it exits the ribosome and makes way for subsequent tRNAs to bind and add relevant amino acids to the peptide chain. When tRNAs encounter a stop codon, addition of amino acids to the peptide chain halts and the nascent peptide chain is released from the ribosome to be further processed. The three-dimensional structure of proteins and their activity is dependent on the linear order of amino acids. Hence assembly of amino acids in their correct order is critical to the production of functional proteins.

1.1.1.4 Protein modifications

When peptide formation is complete at ribosomes, nascent peptides are transported to the rough endoplasmic reticulum and golgi apparatus in the cell where they undergo post-translational modifications (PTMs). PTMs involve attachment of specific chemical groups (for example phosphates, acetyl groups, ubiquitin residues, SUMO protein) to the protein (Han *et al.*, 1992). This leads to alterations in the structure of the protein and has downstream effects on the function of the transcription factor such as their interaction with other cellular proteins, localization to their site of function or degradation (Parekh *et al.*, 1997, Muratani *et al.*, 2003).

Additionally PTMs allow cells to respond to environmental stimuli via PTMs of extant proteins (Wells *et al.*, 2001).

1.1.1.5 Nuclear Localization

Although transcription factors operate in the nucleus of a cell, they are usually held in the cytoplasmic compartment of the cell and only translocated to the nucleus upon receipt of specific cellular cues (Xu *et al.*, 2004). Once activated, the proteins travel from the cytoplasm to the nucleus to reach their target DNA-binding sites. The transport of proteins occurs through openings in the nuclear envelope, known as nuclear pores (composed of nucleoporin proteins) (Cartwright *et al.*, 2000). The mechanism of transport differs among the various families of transcription factors, however in general nuclear import requires the presence of specific molecular signals (amino acid sequences in the transcription factor), mobile (shuttling) carriers and energy (Nigg, 1997). The molecular signals either directly interact with the nucleoporins or interact with other (so called) chaperone proteins that bind to nucleoporins to allow transport of the transcription factor by the consumption of energy. In addition the importance of these transport mechanisms is highlighted by many mutations resulting in human diseases and congenital defects (Brivanlou *et al.*, 2002, Ware *et al.*, 2004, Xu *et al.*, 2004).

There are several mechanisms by which transcription factors can be directed into the nucleus:

- **Nuclear localization signal-mediated nuclear import:** Previous studies investigating the biochemical events that take place during nucleocytoplasmic shuttling have characterized the specific molecular signals within transcription factors that promote nuclear import (Jans *et al.*, 2000). These molecular signals are known as Nuclear Localization Signals (NLSs), which are short stretches of basic amino acids that are either monopartite (a cluster of basic residues preceded by a helix-breaking residue) or bipartite (two clusters of basic residues separated by 9-12 residues) (Hicks *et al.*, 1995, Cokol *et al.*, 2000). They contain a high content of arginine and lysine residues that are recognized by a soluble heterodimeric carrier that consists of proteins importin-α and importin-β. The importin-α/β complex binds

to the NLSs and also associates with nucleoporins (via HEAT repeats in importin-β) allowing nuclear import of the transcription factor (Stewart, 2007).

- **Transport via chaperones:** Various transcription factors have been found to move into the nucleus in an NLS-independent manner (Xu *et al.*, 2004). These transcription factors interact with 'chaperone' proteins that contain NLSs required for nuclear import. For example, the transcription factor STAT itself does not contain an NLS. However its binding partner, interferon-γ, consists of a monopartite NLS; it is therefore through this interaction that STAT is able to translocate into the nucleus (Johnson *et al.*, 1998).
- **Transport into the nucleus via armadillo repeats:** Some transcription factors that do not contain NLSs instead possess armadillo repeats that are similar to the HEAT repeats found in importin-β. These amino acids facilitate nuclear import through recognition of nucleoporins. Armadillo repeats were first identified in the *Drosophila melanogaster* β-catenin orthologue, Armadillo, which is a classic example of a protein transported into the nucleus without a NLS (Peifer *et al.*, 1994).
- **Transport into the nucleus via sumoylation and related mechanisms:** Many transcription factors are targets for sumoylation, an event wherein small ubiquitin related modifier (SUMO) molecules are conjugated to the protein. Although the consequences of sumoylation are diverse, it can regulate the subcellular distribution of a protein. Models demonstrating the mechanisms by which sumoylation facilitates nuclear import have been suggested and in most instances appear to require a NLS (Melchior *et al.*, 2003). In some cases, however, sumoylation of the target proteins is in itself not required for nuclear localization. For example, the enzyme essential for sumoylation, Ubc9, was capable of facilitating nuclear localization of the transcription factor Vsx-1 (Kurtzman *et al.*, 2001). This occurred even when the transcription factor was not sumoylated. Nevertheless an NLS appears to be necessary for this mechanism of transport as well.

1.1.1.6 Mechanisms of nuclear export

Transcription factor function can be dampened by nuclear export. The export process is fundamentally similar to the import process, as the former also requires the presence of specific molecular signals (nuclear exit signals) or proteins that assist the transcription factor in interacting with the nuclear export machinery at the nuclear envelope (Wen *et al.*, 1995). Mutations in these signals lead to accumulation of the transcription factor in the nucleus, which continually promotes gene expression. Hence the biochemical homeostasis is disturbed causing various diseases, such as cancer (Giannini *et al.*, 2004).

1.1.1.7 Dimerization

Some transcription factors require a physical interaction with an identical molecule (homo-) or with another molecule (hetero-) to form a functional dimer in order to bind DNA (Amoutzias *et al.*, 2008). Depending on the type of transcription factor, the choice of partner and the ongoing cellular events, the dimerization process can activate or repress the activity of the transcription factor (Klemm *et al.*, 1998).

1.1.1.8 Dominant-negative interference:

This is a phenomenon where the function of a wild-type protein is abolished due to interference by a mutant isoform (formed via alternate splicing or a genetic mutation) (Herskowitz, 1987, Stamm *et al.*, 2005). The resultant protein derivative can potentially interfere in the molecular functioning of wild-type transcription factor via binding to the target DNA sites, interacting with co-factors, or formation of inhibitory heterodimers (Figure 1.2).

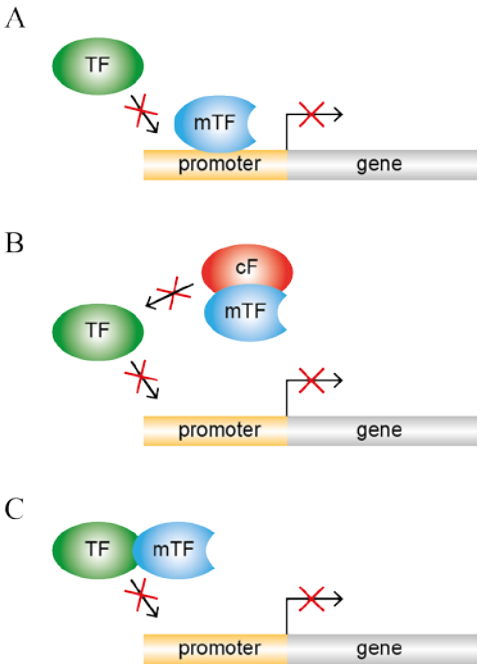


Figure 1.2: Dominant-negative interference. The mutant transcription factor (mTF) can inhibit transcription via the wild-type transcription factor (TF) by **(A)** promoter blocking, **(B)** competing for co-factor (cF), or **(C)** heterodimer formation.

1.1.1.9 Protein Degradation

Since transcription factors are highly active in contrast to other proteins in the cell, maintenance of their steady-state levels needs to be tightly regulated via keeping a balance between synthesis and degradation. This selective and programmed mechanism for protein degradation is provided by the ubiquitin-proteasome system (Desterro *et al.*, 2000). The proteasome is a multicatalytic protease complex that identifies and specifically degrades ubiquitin-tagged proteins. Ubiquitin is a highly conserved protein present in a vast variety of eukaryotic species (Hershko *et al.*, 1998), which is covalently conjugated with lysine residues on proteins that are destined for degradation via the proteasome.

1.2 THE ZIC FAMILY OF PROTEINS

Genes that encode developmental regulatory proteins are generally conserved across the invertebrate-vertebrate divide (Ohno, 1970). For example, genes involved in the development of the *Drosophila* embryo often have vital roles in development of the vertebrate embryo, even though the complexity and developmental mechanisms of the organisms are significantly different (Grinberg *et al.*, 2005). The *Zic* genes are no exception; the vertebrate *Zic* genes were discovered by homology to the *Drosophila odd-paired* (*opa*) gene (Aruga *et al.*, 1994). *Opa* itself was isolated from a genetic screen for mutations that disrupt segmentation of the *Drosophila* embryo. Flies unable to produce functional *opa* gene product lack a portion of every alternate body segment and fall into the group of mutants known as pair-rule mutants (Nusslein-Volhard *et al.*, 1980). Further analysis of *opa* mutant flies revealed it also regulates expression of genes required for the development of visceral mesoderm and midgut morphogenesis (Benedyk *et al.*, 1994).

As observed for many vertebrate genes, evolutionary genome duplication events have expanded the one *Drosophila opa* to multiple paralogs (Grinberg *et al.*, 2005). Mammals have five paralogous *Zic* genes: *Zic1*, *Zic2*, *Zic3*, *Zic4* and *Zic5*, each significantly homologous to the invertebrate *opa* gene and to their respective *Zic* orthologues. The name 'Zic' denotes zinc fingers of the cerebellum (the tissue from which they were first isolated) (Aruga *et al.*, 1994). Mutational analysis in a variety of vertebrate and mammalian organisms has shown that *Zic* genes play important roles in differentiation of ectoderm and mesoderm derived structures during embryogenesis (Houtmeyers *et al.*, 2013). Nonetheless, much remains unknown about the molecular mechanism of *Zic* gene function and regulation. Moreover, the long-standing uncertainty regarding whether ZIC proteins act as bona fide transcription factors or co-factors is only now beginning to be clarified with evidence that they can act as either class of molecule, presumably in a context dependent manner (Ali *et al.*, 2012).

1.2.1 Genomic Arrangement

Sequencing, assembly and annotation of multiple genomes has revealed the *ZIC* genes have an unusual arrangement within the genomes of all vertebrates examined (Houtmeyers *et al.*, 2013). In mammals, the five *Zic* genes reside at three genomic loci. *Zic1* and *Zic4* exist as a divergently transcribed tandem gene pair, as do *Zic2* and *Zic5*, whilst *Zic3* is an X-linked singleton in each species (Figure 1.3) (Ali *et al.*, 2012). Examination of exon-intron boundaries and phylogenetic analysis has revealed that *Zic* genes can be divided into two subgroups (Subgroup A and Subgroup B), based on similarities in nucleotide sequence and protein structure (Aruga *et al.*, 2006). A hypothetical model of *Zic* gene evolution by Aruga *et al.* (2006) proposes that the ancestral single copy *Zic* gene contained one intron and a set of conserved structural domains. Tandem duplication of the ancestral gene and sequence modifications led to the generation of two isoforms: one that acquired an additional intron, while the other gene copy underwent sequence divergence accompanied with the loss of certain structural domains (Houtmeyers *et al.*, 2013). Additional genome duplication events produced eight *Zic* genes organized in four bigene clusters, with each cluster containing a *Zic* gene from Subgroup A (*Zic1*, *Zic2* and *Zic3*) and Subgroup B (*Zic4* and *Zic5*). Nonetheless the maximum number of *Zic* genes identified in any species to date is seven in zebrafish (Grinblat *et al.*, 1998, Toyama *et al.*, 2004), with the varying *Zic* gene copy numbers observed across vertebrate species having arisen by a combination of gene loss and de novo duplication events.

The retention of some *Zic* genes as tandem gene pairs has implications for their function and study. For example, it is likely that the gene pairs share regulatory sequences and have overlapping expression patterns. This provides the potential for functional redundancy and/or other modes of paired activities, such as dominant-negative interference between co-expressed family members. Additionally, the production of animal models in which more than one *Zic* gene is mutated is difficult because the common approach of crossing single mutant strains is impractical due to the low chance of a cross over occurring between the individual mutations.

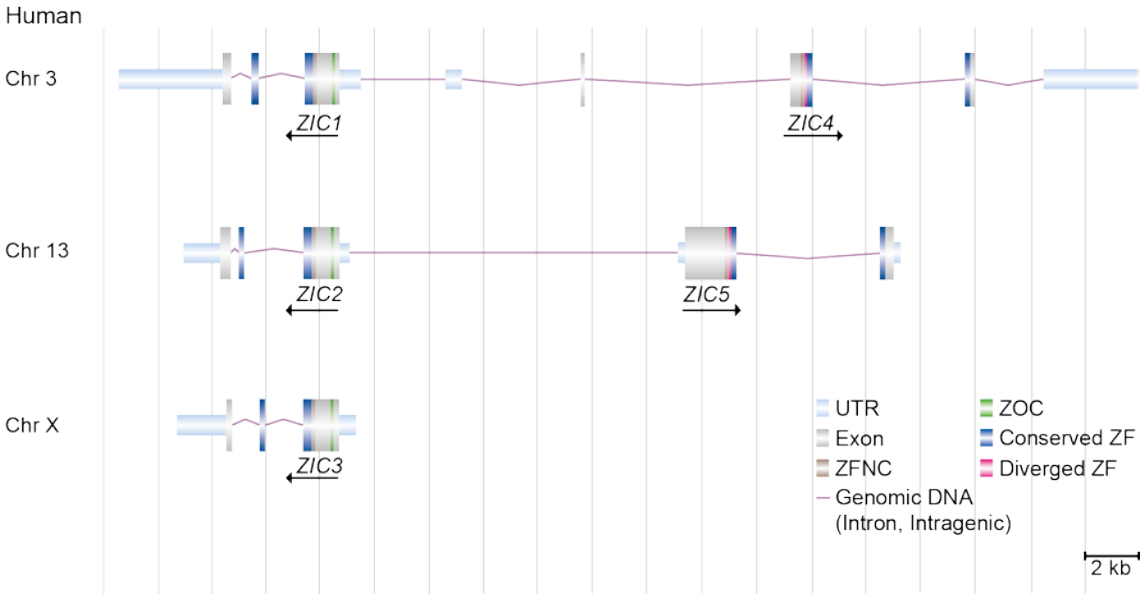


Figure 1.3: Genomic arrangement of vertebrate *ZIC* genes. Illustration of the arrangement of *ZIC* genes on the human genome, as shown in the Ensembl genome browser. *ZIC1* and *ZIC4* are located on chromosome (Chr) 3, *ZIC2* and *ZIC5* are located on chromosome 13, while *ZIC3* is on the X-chromosome. UTR: untranslated region; ZFNC: zinc finger N-terminal conserved region; ZOC: Zic-Opa conserved region; ZF: zinc finger. The figure was adapted from Houtmeyers *et al.* (2013).

1.2.2 Protein Structure and Function

The defining feature of the OPA protein and its vertebrate orthologues (ZIC1-5) is a zinc finger domain (ZFD) that consists of five tandem C_2H_2-type zinc fingers (ZFs) (Figure 1.4). This type of ZF consists of two cysteine and two histidine residues that ligate the zinc (Zn^{2+}) ion, enabling the protein to bind DNA (Iuchi, 2001). The ZFD spans over a third of the gene in the C-terminal half. Sequence comparisons show that the first zinc finger (ZF1) is well conserved amongst ZIC1-3, but is more divergent in ZIC4 and ZIC5 (Aruga *et al.*, 2006). This is part of the basis for the sub-division of the *Zic* genes into structural subclasses with *Zic1-3* belonging to Subgroup A and *Zic4-5* belonging to Subgroup B. The ZIC ZFD possesses not just DNA-binding ability, but also protein binding ability and appears responsible for nuclear localization. Mutations within the ZFD adversely affect the function of these proteins by altering these properties (Mizugishi *et al.*, 2004, Brown *et al.*, 2005, Bedard *et al.*, 2007).

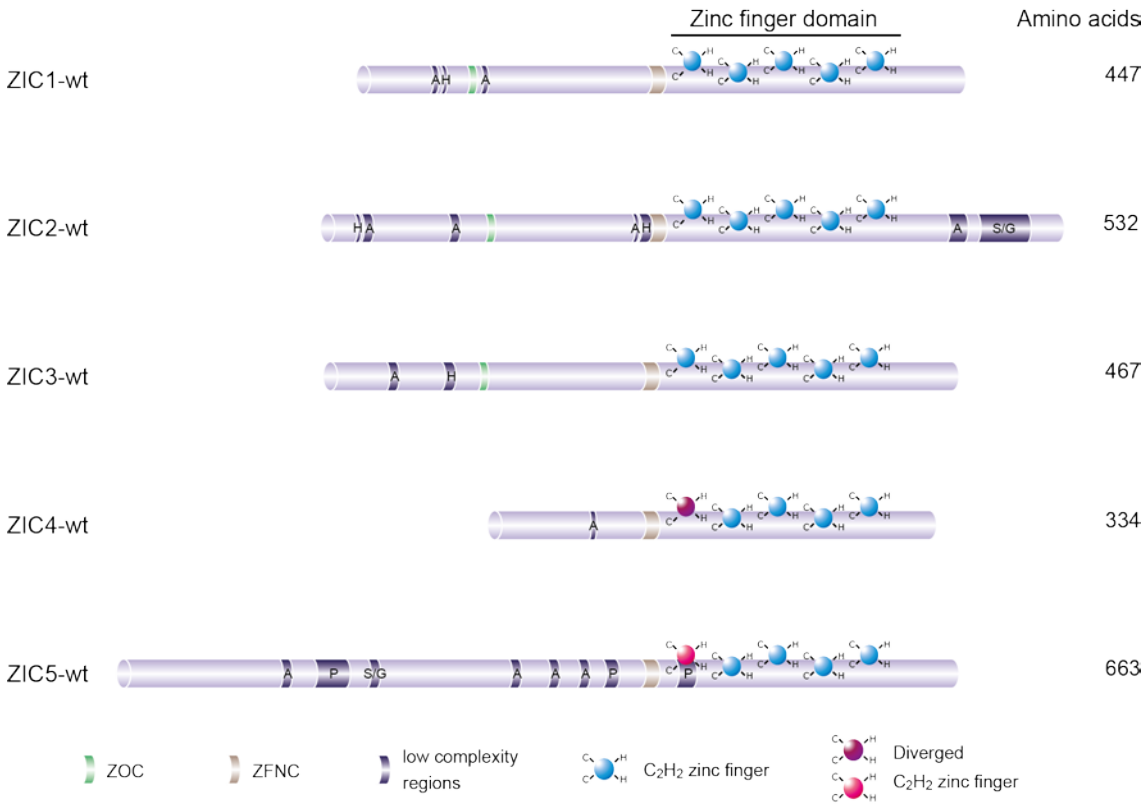


Figure 1.4: Schematic illustration of ZIC proteins. Structural features of the five human ZIC proteins are shown. All contain a C_2H_2-type ZFD that is highly conserved, with only the first zinc fingers of ZIC4 and ZIC5 showing divergence. Each protein also consists of a short, highly conserved ZFNC domain (immediately upstream of the first zinc finger). ZIC1, ZIC2 and ZIC3 also share the ZOC domain. All proteins contain several low complexity regions (A: alanine, H: histidine, P: proline, S/G: serine/glycine). The figure was adapted from Houtmeyers *et al.* (2013).

Biochemical and genetic experiments have shed some light on the ZFs responsible for DNA and protein binding activities. For example, structural analysis of ZIC3 suggests that ZF1 and ZF2 may not be canonical (DNA-binding) C_2H_2-type ZFs. These fingers together form a single structural unit called the tandem $CWCH_2$ motif, which is characterized by the presence of a tryptophan (W) residue between the two canonical cysteines of each ZF (Aruga *et al.*, 2006). This motif is present in the first and second ZFs of a wide range of metazoan species, and may be of biological significance since a missense mutation of the tryptophan in ZF1 of ZIC3 is associated with congenital heart malformation (Chhin *et al.*, 2007). The molecular role of the $CWCH_2$ motif remains unknown, however it is hypothesized that it modulates the DNA-binding capability of the other canonical ZFs and/or participate in protein-protein interactions (Houtmeyers *et al.*,

2013). Furthermore, mutational analysis of *Zic2* has revealed that mutation of ZF4 ablates DNA-binding (Brown *et al.*, 2005) and causes in vivo loss-of-function (Elms *et al.*, 2003b) without necessarily affecting protein-protein interactions (Pourebrahim *et al.*, 2011). Further work is required, however, to clearly delineate individual and co-operative ZF functions.

The N- and C-terminal sequences that flank the ZFs vary significantly among all the family members. Nonetheless there is one short region (14-21 amino acids in length) at the N-terminus of the ZFD (ZFNC) found in all family members across a wide range of species (Aruga *et al.*, 2006) whose function remains unknown. Additionally, a small (9-10 amino acids in length) N-terminally located domain is conserved amongst vertebrate ZIC proteins that belong to Subgroup A. OPA, the *Drosophila* orthologue, also contains this domain hence it was named the ZIC/Odd-paired conserved (ZOC) motif (Aruga *et al.*, 1996, Layden *et al.*, 2010). This domain is required for protein-protein interactions (Mizugishi *et al.*, 2004, Himeda *et al.*, 2013). Other structural features of the ZIC family include low complexity regions (poly-alanine, -histidine, -proline and -serine/glycine tracts) located outside the ZFD. The functional relevance of these regions is still unclear, however expansion mutations of the C-terminal alanine tract of ZIC2 (Brown *et al.*, 2005) and the N-terminal alanine tract of ZIC3 (Wessels *et al.*, 2010) are associated with human pathology.

Immunohistochemical staining has revealed that the protein products of *Zic* genes predominantly localize in cell nuclei (Aruga *et al.*, 1994, Yokota *et al.*, 1996), consistent with a role in gene expression, DNA replication and repair, or regulation of chromatin and other nuclear structures (Aruga, 2004). The ZIC proteins are thought to behave as transcription factors, since the orthologous *opa* is a ZF containing pair-rule gene and all other pair-rule genes are transcription factors. Nonetheless recent evidence shows ZIC proteins can influence gene expression also as a cofactor or by complexing with transcription machinery or chromatin modifying complexes, thus much remains unknown about ZIC protein function.

1.2.2.1 The ZIC proteins function as transcription factors in vitro and in vivo

Perhaps the best evidence that mammalian ZIC proteins act as transcription factors comes from the study of point mutations in the human and mouse *Zic2* genes. A pathogenic human ZIC2 variant (ZIC2-R409P) with a missense mutation in ZF5, displayed diminished affinity to target ZIC DNA-binding sequence and reduced transactivation ability in cell-based assays (Hatayama *et al.*, 2011). The mouse ZIC2 variant (ZIC2-C370S) carries a missense mutation in ZF4 and animals homozygous for this mutation have a phenotype that shifts in the same direction as an allele in which decreased levels of wild-type protein are produced (Elms *et al.*, 2003b). In vitro analysis shows that the ZIC2-C370S protein cannot bind DNA or activate transcription (Brown *et al.*, 2005). These data imply that phenotypes, in either case, are produced due to loss of transcription factor function.

Several studies have attempted to define the consensus DNA binding sequence for ZIC proteins (Table 1.1). The first putative ZIC DNA-binding site was identified via a yeast-one hybrid assay, where human brain cDNAs were used to identify transcription factors that could regulate the expression of the *DOPAMINE RECEPTOR 1A* (D_{1A}) gene (Yang *et al.*, 2000). Electrophoretic mobility shift assay (EMSA) and cell-based reporter assays showed that ZIC2 binds the D_{1A} promoter and represses transcription. Similar experiments identified three ZIC binding sites on the *APOE* promoter and cell-based reporter assays showed ZIC proteins could activate transcription via this promoter (Salero *et al.*, 2001). Another study used cDNA selection and mutational analysis to identify the minimal essential binding site for ZIC1-3 (Mizugishi *et al.*, 2001). However none of these sequences have been validated in vivo. Two recent studies (Badis *et al.*, 2009, Lim *et al.*, 2010) have identified similar consensus ZIC3 DNA-binding sequences (Figure 4.3). Badis *et al.* purified the ZIC3 ZFD and used a universal protein binding microarray (PBM) to identify an optimal ZIC3 binding site. Lim *et al.* performed ChIP-chip on mouse embryonic stem cells (ESCs) and found three different sequences to be over-represented in the presence of ZIC3 (Table 1.1). Additionally Lim *et al.* found ZIC3 bound a specific region in the promoter of the gene *Nanog*, which is a pluripotency regulator. Hence it can be inferred that

ZIC3 is involved in regulating the pluripotency of ES cells. More molecular evidence is required, however, to conclusively determine the regulatory targets of ZIC proteins. Despite the lack of confirmation on the in vivo validity of the identified motifs, these binding sequences have been used to construct in vitro ZIC transactivation assays, which serves as an important tool in assessing the functionality of mutant ZIC proteins and identifying the structural domains necessary for normal protein function (Brown *et al.*, 2005, Ahmed *et al.*, 2013).

Binding region	Sequence (5`-3`)	Assay	Reference
D_{1A} promoter	GACCGCCCCCAGGGCAGGGG	EMSA and cell-based reporter assays	Yang *et al.*, (2000)
APOE promoter (-185/-174)	CTGTGGGGGGTG	EMSA and cell-based reporter assays	Salero *et al.*, (2001)
APOE promoter (-136/-124)	TGCCCTGCTGTGC		
APOE promoter (-65/-53)	CCCTGGGGGAGGG		
Optimised ZIC binding sequence	GGGTGGTC	EMSA (SELEX)	Mizugishi *et al.*, (2001)
Optimised ZIC binding sequence	CACAGCAGG	PBM	Badis *et al.*, (2009)
Optimised ZIC binding sequence	CCCGCTGG	ChIP-chip using mouse ES cells	Lim *et al.*, (2010)
Optimised ZIC binding sequence	CCCAGCGGGG		
Optimised ZIC binding sequence	CCCGCTGTGC		

Table 1.1: ZIC DNA-binding sequences. All studies identified GC-rich sequences as potential ZIC binding sites.

1.2.2.2 The ZIC proteins function as cofactors in vitro

The first evidence that ZIC proteins may act as co-factors came from a candidate approach, based on the fact that the ZIC ZFD is highly homologous to that of the GLI proteins (Aruga, 2004). GLI proteins are vertebrate orthologues of the *Drosophila cubitus interruptus* gene, which is the transcriptional mediator of the developmentally important Hedgehog signaling pathway (Koebernick *et al.*, 2002). The ZIC proteins bind GLI DNA binding sites (GLI-BS), albeit with much lower affinities and do not stimulate transcription from GLI-BS (Mizugishi *et al.*, 2001). Furthermore ZIC and GLI proteins form heterodimers, with both proteins interacting through their ZF3-5 (Koyabu *et al.*, 2001), while no heterodimer formation has been observed between the ZIC family members (Brown *et al.*, 2005). This interaction promotes translocation of GLI proteins, predominantly localized in the cytoplasm, to the nucleus. Consequently this stimulates transcription at GLI-BS above that observed in the presence of only GLI proteins (Mizugishi *et al.*, 2001). Although GLI-ZIC heterodimers have never been purified from an in vivo situation, *Zic3* loss-of-function rescues the digit phenotype in the *Gli3* mouse mutant known as extra toes, implying some functional relationship between these two molecules (Quinn *et al.*, 2012).

More recently ZIC proteins have been implicated in the transcriptional response to canonical WNT signaling (Pourebrahim *et al.*, 2011, Fujimi *et al.*, 2012). The presence of WNT ligand alters transcription of a variety of genes involved in virtually every aspect of embryonic development (Clevers, 2006). Upon binding of WNT ligand to its receptor, β-catenin (a cytoplasmic protein) translocates to the nucleus where it binds its co-activator TCF to activate expression of target genes at TCF DNA-binding sites (Daniels *et al.*, 2005). In the absence of WNT ligand, TCF proteins form a repressor complex at TCF DNA binding sites. In vitro and in vivo reporter assays have demonstrated that ZIC proteins can repress β-catenin mediated transcription (Pourebrahim *et al.*, 2011, Fujimi *et al.*, 2012). Immunoprecipitation revealed that ZIC proteins bind TCF via their ZFD to inhibit TCF-dependent transactivation. To date, there is no genetic evidence of this interaction during murine development.

1.2.3 Role during Embryogenesis

The *Zic* genes have dynamic expression patterns during embryogenesis and in adults their expression has mainly been documented in the central nervous system (Grinberg *et al.*, 2005). In the developing embryo all *Zic* genes are differentially expressed in midline regions, such as the dorsal neural tube (forms the brain and spinal cord) and dorsal part of somites (forms muscles and vertebrae) (Nagai *et al.*, 1997). In addition, distinct combinations of *Zic* genes are expressed in the limb buds, tail buds and the developing eyes.

Zic genes are often expressed in overlapping domains providing the possibility of functional redundancy. Nonetheless, when mutated individually a range of unique phenotypes are found illustrating at least some exclusive functions during mammalian development. For example, loss of *Zic2* function causes holoprosencephaly, heart defects, hindbrain patterning defects, and neural tube defects (Elms *et al.*, 2003b, Warr *et al.*, 2008, Barratt *et al.*, 2014). Mutations in *ZIC3* lead to left-right (L-R) axis malformations and neural tube defects (Gebbia *et al.*, 1997, Purandare *et al.*, 2002, Ware *et al.*, 2004) and lack of *Zic5* causes craniofacial abnormalities and neural tube defects (Furushima *et al.*, 2005). Deletion of *Zic1* in mice leads to cerebellar malformations and axial skeleton abnormalities (Aruga *et al.*, 1998). Heterozygous deletion of *ZIC1* and *ZIC4* leads to a defect in cerebellar development, known as Dandy-Walker malformation (Grinberg *et al.*, 2004). In the case of *Zic* genes, mutations in *Zics1/4*, *Zic2* and *Zic3* in the mouse give rise to the same defects, as associated with human mutations in these genes. Due to practical and ethical constraints the study of developmental processes in humans is hindered. Therefore the mouse is used as a model organism for studying human genetic disorders and mammalian developmental genetics (Malakoff, 2000).

1.3 ZIC3

1.3.1 Gene Location and Inheritance

Zic3 is the only mammalian *Zic* gene family member located on the X-chromosome (Zhu *et al.*, 2007). In zebrafish there is evidence of a *Zic6* remnant, however, in mammals *Zic3* is unpaired (Ali *et al.*, 2012). X-linkage is associated with a specific pattern of inheritance in man and mouse, since in males all X-linked genes have no allelic counterpart; a situation described as hemizygosity. Different progeny classes are generated depending upon whether the mutation is passaged through the mother or father (Figure 1.5). Both hemizygous males and heterozygous females inherit one copy of the mutant allele, however, male carriers generally display phenotypes that are more severe than the females due to a combination of factors: (i) all male cells have only one copy of the *Zic3* gene and hemizygous mutations are functionally equivalent to a homozygous autosomal mutation, and (ii) all female cells have only one active *Zic3* gene and due to the random nature of X-inactivation, females are mosaics of cells that are wild-type or mutant.

A

Maternal Inheritance ($\mathbf{X^mX}$ x $\mathbf{XY}$)

$\mathbf{X^mX}$ Heterozygous female	$\mathbf{X^mY}$ Hemizygous male
$\mathbf{XX}$ Wild-type female	$\mathbf{XY}$ Wild-type male

B

Paternal Inheritance ($\mathbf{XX}$ x $\mathbf{X^mY}$)

$\mathbf{XX^m}$ Heterozygous female	$\mathbf{XY}$ Wild-type male
$\mathbf{XX^m}$ Heterozygous female	$\mathbf{XY}$ Wild-type male

Figure 1.5: X-linked Inheritance. (A) Passage through a female carrier (X^mX) produces a heterozygous female, a hemizygous male and a normal male and female. **(B)** Passage through a male carrier (X^mY) produces females that are obligate heterozygotes and males that cannot inherit the mutation.

1.3.2 *Zic3* expression during murine gastrulation

Expression of *Zic3* mRNA can be detected via whole mount in situ hybridisation (WMISH) in the extraembryonic ectoderm of the mouse embryo as early as 5.0 dpc, when the embryo is still a symmetrical cylinder of cells (Elms *et al.*, 2004). This expression is maintained until the onset of gastrulation. As gastrulation starts (6.5 dpc), expression of *Zic3* shifts distally from the extraembryonic ectoderm to the epiblast of the embryo. When the primitive streak emerges (6.75 dpc), *Zic3* expression is confined to the primitive streak and the anterior ectoderm (prospective head region). As the node forms at the anterior of the primitive streak (7.0 dpc), *Zic3* expression is seen in the primitive streak and surrounding ectoderm but not in the node (Elms *et al.*, 2004). As gastrulation proceeds (7.5 dpc), this expression remains relatively constant with *Zic3* transcripts also found in mesodermal wings of the embryonic region, the primitive streak and the adjacent ectoderm. When headfolds start to develop (7.75 dpc) *Zic3* expression departs from the primitive streak and begins in the node. This expression becomes more prominent as development proceeds and symmetric expression of *Zic3* at the node can be viewed until 8.5 dpc (Purandare *et al.*, 2002). At the completion of gastrulation, *Zic3* transcripts are present at high levels in:

1. the mesoderm (gives rise to cardiac muscles, skeletal muscles, somites and vertebrae)
2. parts of the ectoderm (gives rise to the central nervous system, eyes and skin).

Zic2 and *Zic5* are also expressed prior to the onset of gastrulation and are later found in the mesodermal and ectodermal germ layers (Furushima *et al.*, 2000, Elms *et al.*, 2004). There is also expression overlap during organogenesis where all *Zic* genes are expressed in the dorsal neural tube. *Zic1-3* are co-expressed in the dorsomedial somites and the eye, whereas *Zics2*, *Zic3* and *Zic5* are co-expressed in the limb buds and developing brain, while *Zic2* and *Zic3* are co-expressed in the tail bud (Furushima *et al.*, 2005). Despite the overlap in *Zic3* expression with that of *Zic2* and *Zic5*, analysis of human and mouse mutants shows *Zic3* has a unique role in establishment of the L-R axis.

1.3.3 Regulation of Zic3 Function

1.3.3.1 Via nucleocytoplasmic shuttling

As previously described, nuclear entry of transcription factors is facilitated by the presence of NLSs, which are recognized by and interact with nuclear import receptors on the nuclear membrane (Jans *et al.*, 2000). NLSs usually consist of multiple monopartite or bipartite clusters of positively charged basic amino acids, such as arginine (R), lysine (K) and histidine (H) (Hicks *et al.*, 1995). No such canonical NLS is found in any ZIC protein sequence; instead two recent studies suggest that the ZIC3 ZFD contains an interspersed type NLS, spanning a wide region containing other amino acids (Bedard *et al.*, 2007, Hatayama *et al.*, 2008). For example, Bedard *et al.* (2007) reported that 90% of wild-type ZIC3 (in NIH3T3 cells) localized exclusively in the nucleus, whereas when putative NLS mutations were made in the ZFD, nuclear localization of the mutant proteins was limited to 20-30%. Interspersed NLSs are recognized by the nuclear import machinery (importin-α/β complex) near the NPC and the protein allowed entry to the nucleus via the NPC (Hicks *et al.*, 1995). In addition, the in vivo significance of failed nuclear localization is implied by human genetics. Mutations in the NLS have been identified in human patients with ZIC3-related Heterotaxy, implying the inability of ZIC3 to reach the nucleus causes L-R axis malformation (Ware *et al.*, 2004).

1.3.3.2 Via dominant-negative interference

Two experiments raise the possibility that a second post-translational mechanism of ZIC3 regulation is dominant-negative interference. Firstly, Kitaguchi *et al.* (2000) injected truncated *Xenopus zic3* mRNAs (coding for amino acids 1-214 only) in *Xenopus* blastomeres and allowed development to proceed until L-R axis formation. They observed that overexpression of the mutant construct attenuated the function of the wild-type protein and caused developmental defects (Kitaguchi *et al.*, 2000). Presumably, a naturally occurring short splice variant of *Zic3* (as found for human ZIC3) (Bedard *et al.*, 2011) could modulate ZIC3 function. Mammalian ZIC3 mutant proteins, however, cannot act in a dominant-negative manner against wild-type ZIC3 in

vivo, due to the X-linked location of the gene (as previously described). It can potentially act as a dominant-negative against co-expressed family members such as ZIC2 and ZIC5. This study, however, failed to show any interaction of the mutant ZIC3 with its wild-type counterpart, or binding of the mutant protein to DNA or essential co-factors. Furthermore Brown *et al.* (2005) investigated the possibility that mutant ZIC2 can interact and interfere with its wild-type counterpart. Co-immunoprecipitation and yeast two-hybrid analysis showed that the ZIC proteins do not form dimers (Brown *et al.*, 2005). Nevertheless this study demonstrated, via reporter gene transactivation assays, that some ZIC2 mutants were capable of antagonizing the activity of wild-type ZIC2, in a dose-dependent manner. The mutant proteins evidently retain their DNA-binding or protein binding capabilities. Mutational analysis revealed that a functional ZFD was required for both functions.

1.4 HETEROTAXY

Despite displaying outward symmetry, most organisms have several internal asymmetries, for example humans have both unpaired organs (like the heart or liver) or paired organs in which each one of the pairs have asymmetry (like the lungs). These asymmetries derive from an initial symmetry breaking event that propagates to one side of the embryo, distinguishing left from right (and establishes the embryonic L-R axis). A failure to establish these differences leads to laterality disorders, known as Heterotaxy (Figure 1.6), which is a greek term meaning "other arrangement" (Ware *et al.*, 2006a). Heterotaxies are therefore characterised by abnormal positioning of the thoracic and/or visceral organs and can cause multiple congenital malformations (Sutherland *et al.*, 2009). The major cause of morbidity and mortality are complex cardiovascular malformations that are sometimes accompanied by defects in the pulmonary, gastrointestinal, genitourinary, immune and muscoskeletal systems (Kearns-Jonker, 2006). Epidemiological surveys show Heterotaxy underlies 3% of congenital heart defects and has an estimated prevalence of 1:10,000 live births (Lin *et al.*, 2000). Although the true prevalence may be underestimated due to the wide spectrum of possible phenotypic abnormalities, variable expressivity and incomplete penetrance associated with laterality defects.

Heterotaxy has a complex aetiology, most commonly it is sporadic and both genetic and environmental factors contribute to this condition. Associated environmental factors include maternal diabetes, maternal cocaine use and monozygotic twinning (Kuehl *et al.*, 2002). Mutational analysis in affected children and families has identified associated heterozygous defects in a variety of autosomal genes (*NODAL, CFC-1, ACVRIIB, FOXH1, LEFTYA*) as well as the X-linked *ZIC3* gene (Carmi *et al.*, 1992, Gebbia *et al.*, 1997). Defining the precise role these genes play in L-R axis formation via animal models has advanced the understanding of this disease immensely, although there are some complications due to species specific differences in gene function (Kearns-Jonker, 2006). Additionally, in animal models, more than 100 genes have been shown to be required for the formation of the L-R axis and it is likely that many of these play some role in human Heterotaxy (Sutherland *et al.*, 2009).

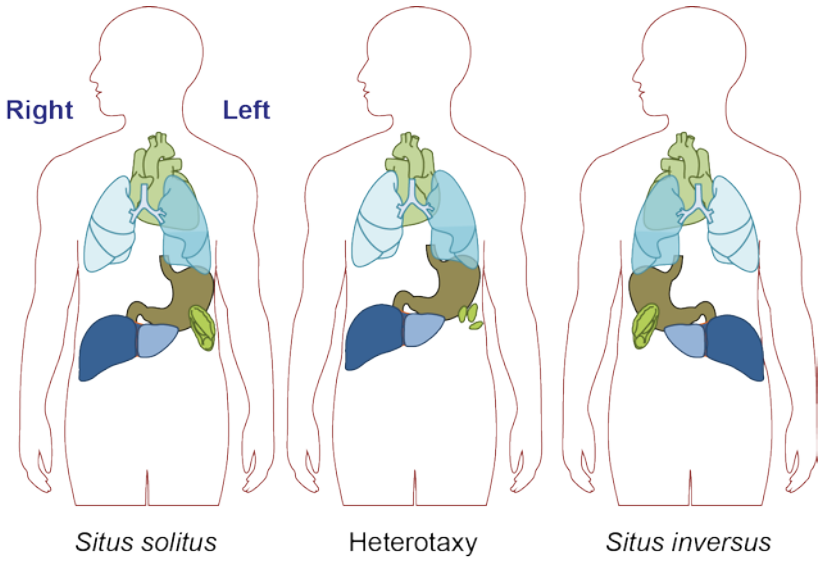


Figure 1.6: *Situs* arrangements. *Situs solitus* shows the normal arrangement of visceral organs, with the heart, stomach and spleen on the left side, and the liver on the right. Heterotaxy shows the heart and visceral organs oriented independently of each other. *Situs inversus* represents a mirror image reversal of all organ asymmetry. This figure was obtained from R. Arkell.

ZIC3 was the first gene implicated in human Heterotaxy (Gebbia *et al.*, 1997) and remains the gene most frequently associated with this condition. Mutations within *ZIC3* have been found to contribute to 5% of all familial cases of Heterotaxy and 1% to sporadic cases of Heterotaxy (Ware *et al.*, 2004). To date twenty-three different mutations in *ZIC3* have been reported: twelve missense, ten nonsense, one frameshift (Gebbia *et al.*, 1997, Megarbane *et al.*, 2000, Ware *et al.*, 2004, D'Alessandro *et al.*, 2011, Cowan *et al.*, 2014). Males are hemizygous for the mutation and manifest *situs ambiguus* (Heterotaxy) with varying degrees of severity. However some male carriers (from a family with a history of *ZIC3*-related congenital defects) have been reported to be phenotypically normal (Megarbane *et al.*, 2000, D'Alessandro *et al.*, 2011), indicating that other proteins might be able to compensate for ZIC3 function. Females that are heterozygous for the mutation are usually asymptomatic. Nevertheless in one family some females carrying the mutation have *situs inversus* (Gebbia *et al.*, 1997). The different phenotypes in females can be due to skewed X-inactivation where one X-chromosome carrying normal *ZIC3* is inactivated while the other X-chromosome carrying the mutated allele remains active (Chhin *et al.*, 2007). This pattern of inheritance exhibited by *ZIC3* and the fluctuating phenotypes observed makes

diagnosis difficult. Specifically it is difficult to pinpoint the particular tissues and the exact stage of development where ZIC3 function is required using human clinical data.

1.4.1 Mouse models of *Zic3* dysfunction

Mice present an excellent system for analysing the function of genes required for mammalian development, providing function of the gene of interest is well conserved between man and mouse. There is ample evidence that this is the case for *ZIC3*/*Zic3*, since multiple strains of mice are available with *Zic3* loss-of-function mutations (Table 1.2) and in each case mutant embryos display similar anatomic abnormalities as observed in human patients of Heterotaxy (Purandare *et al.*, 2002).

MGI Allele ID	Allele Symbol	Allele Name	Synonyms	Allele Type	Allele Attributes
MGI: 1856679	*Zic3^{Bn}*	bent tail	*Bn*	spontaneous	large deletion
MGI: 3043028	*Zic3^{Ka}*	katun	*Ka*	spontaneous	nonsense
MGI: 5476827	*Zic3$^{tm1.1Smwa}$*	targeted mutation 1.1, Stephanie M. Ware	*Zic3^{flox}*	targeted	conditional ready, no functional change
MGI: 2180720	*Zic3^{tm1Bca}*	targeted mutation 1, Brett Casey	*Zic3^{-}*, *Zic3^{null}*	targeted	null/knockout, reporter
MGI: 3698161	*Zic3^{tm1Jwb}*	targeted mutation 1, John W. Belmont	*Zic3^{neo}*, *Zic3^{null}*	targeted	null/knockout
MGI: 5470150	*Zic3$^{tm2.1Jwb}$*	targeted mutation 2.1, John W. Belmont	*Zic3^{flox}*	targeted	conditional ready, no functional change

Table 1.2: Murine alleles for *Zic3* (downloaded from MGI: http://www.informatics.jax.org).

Much is known about the molecular circuitry that establishes the murine L-R embryonic axis (Figure 1.7) and analysis of embryos null for *Zic3* shows they have aberrant expression of *Nodal* and *Pitx2*, at 8.5 dpc. It was initially proposed that *Zic3* may control *Nodal* expression during L-R patterning (Ware *et al.*, 2006b). Recent studies, however, using conditional inactivation of *Zic3* indicate it acts earlier in development than initially proposed (Jiang *et al.*, 2013, Sutherland *et al.*, 2013). ZIC3 function in the primitive streak of the gastrula is needed to prevent Heterotaxy.

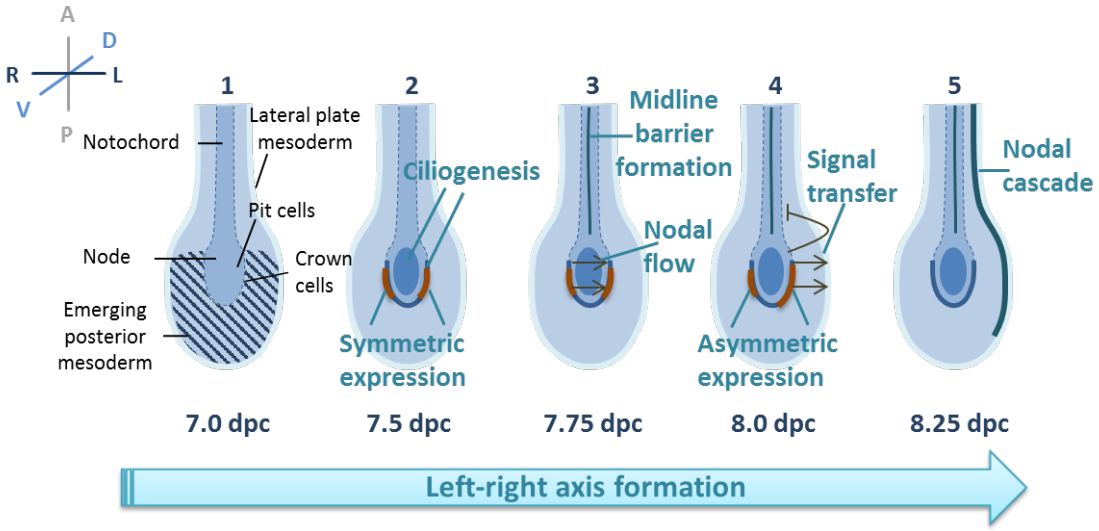


Figure 1.7: Schematic representation of L-R axis formation in the mouse. 1: the node is induced to form at mid-gastrulation (7.0 dpc) and in the next 24 hours develops into a shallow, crescent shaped depression on the ventral side of the embryo. The cells within the depression are termed pit cells, whereas those that form the raised surface that rings the pit are called crown cells (Norris, 2012). **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 first a disorganised, then laminar, leftward flow of extracellular fluid within the node. It is posited that this leftward nodal-flow is sensed by crown cell cilia, prompting a Ca^{2+} flux in the left 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 asymmetrically expressed in the node crown). WNT 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 obtained from R. Arkell.

Zic3 is thought to be instrumental in other developmental process as it is required to maintain pluripotency in ESCs (Lim *et al.*, 2007) and is expressed in an array of developing embryonic tissues, such as those that form the central nervous system, muscles and aspects of the visual system (Herman *et al.*, 2002). Precisely how *Zic3* acts to prevent Heterotaxy and its relevance in other developmental processes is yet to be discovered.

1.5 SCOPE OF THESIS

Given the criticality of ZIC proteins for embryogenesis, it is important to understand how ZIC proteins function at the molecular level. This includes not only identifying target DNA-binding sites or protein partners, but also characterisation of structural domains required for particular functions. This thesis aims to:

1 determine the importance of the ZFD and C-terminus for ZIC3 function, using a novel murine allele

2 design a new ZIC-specific transactivation assay

3 identify regions within the ZIC3 protein involved in transactivation

CHAPTER 2

MATERIALS AND METHODS

2.1 GENOTYPING OF KATUN MICE

2.1.1 Genomic DNA Extraction

The katun mouse colony was maintained at the Australian Phenomics Facility at The Australian National University according to Australian Standards for Animal Care under protocols approved by The Australian National University Animal Ethics and Experimentation Committee.

Mice were genotyped (sections 2.1.2 and 2.1.4) by PCR screening of genomic DNA extracted from ear biopsy tissue. Ear notches were collected in 50 µL of lysis solution [50 mM Tris(hydroxymethyl)aminomethane (Tris)-HCl (pH 8.5), 1 mM Ethylenediaminetetraacetic acid (EDTA), and 5% Tween 20] containing 2 µL of Proteinase K (10 µg/µL; Sigma Aldrich, Cat. No. P2308) and incubated at 55°C for 60 mins, followed by 95°C for 10 mins to inactivate Proteinase K. Tissue debris was pelleted by centrifuging at 2,000 g for 5 mins and each sample was diluted 1:10 in Analar H_2O. Embryos were genotyped using a fragment of extra embryonic tissue/ectoplacental cone (7.5 dpc) or yolk sac (8.5 dpc and 9.5 dpc). Tissue was collected in 20 µL (7.5 dpc), 25 µL (8.5 dpc) or 35 µL (9.5 dpc) of lysis buffer (50 mM Tris-HCl, pH 8.5, 1 mM EDTA, 0.5% Tween 20) containing 1 µL of Proteinase K (10 µg/µL), followed by incubation at 55°C for 25 mins and 95°C for 5 mins to inactivate Proteinase K.

2.1.2 Genotyping PCR and primers

PCR reactions for optimising genotyping assay involved two buffers: ThermoPrime ReddyMix PCR Mastermix (Thermo Scientific; Cat. No. AB-0575/DC/LD/B) and ImmoMix™ (Bioline; Cat. No. BIO-25020). All reactions were carried out in a 15 µL volume in the presence of 0.67 µM of each oligonucleotide. The ThermoPrime buffer led to final reaction conditions of 75 mM Tris-HCL (pH 8.8 at 25°C), 20 mM Ammonium Sulphate $(NH_4)_2SO_4$, 0.01% (v/v) Tween 20, 1.5 mM Magnesium Chloride ($MgCl_2$) and also contained an inert gel loading dye. ImmoMix™ led to final reaction conditions of 67 mM Tris-HCL (pH 8.3 at 25°C), 16 mM $(NH_4)_2SO_4$, 0.01% (v/v) Tween 20 and 1.5 mM $MgCl_2$. PCRs to be analyzed via gel electrophoresis were performed in 96-Well Clear, Flat

Top PCR plates (Axygen; Cat. No. PCR-96-FLT-C) and the plate was sealed with an Easy Pierce Heat Sealing Film (Axygen; Cat. No. MF-111). All PCRs were performed in an Eppendorf Mastercycler® using two Touchdown PCR programs: TD60 and TD65 (Table 2.1). Information on primers used for all reactions can be found in Table 2.2.

Program	TD60		TD65	
Initial denaturation	94°C; 4 mins		94°C; 4 mins	
Denature	94°C; 30 secs	x19 cycles	94°C; 30 secs	x19 cycles
Anneal	60°C; 30 secs (decreasing 0.5°C per cycle)		65°C; 30 secs (decreasing 0.5°C per cycle)	
Extend	72°C; 30 secs		72°C; 30 secs	
Denature	94°C; 30 secs	x19 cycles	94°C; 30 secs	x19 cycles
Anneal	50°C; 30 secs		55°C; 30 secs	
Extend	72°C; 30 secs		72°C; 30 secs	
Polish	72°C; 7 mins		72°C; 7 mins	

Table 2.1: Thermal profiles of PCR programs used for optimization of genotyping assay.

Primer	Sequence (5`- 3`)
Ark209_F	CTA CTT GCT CTT TCC TGG
Ark241_F	GCG CCT TCT TCC GTT ACA TG
Ark242_R	AGC CTC CTC GAT CCA CTT ACA G
Ark999_R	TGC TCC ATG GTA ACA TGC
Ark1002_F	GAG GTC ATG AAG GTC AG
Ark1003_R	GGG CAT AAA CTT TCC AG
Ark1085_F	CCT TCT TCC GTT ACA TGC G
Ark1086_R	CTG AGC CTC CTC GAT CC

Table 2.2: Primers used for genotyping assays. F: forward primer; R: reverse primer.

2.1.3 Agarose Gel Electrophoresis

After PCR amplification or digestion of DNA via restriction enzymes, the size of the resulting bands was visualized using agarose gel electrophoresis. Size of the fragment determined percentage of agarose in solution. Agarose gels were made by dissolving UltraPure™ Agarose (Life Technologies; Cat. No. 16500-500) in TBE buffer (0.1 M Tris-HCl, 0.09 M Boric acid and 0.001 M EDTA; Amresco). The mixture was heated in a microwave to dissolve the agarose, followed by addition of RedSafe™ Nucleic Acid Staining Solution (1:20,000 dilution; Intron Biotechnology, Cat. No. 21141). Dissolved agarose solution was poured into a cassette and allowed to set for 60 mins.

DNA was prepared for electrophoresis by adding 1 µL of 5x loading dye [20% glycerol (Merck), 19.2% 0.5 M Na_2EDTA (Sigma-Aldrich) and 0.001% bromophenol blue (Sigma-Aldrich)] to every 5 µL of sample. Samples were loaded in separate wells on the agarose gel, with 0.6 µg of 1 Kb Plus DNA Ladder (Invitrogen™; Cat. No. 10787-026) added on either side of the sample group. Gels were electrophoresed at 5–6 Vcm^{-1} for 25-40 mins in 1x TBE and viewed under UV light using a Gel Doc XR System (Bio-Rad).

2.1.4 Genotyping assays

Genotyping of animals and embryos from the katun mouse colony was done using the following techniques:

2.1.4.1 Allelic Discrimination

PCR for allelic discrimination was done using TaqMan® Universal PCR Master Mix (Life Technologies; Cat. No. 4304437). All reactions were carried out in a final volume of 10 µL with ~30 ng of digested ear notch DNA in the presence of 0.9 µM of primers Ark241-Ark242 (Table 2.2). Additionally 0.25 µM of allele-specific probes (Applied Biosystems®) were used to identify the wild-type allele (VIC-CAT CAA GCA GGA GCT G-MGBNFQ) and katun allele (FAM-CAT CAA GCA GTA GCT G-MGBNFQ) (Figure 2.1). Reactions were set up in 96-well Half-Skirted PCR Microplates (Axygen®; Cat. No. PCR-96-LP-AB-C) covered with Axygen Microplate Sealing Film (Fisher Scientific; Cat. No. UC500) and performed using the StepOnePlus™ Real-Time PCR System (Applied Biosystems®). The StepOne Software (version 2.2.2; Applied Biosystems®) was used to run the assay using the following conditions: an initial pre-PCR read at 60°C for 30 secs to record background fluorescence, followed by 95°C for 10 mins to denature the template and a cycling stage of 95°C for 15 secs and of 60°C for 1 min for 50 cycles. A post-PCR read was performed at 60°C for 30 secs to collect data after completion of the PCR. Data was analysed using the same software that records the pre- and post-PCR reads and calculates normalized dye fluorescence (ΔRn) from the wild-type and mutant alleles as a function of cycle number. Based on this data the software called the sample as homozygous for either wild-type or mutant allele, or heterozygous with both alleles.

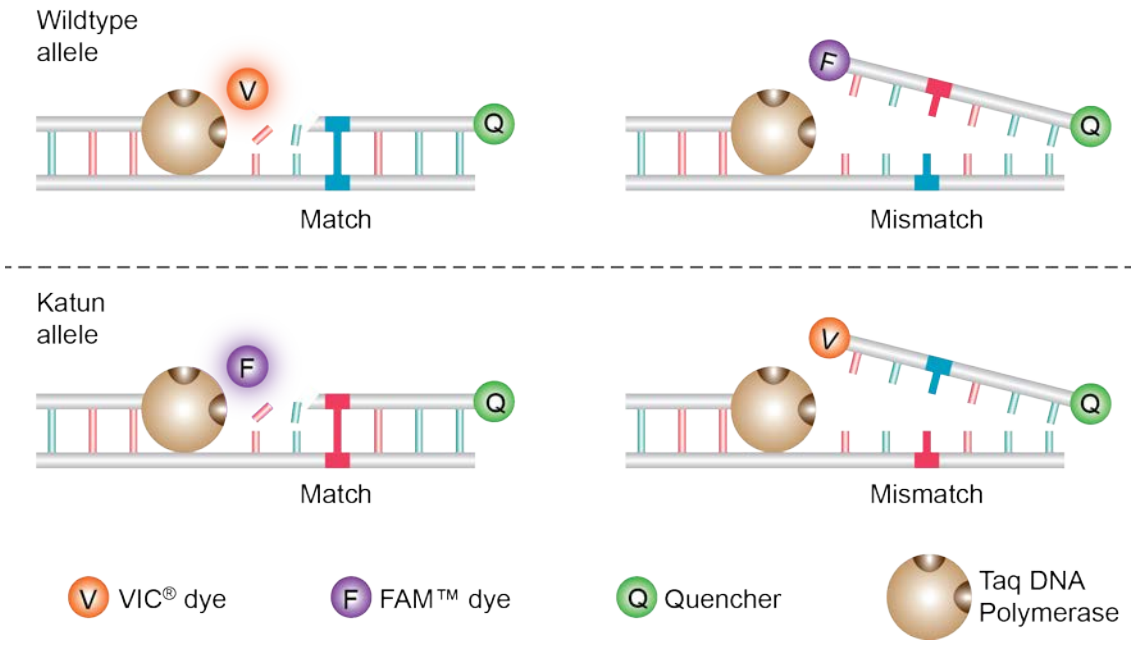


Figure 2.1: Allelic discrimination using fluorogenic probes. Wild-type and katun alleles were distinguished using non-extendable oligonucleotide fluorogenic probes (Livak, 1999). The VIC® (4,7,2'-trichloro-7'-phenyl-6-carboxyfluorescein) probe complemented the wild-type allele, while the FAM™ (6-carboxyfluorescein) probe complemented the katun allele. The probes differed by only one nucleotide (at the site of katun mutation), and at the 5' end were labelled with the reporter (VIC® or FAM™) dye and at the 3' end with the quencher (MGBNFQ: minor groove binding non-fluorescent quencher). The proximity of the reporter dye to the quencher reduced fluorescence via the reporter. During PCR as the forward primer extended the target specific probe was degraded by the 5'-3' exonuclease activity of the Taq DNA polymerase (Arya *et al.*, 2005), which released the reporter dye allowing it to fluoresce. Both probes were included in the same well for each PCR reaction thus fluorescence was only observed when target allele for the probe was present (McGuigan *et al.*, 2002). As PCR cycles increased, the fluorescence intensity increased that was proportional to the accumulation of the PCR product. The fluorescence was measured using a Real-Time PCR machine. This figure was adapted from Applied Biosystems' Allelic Discrimination Getting Started Guide.

2.1.4.2 High Resolution Melt Analysis

PCRs were carried out using ImmoMix™ (Bioline; Cat. No. BIO-25020) and included the LC Green® Plus+ Melting Dye (Idaho Technology Inc.; Cat. No. BCHM-ASY-0005). All reactions were carried out in a final volume of 15 µL with ~30 ng of digested ear notch DNA in the presence of 0.67 µM of primers (Table 2.2). Reactions were set up in Hard-Shell® 96-well PCR Plates (BioRad; Cat. No. HSP-9665) covered with Axygen Microplate Sealing Film (Fisher Scientific; Cat. No. UC500). To avoid evaporation during the HRM process each reaction was covered with ~10 µL of mineral oil prior to PCR. On completion of PCR reaction, the 96-well plate containing PCR products was placed directly into a Light Scanner HR 96 (Idaho Technologies Inc.) and samples melted from 60 to 95°C at a rate of 0.1 $secs^{-1}$. The LC Green® dye specifically binds to double-stranded DNA and emits fluorescence that is captured by the Light Scanner HR 96 instrument. As temperature increases double-stranded DNA is converted to single-stranded, which dissociates LC Green from DNA resulting in a decrease in fluorescence (Figure 2.2). Since melting of DNA is dependent on sequence and length, each amplicon has a unique melt profile. The data were analysed with LightScanner software (Idaho Technologies Inc.).

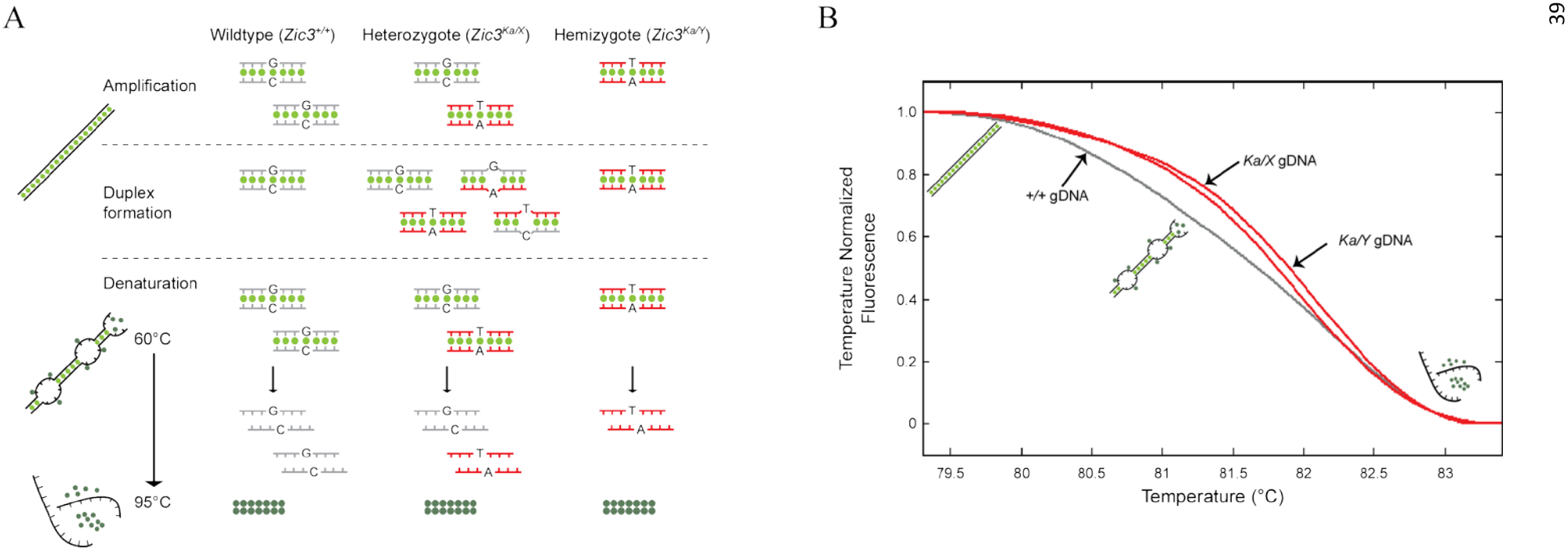


Figure 2.2: Schematic illustration of HRMA. (A) DNA fragments containing the site of mutation were amplified from genomic DNA. Each PCR reaction included a fluorescent DNA-intercalating dye (LC Green®) that specifically bound DNA duplexes and emitted fluorescence (light green circle). Wild-type (*Zic3*$^{+/+}$) and hemizygous (*Zic3*$^{Ka/Y}$) animals produced homoduplexes, while heterozygous (*Zic3*$^{Ka/X}$) animals produced two homoduplexes and two heteroduplexes. For HRMA amplicons were denatured by heating to 95°C to produce single-stranded DNA strands, which caused detachment of the dye and a decrease in fluorescence (dark green circle). **(B)** Shifted melting curves showing melt profiles of amplicons generated from wild-type (*+/+*), heterozygote (*Ka/X*) and hemizygote (*Ka/Y*) genomic DNA (gDNA). The figure was adapted from (Meistertzheim *et al.*, 2012).

2.1.5 Reverse-Transcriptase (RT)-PCR

Embryos to be used for RT-PCR were dissected, staged (by R. Arkell) and individually frozen in 96-well plates (Axygen; Cat. No. PCR-96-FLT-C) on dry ice and stored at -80°C. Upon geno- and sex-typing of the corresponding embryo tissue, the collected embryos were pooled in three genotype classes (*Zic3*$^{+/+}$, *Zic3*$^{Ka/X}$ and *Zic3*$^{Ka/Y}$). At 7.5 dpc each pool consisted of 10 embryos, at 8.5 dpc of 4 embryos and at 9.5 dpc of 2 embryos. RNA was extracted from each sample (by R. Ali) and DNase treated using the RNAqueous®-4PCR kit (Ambion™; Cat. No. AM1914). Quantification of RNA amount was done by Nanodrop spectrophotometry. Approximately 500 ng of RNA template was included in a random primed cDNA first strand synthesis reaction using the SuperScript® VILO™ cDNA synthesis kit (Invitrogen™; Cat. No. 11754-050). A RT negative control synthesis reaction was performed in parallel from each RNA sample. To confirm absence of contaminating genomic DNA in the original RNA samples, amplification from each cDNA sample was performed using primers Ark364 (5`- TCG GAC AAG CCC TAT ATC TG -3`; exon 2 of *Zic3*) and Ark311 (5`- GTT TGC AGA AGC TAT AGC GG -3`; exon 3 of *Zic3*) using ThermoPrime ReddyMix PCR Mastermix (Thermo Scientific; Cat. No. AB-0575/DC/LD/KCL/B) [10 mM Tris-HCl (pH 8.3 at 25°C), 50 mM Potassium Chloride, 1.5 mM $MgCl_2$] and the TD60 PCR thermal cycling conditions (Table 2.1) with the following modification: extension during cycles was carried out at 72°C for 45 secs. PCR products were analysed by agarose gel electrophoresis (section 2.1.3). Amplification from genomic DNA produced a 1006 bp product while that from cDNA resulted in a 162 bp product. For allele specific PCR each cDNA (0.5 µL) was used in the HRMA assay (section 2.1.4.2).

2.2 PLASMIDS

2.2.1 Purchased plasmids

pENTR-3C: pENTR™3C 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 (Figure 2.3A). For cloning of inserts, the vector was linearized with *EcoR*I [New England BioLabs® (NEB)] and dephosphorylated with Antarctic Phosphatase (section 2.3.5). The MCS is flanked by *att*L sites required for Gateway® LR recombination cloning (section 2.3.7.2). A Kanamycin resistance gene is included for selection in *E. coli*.

V5-DEST: pcDNA3.1/nV5-DEST (Invitrogen™; Cat. No. 12290-012) 'Destination' vector used for Gateway® Recombination Cloning. Contains a human cytomegalovirus immediate-early (CMV) promoter for high-level expression of CDSs. An N-terminal V5-epitope tag is present, followed by *att*R sites that surround the *ccd*B gene (Figure 2.3B) required for Gateway® LR recombination cloning (section 2.3.7.2). An Ampicillin resistance gene is included for selection in *E. coli*.

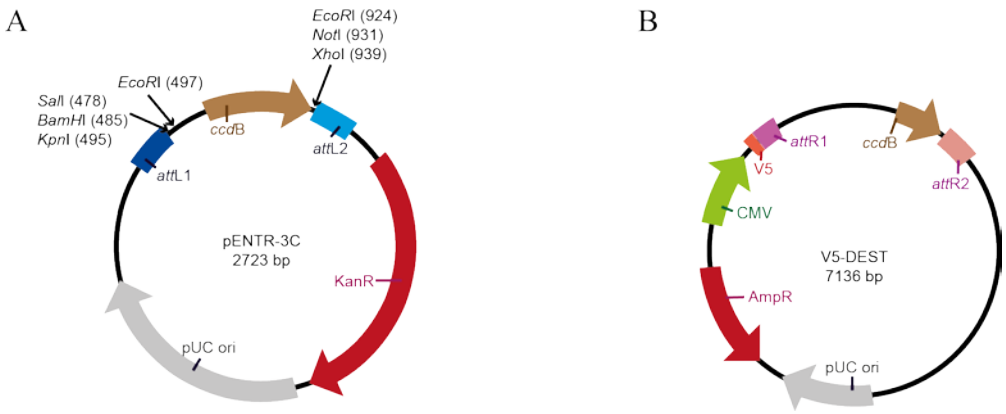


Figure 2.3: Plasmids maps of vectors used for Gateway recombination cloning. (A) Entry vector, pENTR-3C and **(B)** Destination vector, V5-DEST. *att*L1, *att*L2, *att*R1, *att*R2: recombination sites; *ccd*B: negative selection gene; pUC ori: origin of replication; KanR: Kanamycin resistance gene; AmpR: Ampicillin resistance gene; CMV: human cytomegalovirus promoter; V5: epitope tag**.**

EGFP-C1: pEGFP-C1 (Clontech; Cat. No. 6084-1), used to create EGFP-tagged expression clones. Contains the CMV promoter upstream of an Enhanced Green Fluorescent Protein (EGFP; modified GFP for brighter fluorescence and higher expression in mammalian cells) CDS, followed by a MCS. A Kanamycin resistance gene is included for selection in *E. coli*.

B:*luc2*: pGL4.20[*luc2*/Puro] reporter vector (Promega; Cat No. E675A), used for making reporter constructs. Inserts (genomic promoter or synthetic enhancers) were cloned into the vector using restriction enzymes sites in the MCS. An Ampicillin resistance gene is included for selection in *E. coli*.

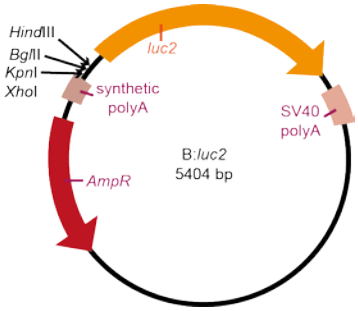


Figure 2.4: Plasmid map for B:*luc2*. Enzymes used for cloning inserts are shown. *luc2*: synthetic firefly (*Photinus pyralis*) luciferase CDS optimized for expression in mammalian cells; SV40 polyA: signals termination of transcription by RNA polymerase II and adds a polyA tail at the 3`-end of the RNA transcript; AmpR: Ampicillin resistance gene; synthetic polyA: transcriptional pause site for reducing the effects of spurious transcription on luciferase expression.

2.2.2 Wild-type expression constructs

pENTR-ZIC3-wt: full length human *ZIC3* CDS was recovered from the HA-ZIC3-wt expression plasmid, a kind gift from Dr Stephanie Ware (Ware *et al.*, 2004). PCR amplification (section 2.3.8.1) with primers Ark1152_F (5`- ATC CGG TAC Cga att cAC CCT CTC TCA CTT CGG -3`) and Ark1153_R (5`- GTG CGG CCG Cga att cCC GCT CTA GAA CTA GTG G -3`) generated *ZIC3* amplicons that were subsequently cloned into pENTR-3C. Each primer includes a 15 bp region of homology (underlined) with the pENTR-3C vector and an additional base pair that regenerated the *EcoR*I site (lower case) for cloning with the In-Fusion™ Dry-Down PCR Cloning kit (Clontech; Cat. No. 639609) (section 2.3.7.1).

V5-ZIC3-wt: mammalian expression construct containing V5-tagged human *ZIC3* wild-type CDS, downstream of a CMV promoter. Plasmid was constructed with the Gateway® LR recombination cloning system (section 2.3.7.2) using pENTR-ZIC3-wt and V5-DEST.

EGFP-ZIC3-wt: full length human *ZIC3* CDS was recovered from HA-ZIC3-wt via PCR (section 2.3.8.1) using primers Ark1208_F (5`- GAG CTC AAG CTT Cga att cTA CCC TCT CTC ACT TCG G -3') and Ark1209_R (5`- TAC CGT CGA CTG CAg aat tcC CGC TCT AGA ACT AGT G -3`). Each primer includes a 15 bp region of homology (underlined) with the EGFP-C1 vector and the *EcoR*I site (lower case) for cloning with the In-Fusion™ Dry-Down PCR Cloning kit (Clontech; Cat. No. 639609) (section 2.3.7.1).

pENTR-ZIC2-wt: full length human *ZIC2* CDS was amplified from pcDNA-ZIC2, a generous gift from Prof. M. Maral Mouradian (Yang *et al.*, 2000). PCR amplification (section 2.3.8.1) with primers Ark1150_F (5`- ATC CGG TAC Cga att cAG TGT GGT GGA ATT CCT GGC C -3`) and Ark1168_R (5`- GTG CGG CCG Cga att cGA GGG TTA GGG ATA GGC TTA C -3`) generated *ZIC2* amplicons that were subsequently cloned into pENTR-3C. Each primer includes a 15 bp region of homology (underlined) with the pENTR-3C vector and an additional base pair that regenerated the *EcoR*I site (lower case) for cloning with the In-Fusion™ Dry-Down PCR Cloning kit (Clontech; Cat. No. 639609) (section 2.3.7.1).

V5-ZIC2-wt: mammalian expression construct containing V5-tagged human *ZIC2* wild-type CDS, downstream of a CMV promoter. Plasmid was constructed with the Gateway® LR recombination cloning system (section 2.3.7.2) using pENTR-ZIC2-wt and V5-DEST.

pENTR-ZIC5-wt: full length human *ZIC5* CDS was excised from pCMV6-XL5-ZIC5 (Origene) via *EcoR*I digestion and ligated into pENTR-3C using T4 DNA ligase (NEB; Cat. No. M0202L). Constructed by M. Zavortink.

V5-ZIC5-wt: mammalian expression construct containing V5-tagged human *ZIC5* wild-type CDS, downstream of a CMV promoter. Plasmid was constructed (by R. Ali) with the Gateway® LR recombination cloning system (section 2.3.7.2) using pENTR-ZIC5-wt and V5-DEST.

pENTR-Cdx2: full length mouse *Cdx2* CDS was amplified from pcDNA-Cdx2, a generous gift from Prof. Stefan Broer (RSB, ANU). PCR amplification with primers Ark1420_F (5`- Tgg atc cAT GGT GAG GTC TGC TCT G -3`) and Ark1168 (5`- Act cga gAG CTG TTC GTG GGT AGG A-3`) generated *Cdx2* amplicons containing restriction sites (lower case letters in primer sequence): *BamH*I on 5`-end and *Xho*I on 3`-end. Amplicons and pENTR-1A 'Entry' vector were subjected to restriction enzyme digested, followed by ligation. This plasmid was made by K. Barratt.

V5-Cdx2: mammalian expression construct containing V5-tagged mouse *Cdx2* wild-type CDS, downstream of a CMV promoter. Plasmid was constructed (by K. Barratt) with the Gateway® LR recombination cloning system (section 2.3.7.2) using pENTR-Cdx2 and V5-DEST.

2.2.3 Mutant ZIC3 constructs

2.2.3.1 Pathogenic variants

EGFP-ZIC3-katun: full length human *ZIC3-katun* CDS was recovered from HA-ZIC3-katun (made by M. Zavortink) via PCR (section 2.3.8.1) using primers Ark1208_F-Ark1209_R, and the plasmid generated using the same method as for EGFP-ZIC3-wt.

PTC-containing mutants (pENTR-ZIC3-katun, pENTR-ZIC3-E155X, pENTR-ZIC3-C268X, pENTR-ZIC3-Q292X, pENTR-ZIC3-1507insTT, and pENTR-ZIC3-K408X): all contained a nonsense mutation, except pENTR-ZIC3-1507insTT that had a frame-shift (mutation occurred after ZF2 that installed a stop codon 252 bases downstream in ZF5). Construct names show the particular amino acid (E155, C268, Q292 and K408) mutated to a stop codon (X) with numbers indicating length of the translated protein product. All mutations were installed using site-directed mutagenesis (section 2.3.3.1). pENTR-ZIC3-E155X was made by K. Diamand.

ZF1 mutants (pENTR-ZIC3-C253S, pENTR-ZIC3-W255G, and pENTR-ZIC3-H286R): each mutant contained a missense mutation at a conserved residue in ZF1 (C253, W255, H286), and were made by K. Diamand using site-directed mutagenesis (section 2.3.3.1).

2.2.3.2 Synthetic variants

ZF mutants (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 2.3.3.1). All constructs, except pENTR-ZIC3-C365S, were made by K. Diamand.

Deletion mutants (pENTR-ZIC3-ZOCdel, pENTR-ZIC3-ZFNCdel, pENTR-ZIC3-Ndel): mutants were missing either the ZOC domain (ZIC3-ZOCdel), the ZFNC domain (ZIC3-ZFNCdel) or the whole N-terminus region preceding the ZFD (ZIC3-Ndel). Deletions were made by overlap extension PCR (section 2.3.3.2).

All 'Entry' clones (pENTR-) were transferred to the V5-DEST expression vector via a Gateway® LR Clonase reaction (Life Technologies) (section 2.3.7.2) to generate **V5-ZIC3-katun**, **V5-ZIC3-E155X**, **V5-ZIC3-C268X**, **V5-ZIC3-Q292X**, **V5-ZIC3-1507insTT**, **V5-ZIC3-K408X**, **V5-ZIC3-C253S**, **V5-ZIC3-W255G**, **V5-ZIC3-H286R**, **V5-ZIC3-C268S**, **V5-ZIC3-C302S**, **V5-ZIC3-C335S**, **V5-ZIC3-C365S**, **V5-ZIC3-C393S**, **V5-ZIC3-ZOCdel**, **V5-ZIC3-ZFNC** and **V5-ZIC3-ZFD**.

2.2.4 Reporter constructs

2.2.4.1 Genomic promoters

B:*luc*$^{+}$:APOE: pXP2-APOE (-189/+1) was a gift from Prof. Francisco Zafra (Salero *et al.*, 2001). The vector contains -189 to +1 bases of the *Apolipoprotein E* (*APOE*) gene promoter, upstream of the *Photinus pyralis* luciferase gene. The empty pXP2 vector is 'promoterless' hence the expression of luciferase protein is driven by stimulation of the *APOE* promoter by binding of a transcription factor.

B:*luc2*:APOE and B:*luc2*:APOE:β-globin: the *APOE* promoter was PCR amplified (section 2.3.8.1) from B:*luc*$^{+}$:APOE using primers Ark1281_F (5`- ATA TTg gta ccA AGC TCA GAT CCA AGC TTG GGA CTG TGG G- 3`) and Ark1282_R (5`- GGA ATG aga tct TCA CAT CTC GAG AGG ACT CAA GGA TCC C- 3`), to generate *APOE* amplicons containing restriction sites (lower case letters in primer

sequence): *Kpn*I on 5`-end and *Bgl*II on 3`-end. The B:*luc2* and B:*luc2*:β-globin (section 2.2.4.2) vectors and *APOE* amplicons were subjected to restriction enzyme digests (sections 2.3.4 and 2.3.8.2), followed by ligation (section 2.3.8.3).

B:*luc2*:Nanog and B:*luc2*:Nanog:β-globin: the *Nanog* promoter was PCR amplified (section 2.3.8.1) from mouse genomic DNA using primers Ark1543_F (5`- TTg gta ccC TGG GTC ACC TTA CAG C -3`) and Ark1544R (5`- CGa gat ctT ATT CTC CCA GGC ACC C -3`), to generate *Nanog* amplicons containing restriction sites (lower case letters in primer sequence): *Kpn*I on 5`-end and *Bgl*II on 3`-end. The B:*luc2* and B:*luc2*:β-globin (section 2.2.4.2) vectors and *Nanog* amplicons were subjected to restriction enzyme digests (sections 2.3.4 and 2.3.8.2), followed by ligation (section 2.3.8.3).

2.2.4.2 Minimal promoters

B:*luc2*:β-globin, B:*luc2*:c-fos and B:*luc2*:TK: to preserve the MCS of B:*luc2*, the β-globin, c-fos and TK minimal promoters were cloned into the B:*luc2* vector using a single restriction enzyme digest. The B:*luc2* vector was linearized using *Hind*III (sections 2.3.4 and 2.3.8.2) and treated with Antarctic phosphatase (section 2.3.5). The minimal promoters were amplified using primers (Table 2.3) that added *Hind*III restriction sites on either side of the amplicons. Following digestion with *Hind*III, amplicons were ligated (section 2.3.8.3) into the linearized and dephosphorylated B:*luc2* vector. Correct orientation of promoter in the ligated plasmids was determined by DNA sequencing (section 2.3.3.4).

Promoter	Primer	Sequence (5`- 3`)	Template
β-globin	Ark1510_F Ark1507_R	TGa agc ttA TGA ACT AGT GGA TCC C CTa agc ttT CGG CTA GAA GCA AAT G	pKS:β-globin:lacZ
c-fos	Ark1515_F Ark1516_R	CAa agc ttC AAC CCC AGT GAC GTA GG GCa agc ttG CTC TAG AGA GAC ACT GG	TOPflash:c-fos
TK	Ark1513_F Ark1514_R	CGa agc ttA CGA CTC TAG AGG ATC CG GTa agc ttA TGA TCT GCG GCA CGC TG	TOPflash:TK

Table 2.3: Primers used for amplifying minimal promoters. Lower case letters show recognition site of *Hind*III enzyme. F: forward primer; R: reverse primer. The TOPflash:cfos plasmid was a gift from Dr. Sabine Tejpar (Department of Oncology, Katholieke Universiteit Leuven, Belgium). The TOPflash:TK plasmid was purchased from Upstate Biotechnology.

2.2.4.3 Synthetic enhancers

The ZIC DNA-binding sites identified by Salero *et al.* (2001) and Lim *et al.* (2010) (Figure 2.5A) were used to design six synthetic enhancer motifs (Figure 2.5B).

A

Binding site	Sequence (5`- 3`)	Reference
APOE-M1	GGA CTG TGG GGG GTG GTC AA	Salero *et al.* (2001)
APOE-M2	CTC CCT CTG CCC TGC TGT GC	
APOE-M3	CTA TCC CTG GGG GAG GGG GC	
Z3M1	CCC GCT GG	Lim *et al.* (2010)
Z3M2	CCC AGC GGG G	
Z3M8	CCC GCT GTG C	

B

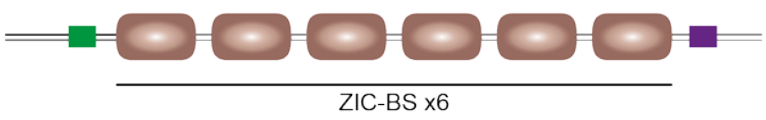


Figure 2.5: ZIC-specific synthetic enhancer motifs. (A) ZIC DNA-binding sites used for making synthetic enhancer motifs, which contained **(B)** six copies a particular binding site (ZIC-BS) separated by a 5-nucleotide spacer (5`- TAG AA -3`). The green and purple boxes denote restriction enzyme sites used for cloning.

B:*luc2*:APOE-M1, B:*luc2*:APOE-M2, and B:*luc2*:APOE-M3: motifs were ordered as cloned inserts in the pUC57 vector (Ampicillin resistant) (GenScript). Plasmids were transformed into *E. coli* by electroporation (section 2.3.1) and isolated (section 2.3.2). APOE-M1 and APOE-M3 were released from the pUC57 vector using *Kpn*I and *Bgl*II restriction enzymes, while APOE-M2 was released using *Kpn*I and *Hind*III (section 2.3.4). Inserts were separated from digested vector by agarose gel electrophoresis (section 2.1.3). The digested inserts were excised from the gel and extracted using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel; Cat. No. 740609.50), followed by ligation (section 2.3.8.3) into B:*luc2* (previously digested with the appropriate combination of restriction enzymes).

B:*luc2*:Z3M1, B:*luc2*:Z3M2, and B:*luc2*:Z3M8: motifs were ordered as forward and reverse oligomers (Gene Link™), with restriction enzyme sites on either side of the synthetic enhancer motif: *Kpn*I on 5`-end and *Hind*III on 3`-end. The complimentary oligonucleotides were mixed in a 1:1 molar ratio and diluted to a concentration of 1 pmol/µL in oligo annealing buffer [10 mM Tris (pH 8.0), 0.1 mM EDTA, and 50 mM Sodium Chloride (NaCl)]. Annealing was carried out in Eppendorf Mastercycler® using the following program: 95°C for 5 mins, followed by 1°C/min decrease in temperature for 70 mins. Annealed oligomers were precipitated (section 2.3.6), followed by digestion with *Kpn*I and *Hind*III (section 2.3.4) and ligated (section 2.3.8.3) into B:*luc2* (previously digested with *Kpn*I and *Hind*III).

B:*luc2*:Z3M2:β-globin, B:*luc2*:Z3M2:c-fos, B:*luc2*:Z3M2:TK and B:*luc2*:Z3M8:β-globin: the B:*luc2*:Z3M2 and B:*luc2*:Z3M8 vectors were digested using *Hind*III (sections 2.3.4 and 2.3.8.2) and treated with Antarctic phosphatase (section 2.3.5). The minimal promoters (previously amplified; section 2.2.4.2) were ligated (section 2.3.8.3) into (linearized and dephosphorylated) B:*luc2*:Z3M2 and B:*luc2*:Z3M8 vectors. Correct orientation of promoter in the ligated plasmids was determined by DNA sequencing (section 2.3.3.4).

2.3 MOLECULAR CLONING

2.3.1 Bacterial transformation

DNA was transformed into bacteria via heat-shock treatment or electroporation. Plasmids generated from In-Fusion™ reactions and site-directed mutagenesis were transformed using heat-shock (according to manufacturer's instructions). All other transformations were done by electroporation using DH5α electro-competent *Escherichia coli* (*E. coli*) cells.

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. The mixture was transferred to an ice-cold cuvette and shocked at 2.5 kV using a BioRad *E. coli* Pulser (BioRad). Immediately after pulsing 500 µL of Luria-Bertani (LB) Broth media (Bacto; Cat. No. 244620) were added to revive the bacteria and solution was incubated for 1 hr 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 hours at 37°C.

2.3.2 Plasmid isolation

Depending on the amount of DNA required, plasmid DNA was isolated using two methods.

1. To screen for mutagenized or cloned plasmids, single colonies from LB Agar plates were picked and each incubated in 4 mL of LB Broth (containing 0.1 mg/mL of the appropriate antibiotic) at 37°C for 16 hours on a rotating platform. 3 mL of this overnight culture were then used to isolate plasmid DNA using High Pure Plasmid Isolation kit (Roche; Cat. No. 11754785001) (average yield: 7-10 µg).
2. To obtain a higher quantity and/or purity of plasmid DNA, single colonies from LB Agar plates were picked and each incubated in 2 mL of LB Broth (containing 0.1 mg/mL of the appropriate antibiotic) at 37°C for 6-8 hours on a rotating platform. This culture was added to 100 mL of LB Broth (containing 0.1 mg/mL of the appropriate antibiotic) and

incubated at 37°C for 16 hours on a rotating platform. Plasmid DNA was isolated using NucleoBond® Xtra Midi (Macherey-Nagel; Cat. No. 740410.50).

In each case DNA was resuspended in Analar H_2O, instead of the supplied buffer.

2.3.3 PCR based mutagenesis and screening

2.3.3.1 Site-directed mutagenesis

Point mutations or insertions in *ZIC3* CDS were installed using the QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies; Cat. No. 210518) by following manufacturer's instructions. Mutagenesis PCRs were performed in a 50 µL volume using 100 ng of pENTR-ZIC3-wt and 0.25 µM of each primer (Table 2.4). Cycling parameters are shown in Table 2.5. Upon completion of PCR, samples were treated with 2 µL of *Dpn*I enzyme (supplied with kit) by incubation at 37°C for 1 hour. 2 µL of the reaction mix was then transferred into 45 µL XL10-Gold® ultracompetent bacteria cells (supplied with kit) and transformed via heat shock.

Mutation	Primer	Sequence (5`- 3`)
ZIC3-katun	Ark1006_F	GCC TAT CAA GCA G**T**A GCT GTC GTG CAA GTG
	Ark1007_R	CAC TTG CAC GAC AGC T**A**C TGC TTG ATA GGC
ZIC3-E155X	Ark1655_F	GGC ATC CCC **T**AG CCC CCT AGC TAC TTG
	Ark1656_R	CAA GTA GCT AGG GGG CT**A** GGG GAT GCC
ZIC3-C268X	Ark1399_F	CGG CCC AAG AAG AGC TG**A** GAC CGG ACC TTC AGC
	Ark1400_R	GCT GAA GGT CCG GTC **T**CA GCT CTT CTT GGG CCG
ZIC3-Q292X	Ark1401_F	GTG GGG GGC CCG GAG **T**AG AAC AAC CAC GTC TGC
	Ark1402_R	GCA GAC GTG GTT GTT CT**A** CTC CGG GCC CCC CAC
ZIC3-1507insTT	Ark1403_F	CCG AGT GCA CAC GGG C**TT** GAG AAG CCC TTC CCA
	Ark1404_R	TGG GAA GGG CTT CTC **AA**G CCC GTG TGC ACT CGG
ZIC3-K408X	Ark1405_F	CTG CGC AAA CAC ATG **T**AG GTT CAT GAA TCT CAA
	Ark1406_R	TTG AGA TTC ATG AAC CT**A** CAT GTG TTT GCG CAG
ZIC3-C253S	Ark1649_F	CAG GAG CTG TCG T**C**C AAG TGG ATC GAC
	Ark1650_R	GTC GAT CCA CTT G**G**A CGA CAG CTC CTG
ZIC3-W255G	Ark1651_F	CGT GCA AG**G** GGA TCG ACG AGG CTC
	Ark1652_R	GAG CCT CGT CGA TCC **C**CT TGC ACG
ZIC3-W255G	Ark1653_F	GTC ACC ATG GAG C**G**T GTG GGG GGC
	Ark1654_R	GCC CCC CAC A**C**G CTC CAT GGT GAC
ZIC3-C268S	Ark1608_F	CCC AAG AAG AGC T**C**C GAC CGG ACC TTC AGC
	Ark1609_R	GCT GAA GGT CCG GTC G**G**A GCT CTT CTT GGG
ZIC3-C302S	Ark1610_F	GCT ACT GGG AGG AGT **C**CC CCC GGG AG
	Ark1611_R	CTC CCG GGG G**G**A CTC CTC CCA GTA GC
ZIC3-C335S	Ark1612_F	GCT ACT GGG AGG AGT **C**CC CCC GGG AG
	Ark1613_R	CTC CCG GGG G**G**A CTC CTC CCA GTA GC
ZIC3-C365S	Ark1179_F	GAA TTT GAA GGC **A**GT GAC AGA CGC TTT GC
	Ark1180_R	GCA AAG CGT CTG TCA C**T**G CCT TCA AAT TC
ZIC3-C393S	Ark1614_F	CCC TAT ATC TGC AAA GTG T**C**C GAC AAG TCC TAC ACG
	Ark1615_R	CGT GTA GGA CTT GTC G**G**A CAC TTT GCA GAT ATA GGG

Table 2.4: Primers used for site-directed mutagenesis. Primers were designed based on manufacturer guidelines of the QuikChange Lightning Site-Directed Mutagenesis kit. Red letters represent the mutated base(s). F: forward primer; R: reverse primer.

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

Table 2.5: Thermal cycling parameters for site-directed mutagenesis.

2.3.3.2 Overlap extension PCR to create deletion mutants

The most critical aspect of using overlap extension PCR, for deleting regions within a CDS, is primer design (Table 2.6). Primers are constructed to amplify the entire plasmid sequence, 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 2.6). Primers bind template via their 3`-ends, while their 5`-ends contain complimentary overhangs. Thus the 5`-end of mutant strand from each primer contains a region of homology (at least 15-16 bp) that allows annealing into a double-stranded DNA fragment.

Mutation	Primer	Sequence (5`- 3`)
ZIC3-ZOCdel	Ark1644_F	CCT TCA ACT CAA GCT CCG GGC TCA GTG A
	Ark1643_R	CGG AGC TTG AGT TGA AGG CGG CAG AG
ZIC3-ZFNCdel	Ark1648_F	ACG GTC CTC TGT CGT GCA AGT GGA TCG
	Ark1647_R	ACG ACA GAG GAC CGT GGT GGG CCG
ZIC3-Ndel	Ark1675_F	CGA ATT CAT GTG CAA GTG GAT CGA CGA G
	Ark1676_R	CTT GCA CAT GAA TTC GGT ACC GGA TCC AG

Table 2.6: Primer design for creation of deletion mutants using overlap extension PCR. Nucleotides in black show sequence that binds to template. Nucleotides in purple show overhangs and the underlined sequence is the region of complementarity between a primer pair. The green nucleotides were included in sequence to install the start codon (ATG), since the ZIC3-Ndel mutant lacks the first methionine due to deletion of the whole N-terminus. F: forward primer; R: reverse primer.

Mutagenesis PCRs were performed with the *PfuUltra* II Hotstart PCR Master Mix (Agilent Technologies; Cat. No. 600850) in a 50 µL volume using 10 ng of pENTR-ZIC3-wt, 0.2 µM of each primer and the cycling conditions shown in Table 2.7. When reaction was complete DNA was precipitated (section 2.3.6) followed by digestion with *Dpn*I enzyme (NEB; Cat. No. R0176) for 2 hours. DNA was then precipitated (section 2.3.6) and transformed into *E. coli* (section 2.3.1). Bacterial colonies were screened for the correct plasmid via colony PCR (section 2.3.3.3) and/or DNA sequencing (section 2.3.3.4).

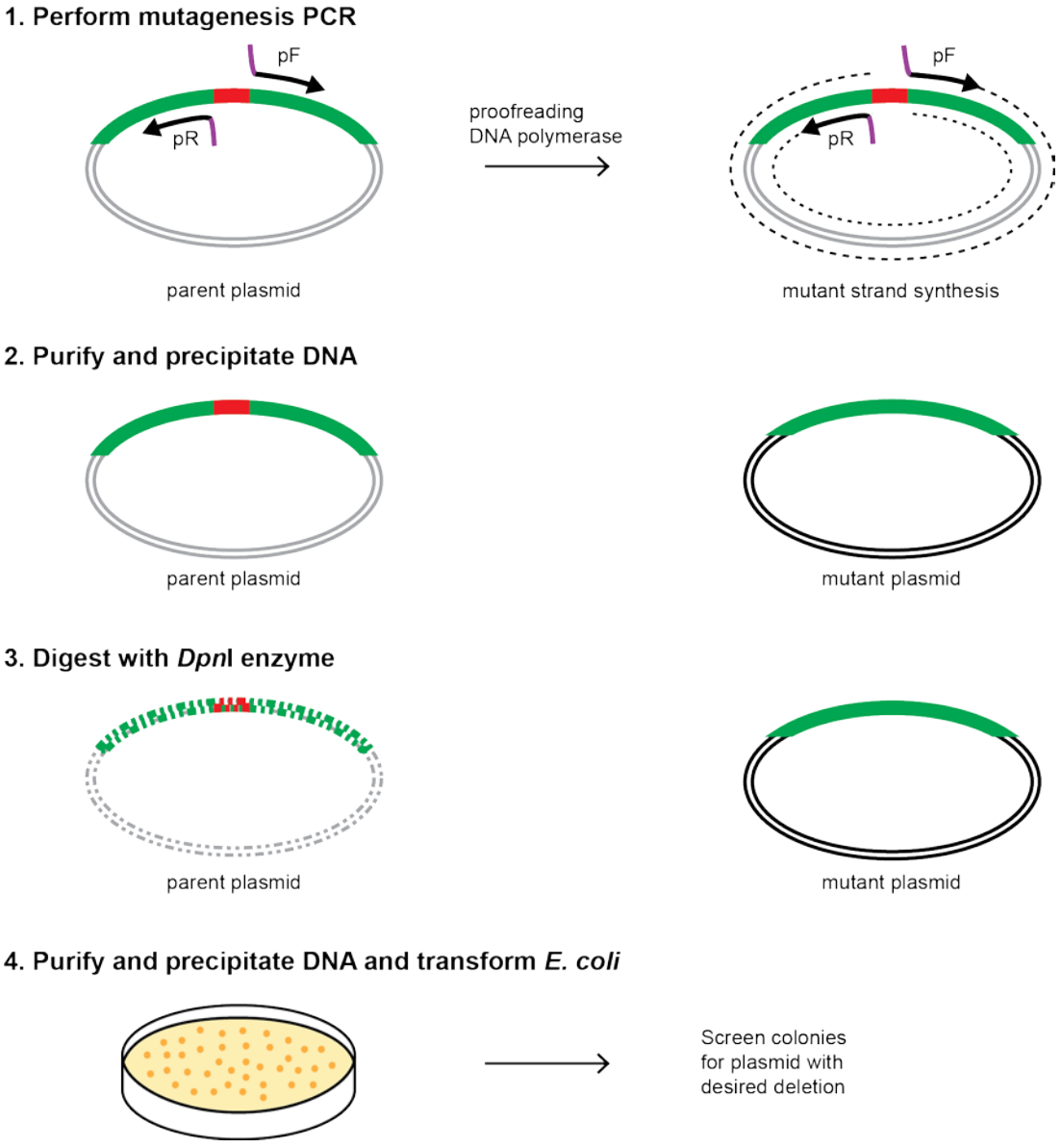


Figure 2.6: Creating deletion mutants by overlap extension PCR. 1: primers (pF: forward primer; pR: reverse primer) are designed to have their 5`-ends facing each other and 3`-ends oriented such that the DNA polymerase extends the entire plasmid, except the region to be deleted. PCR is performed with a proofreading DNA polymerase. **2:** PCR products are precipitated (to isolate DNA form reaction mixture) and **3:** treated with *Dpn*I to digest the template methylated plasmid (Lacks *et al.*, 1975). **4:** Following another round of DNA precipitation, bacteria are transformed with the newly synthesized mutant plasmid. Screening of colonies can be done via colony PCR and/or DNA sequencing.

Program	Temperature	Time	Cycles
Initial denaturation	95°C	2 mins	1
Denature	95°C	20 secs	30
Anneal	60°C	20 secs	
Extend	72°C	57 secs	
Polish	72°C	3 mins	1

Table 2.7: Thermal cycling parameters for overlap extension PCR.

2.3.3.3 Bacterial colony screening

To identify bacterial colonies containing the desired plasmids, colony PCRs were performed. Single colonies were picked from agar plates using a sterile pipette tip and resuspended in 20 µL of Analar H_2O. PCR reactions were performed using the ThermoPrime ReddyMix PCR Mastermix (Thermo Scientific; Cat. No. AB-0575/DC/LD/B), 0.67 µM of primers (Table 2.8) and 5 µL of resuspended bacteria (as template) in a 20 µL reaction volume. Cycling parameters are shown in Table 2.8B.

A

Parent plasmid	Primer	Sequence (5`- 3`)	T_A
B:*luc2*	Ark1566_F Ark1567_R	CAT TTC TCT GGC CTA ACT GG AAC AGT ACC GGA TTG CCA AG	55°C
pENTR-ZIC3-wt	Ark1673_F Ark1674_R	GAC TGG ATC CGG TAC CG GTC TAG ATA TCT CGA GTG CG	50°C

B

Program	Temperature	Time	Cycles
Initial denaturation	95°C	10 mins	1
Denature	95°C	30 secs	30
Anneal	50 - 55°C	30 secs	
Extend	72°C	30 - 90 secs	
Polish	72°C	7 mins	1

Table 2.8: Bacterial colony PCR parameters. (A) Primers used for screening of colonies containing the desired plasmids from section 2.2.4 (B:*luc2* reporters) or section 2.2.3.2 (ZIC3 deletion mutants). T_A: annealing temperature. **(B)** Cycling parameters for colony PCR. Annealing temperature varied depending on the primers used. Extension time (30 secs per 0.5 kb) varied depending on the length of DNA to be amplified.

2.3.3.4 DNA sequencing

150-200 ng of plasmid DNA or 3-20 ng of PCR product was sequenced using 3 pmol primer and 1 µL of BigDye® Terminating mixture (ThermoFisher Scientific) in Sequencing Buffer (70 mM Tris-HCl and 1.75 mM $MgCl_2$). The reaction was run in an Eppendorf Mastercycler® using the cycling parameters shown in Table 2.9.

Program	Temperature	Time	Cycles
Initial denaturation	94°C	5 mins	1
Denature	96°C	10 secs	30
Anneal	50°C	5 secs	
Extend	60°C	4 mins	

Table 2.9: Thermal cycling parameters for DNA sequencing PCR.

Upon completion of sequencing PCR, the amplified DNA was precipitated to remove excess dye and nucleotides. To each well containing amplified DNA, 2 µL of 125 mM EDTA (Ajax Chemicals) at pH 8.0, 3 µL of 3 M Sodium Acetate (Sigma-Aldrich; Cat. No. S2889) at pH 5.2 and 50 µL of 100% ethanol were added. After a gentle vortex the reactions were incubated for 15 mins at room temperature in a dark environment and DNA was pelleted by centrifugation at 3,220 g for 30 mins at 4°C. To remove supernatant, the PCR plate was flipped (face-down) and spun at 100 g for 1 min at 4°C. When residual ethanol had evaporated, 20 µL of HiDi Formamide (Applied Biosystems™; Cat. No. 4311320) were added to each reaction. Sequencing samples were submitted to the RSB Sequencing Facility for processing using their ABI 3730 sequencer (Applied Biosystems). The resulting sequences were analysed using the Geneious software (Biomatters; version 5.5.9).

2.3.4 Restriction Enzyme Digests

DNA was digested with restriction enzymes for preparation of inserts and vectors for cloning or analysis of plasmids post-ligation. For plasmids 1 µg of was digested using 5 U of appropriate restriction enzyme (Table 2.10), while for PCR products 1 µg was digested using 1 U of enzyme. Reactions were performed with the appropriate NEBuffer (10X stock), Bovine Serum Albumin (BSA; 10X stock) and enough Analar H_2O such that the final volume was 10-20 times greater than the volume of restriction enzyme added. Reactions were incubated at 37°C for 1-3 hours.

Enzyme	Target site	Cat. No.	NEBuffer
*Bgl*II	A/GATCT	R0144	NEB2; in DD with *Kpn*I
*EcoR*I	G/AATTC	R0101	*EcoR*I
*Hind*III	A/AGCTT	R0104	NEB2
*Kpn*I	GGTAC/C	R0142	NEB2; in DD with *Bgl*II or *Hind*III
*Dpn*I	GA^{m}/TC	R0176	NEB2

Table 2.10: Restriction enzymes used to digest DNA. All enzymes were purchased from New England BioLabs® (NEB). *Dpn*I enzyme digests methylate DNA. m: methyl (CH_3) group; DD: double digest.

2.3.5 Antarctic phosphatase treatment

For ligations involving a single restriction enzyme digest, vectors were dephosphorylated using Antarctic phosphatase (NEB; Cat. No. M0289), to prevent re-ligation. Upon completion of digestion, the enzyme was heat-inactivated (by incubating at 65°C for 20 mins) and 2 U of Antarctic phosphatase were added for every 1 µg of digested vector. Appropriate volume of the Antarctic phosphatase buffer (NEB; Cat. No. B0289) was added and the reaction incubated at 37°C for 30 mins, followed by heat-inactivation of the phosphatase 65°C for 30 mins.

2.3.6 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 vortexed vigorously for 5 secs and centrifuged at 16,000 g for 20 mins. Supernatant was removed and 180 µL of 70% Ethanol at -20°C were added followed by another spin at 16,000 g for 2 mins. Supernatant was removed and tubes were left on the bench to air-dry. Once no Ethanol could be seen in the tubes, 20 µL of Analar H_2O were 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 µL of Analar H_2O.

2.3.7 Recombination based cloning

2.3.7.1 In-Fusion™ Dry-Down PCR Cloning

In-Fusion™ cloning allows directional cloning of DNA fragments into any vector, using a proprietary enzyme that recombines DNA fragments containing (at least) 15 bp of homology with the site of insertion in the linearized cloning vector. *ZIC* CDS were cloned into pENTR-3C and/or EGFP-C1 vectors using the In-Fusion™ Dry-Down PCR Cloning kit, by following manufacturer's instructions. Prior to the In-Fusion™ reaction, inserts (*ZIC* CDS) and vectors (pENTR-3C or EGFP-C1) were digested with *EcoR*I and dephosphorylated using Antarctic phosphatase. The digested fragments (200 ng of insert; 100 ng of vector) were added to the In-Fusion™ Dry-Down and the reaction volume brought up to 10 µL. Reaction mix was incubated first at 37°C for 15 mins, and then at 50°C for 15 mins before transferring on ice. The reaction was then transformed into bacterial cells via heat-shock.

2.3.7.2 Gateway® Recombination Cloning Technology

To obtain V5-epitope tagged expression plasmids the Gateway® Recombination Cloning Technology (Life Technologies) was used to transfer CDS from the 'Entry' (pENTR) vector to the 'Destination' (V5-DEST) expression vector, by following manufacturer's instructions. This method is based on the site-specific recombination properties of bacteriophage lambda (Landy, 1989) and relies on the presence of recombination sites (*att*L: pENTR and *att*R: V5-DEST) on vectors, and enzyme (e.g. LR Clonase) that catalyses the recombination reaction (Figure 2.7).

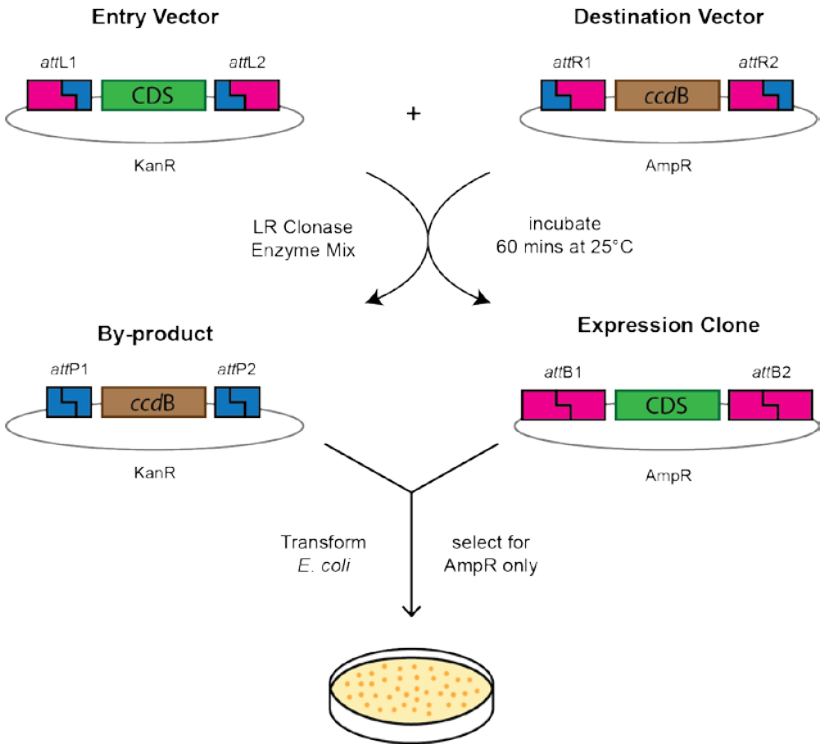


Figure 2.7: Gateway® LR Recombination Cloning. DNA sequences are exchanged between vectors using site-specific recombination. The 'Entry' vector contains the CDS to be transferred flanked by *att*L1 and *att*L2 sites and a Kanamycin resistance (KanR) CDS. The 'Destination' vector contains the *ccd*B gene flanked by *att*R1 and *att*R2 sites and an Ampicillin resistance (AmpR) CDS. The vectors are mixed with the LR Clonase Enzyme mix and incubated at 25°C for 60 mins. The LR Clonase enzyme catalyses the recombination between the *att*L and *att*R sites of the vectors, allowing transfer of CDS into the 'Destination' vector to create the desired Expression clone. DNA from the reaction is precipitated (section 2.3.6), transformed into DH5α bacteria and plated onto agar plates containing Ampicillin. Bacteria carrying the 'Entry' vector or by-product will be non-viable due to lack of AmpR. Bacteria transformed with the original 'Destination' plasmid will be non-viable due to lethality of the *ccd*B gene for the DH5α strain. Thus the only surviving bacteria on the plate will be carrying the desired Expression clone.

2.3.8 Ligation based cloning

2.3.8.1 Insert preparation

Inserts were prepared via PCR amplification. Five PCR reactions were performed to amplify the region of interest from template. Each PCR was performed in 10 µL volume using 10 ng of template, 0.25 µM of each primer and the cycling condition shown in Table 2.11.

Program	Temperature	Time	Cycles
Initial denaturation	94°C	10 mins	1
Denature	94°C	15 secs	35
Anneal	50 - 60°C	30 secs	
Extend	72°C	30 secs	
Polish	72°C	7 mins	1

Table 2.11: Thermal cycling parameters for insert amplification. Annealing temperature varied depending on the primers used (listed in section 2.2)

Reactions were pooled and DNA precipitated using Ethanol precipitation (section 2.3.6). Amplified DNA was dissolved in Analar H_2O and separated from template using agarose gel electrophoresis (section 2.1.3). The correct sized fragment was excised from agarose gel and DNA extracted using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel; Cat. No. 740609.50). The gel-extracted PCR product was subjected to a restriction digest (1 µg of PCR product was cut using 1 U of appropriate restriction enzyme, by incubating at 37°C for 60 mins), followed by agarose gel electrophoresis to obtain digested and purified insert fragments.

2.3.8.2 Vector preparation

10 µg of vector DNA was digested using 50 U of appropriate restriction enzyme by incubating at 37°C for 3 hours, followed by Antarctic phosphatase treatment (when required). DNA was purified using Ethanol precipitation (section 2.3.6) and digested vector was separated from undigested vector using agarose gel electrophoresis (section 2.1.3) by running at 5 Vcm^{-1} for 10 mins, followed by 0.5 Vcm^{-1} for 16 hours. Digested vector was excised from agarose gel and DNA extracted using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel; Cat. No. 740609.50).

2.3.8.3 Ligation reaction

All ligation reactions were performed using T4 DNA Ligase (NEB; Cat. No. M0202L) in a 20 µL volume. Each ligation reaction was carried out using 100 U of Ligase and 50 ng of digested vector. Amount of insert was calculated using the following formula:

$$\text{Amount of insert (ng)} = \frac{\text{size of insert (bp)}}{\text{size of vector (bp)}} \times 50 \text{ ng (amount of vector)}$$

Ligations were performed using insert:vector ratios of 3:1 and 10:1. When all components were added reactions were incubated at 22°C for 20 mins. Ligated plasmid DNA was precipitated by Ethanol precipitation (section 2.3.6) and transformed into DH5α *E. coli* by electroporation (section 2.3.1).

2.4 CELL CULTURE

2.4.1 Cell lines and culture conditions

Mammalian cell lines African green monkey kidney fibroblast-like cell line (COS-7), Mouse embryonic fibroblast cells (NIH3T3), Human embryonic kidney cells (HEK293T), Human primary glioblastoma (U87) and Mouse embryonic mesenchymal cell line (C3H-10T1/2) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma-Aldrich; Cat. No. D6429). Chinese Hamster Ovarian cell line (CHO) was cultured in Ham's F-12 Nutrient Mix (Life Technologies; Cat. No. 11765-054). All culture media were supplemented with 10% (v/v) fetal bovine serum (Gibco™; Cat. No. 10100-147), 2 mM L-Glutamine (Gibco™; Cat. No. 25030-081) and 0.1 mM non-essential amino acid solution (Gibco™; Cat. No. 11140-50), and cells were grown in a humidified incubator at 37°C with 5% CO_2.

For routine passaging, plating medium was discarded and cells were washed once with 1x Phosphate Buffered Saline (PBS) solution (140 mM NaCl, 3 mM KCl, 10 mM phosphate buffer pH = 7.4; Amresco). The wash solution was discarded and cells were dissociated from the growth surface by adding 0.5 g/L trypsin (Gibco™; Cat. No. 15400-054) and incubating at 37°C for 5 mins. Supplemented DMEM was added to inactivate the trypsin and the entire cell/DMEM solution was repeatedly pipetted to form a single cell suspension. The desired amount of cells were then transferred to new tissue-culture treated plasticware containing fresh supplemented DMEM.

All experiments reported in Chapter 3, 4 and 5 were performed using HEK293T cells (see Appendix 1)

2.4.2 Transfection

Cells were transfected at 60-90% confluency using Lipofectamine™ 2000 (Invitrogen™; Cat. No. 11668-019) as per manufacturer's guidelines. Depending on the type of experiment, the size of tissue-culture treated plasticware used varied, thus amount of DNA and Lipofectamine™ 2000 had to be scaled (Table 2.12).

Dish type	DMEM (µL)	Lipofectamine™ 2000 (µL)	total DNA (µg)
96 well plate	2x 25	0.5	0.6
12 well plate	2x 100	4	1.6
6 well plate / 35 mm dish	2x 250	10	4
10 cm dish	2x 1500	60	24

Table 2.12: Transfection amounts based on culture plate. The total cell yield varied depending the dish/plate used for transfections, thus Lipofectamine™ 2000 and DNA amounts were adjusted accordingly. Initially separate dilutions were made for Lipofectamine™ 2000 and DNA e.g. in a 12-well plate, 4 µL of Lipofectamine™ 2000 were diluted in 100 µL of DMEM (no supplements) and 1.6 µg of DNA were diluted in another 100 µL DMEM (no supplements).

Most transfections were performed in a 12-well plate. Lipofectamine™ 2000 and DNA were individually diluted in DMEM (no supplements) and incubated at room temperature for 5 mins. The dilutions were mixed and incubated at room temperature for 20 mins. The Lipofectamine™ 2000 and DNA mixture was then added in a drop-wise manner to each well. The plate was gently rocked to mix and returned to the incubator.

2.4.3 Immunofluorescence (IF)

2.4.3.1 Staining and microscopy

Cells were grown in 35 mm tissue-culture dishes (Corning; Cat. No. CLS430165) that contained sterile 13 mm coverslips (ProSciTech; Cat. No. G402) and transfected with relevant expression constructs. 24 hours post-transfection, culture media was removed and replaced with 1x PBS. Coverslips were removed and placed in a humidifying chamber. Cells were fixed with 4% paraformaldehyde/PBS solution (ProSciTech; Cat. No. C004) at room temperature for 30 mins. Coverslips were washed thrice with 1x PBS and cells were then permeabilized with 0.25% Triton™ X-100/PBS (Sigma-Aldrich; Cat. No. T9284). Coverslips were washed thrice with 1x PBS and then incubated with 5% skim milk/PBS solution (IF blocking buffer) for 16 hours at 4°C. After blocking, cells were exposed to primary antibodies [mouse α-V5 (Anti-Xpress™; Cat. No. R960-

25) at a 1:200 dilution; or rabbit α-GFP (Cell Signalling; Cat. No. #2555) at a 1:300 dilution, and rabbit α-LaminB1 (Abcam; Cat. No. ab16048) at a 1:1000 dilution] diluted in IF blocking buffer, and incubated at room temperature for 1 hour. Coverslips were then washed six times with IF blocking buffer and incubated with the appropriate fluorescent dye (Alexa Fluor® 488 or Alexa Fluor® 594) conjugated secondary antibody (α-mouse and α-rabbit at a 1:500 dilution; Molecular Probes, Invitrogen) diluted in IF blocking buffer, and incubated at room temperature for 1 hour. Transfected proteins (V5- or EGFP-tagged) and LaminB1 were detected using different fluorophores. Coverslips were washed six times with 1x PBS, before mounting on to clear white glass slides (ProSciTech; Cat. No. G300FB) with anti-fade mounting agent [2.4 mM of n-propyl gallate (NPG, Sigma-Aldrich; Cat. No. P3130) in a 1:1 solution of glycerol and 1x PBS]. Edges of coverslips were sealed with nail polish. Cells were viewed using the LSM 5 Pascal (ZEISS) confocal microscope.

2.4.3.2 Quantification of subcellular localization

Subcellular distribution of transfected ZIC proteins in 20-100 transfected cells was scored blind per experiment. For each cell, the nuclear and cytoplasmic compartments were traced (Figure 2.8) 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). 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 (N_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. For statistical analysis, GenStat (VSN International) was used to perform a non-orthogonal factorial analysis of variance (ANOVA) of the data. Images were assembled in Adobe Illustrator CS5.

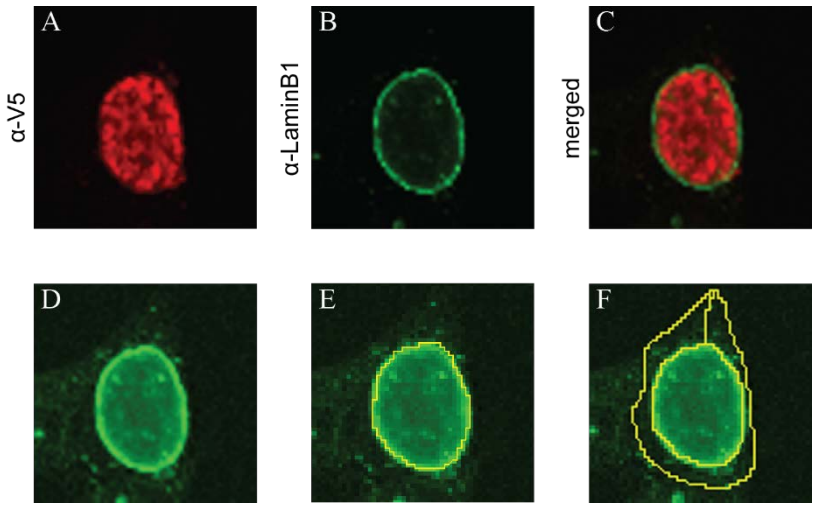


Figure 2.8: Quantifying subcellular localization. HEK293T cells were transfected with V5-ZIC3-wt and immunostained with **(A)** α-V5 (red) to detect the over-expressed ZIC protein, and **(B)** α-LaminB1 (green) to mark the nuclear envelope. **(C)** Merged image shows subcellular localization of the protein. **(D)** Exposure of the image shown in **B** was increased (using Adobe Photoshop CS5) to highlight the cytoplasmic boundary. **(E, F)** ImageJ software was used to encircle the perimeter and measure the area occupied by the **(E)** nuclear and **(F)** cytoplasmic compartments. Measurement of fluorescence intensity from each compartment was made by switching to the red channel **(A).**

2.4.4 Cell Lysis and Subcellular Fractionation

Cells were grown in 35 mm tissue-culture dishes or flat bottom cell-culture treated 12-well plates (Costar®, Cat. No. CLS3513), transfected with appropriate combination of plasmids and harvested for analysis. Proteins were extracted from cells and fractionated using the NE-PER™ Extraction kit (Thermo Scientific™; Cat. No. 78833) with some modifications to the manufacturer's protocol. 24 hours post-transfection cells were harvested in DMEM and centrifuged at 500 g for 5 mins. Supernatant was removed and (following volumes are for a 12-well plate) 50 µL of ice-cold CERI buffer with protease inhibitors (cOmplete™, EDTA-free Protease Inhibitor, Roche; Cat. No. 11873580001) added to the cell pellet, followed by vigorous vortexing for 15 secs and incubation on ice for 10 mins. 2.75 µL of CERII buffer were added, followed by vigorous vortexing for 5 secs and incubation on ice for 1 min. The cell solution was then centrifuged at 20,000 g for 15 mins at 4°C to pellet the nuclei and separate the cytoplasmic extract (supernatant). When spin was complete, supernatant was transferred to a clean pre-

chilled tube and ice-cold NERI buffer with protease inhibitors added to the pellet. The pellet was vortexed vigorously for 15 secs and incubated on ice for 40 mins, with 15 secs of vortexing every 10 mins. The tubes were then centrifuged at 20,000 g for 25 mins at 4°C to separate the nuclear proteins (supernatant) from debris (pellet). When spin was complete, supernatant was transferred to a clean pre-chilled tube. Volumes of reagents were scaled up for cells harvested from 35 mm tissue-culture dish.

2.4.5 Sodium Dodecyl Sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins in cell lysates were separated based on molecular weight (MW) on acrylamide gels. A 10% gel was used for all experiments (Table 2.13). Constituents of the running gel were mixed and cast on to the Mini-PROTEAN® casting apparatus (Bio-Rad), overlaid with hydrated butanol (Sigma-Aldrich; Cat No. B7906) and left to polymerize at room temperature for 60 mins. Butanol was then removed and top of the running gel washed with MilliQ water. The gel was then overlaid with a 3.75% stacking gel (Table 2.13) and a comb was inserted to create wells for loading samples. This was left to polymerize for another 60 mins.

Reagent	Volume
Running Gel (10%)	
40% Acryl:Bisacryl (Sigma-Aldrich; Cat. No. A7168)	3.125 mL
1.5 M Tris (pH 8.8) (Sigma-Aldrich; Cat. No. 252859)	3.125 mL
10% SDS solution (Sigma-Aldrich; Cat. No. L4390)	125 µL
Deionized H_2O	6 mL
10% Ammonium persulphate (Sigma-Aldrich; Cat. No. A3678)	125 µL
N,N,N',N'-Tetramethylethylenediaminse (Sigma-Aldrich; Cat. No. T9281)	125 µL
Stacking Gel (3.75%)	
40% Acryl:Bisacryl (Sigma-Aldrich; Cat. No. A7168)	235 µL
0.5 M Tris (pH 8.8) (Sigma-Aldrich; Cat. No. 252859)	625 µL
10% SDS solution (Sigma-Aldrich; Cat. No. L4390)	25 µL
Deionized H_2O	1.6 mL
10% Ammonium persulphate (Sigma-Aldrich; Cat. No. A3678)	50 µL
N,N,N',N'-Tetramethylethylenediaminse (Sigma-Aldrich; Cat. No. T9281)	5 µL

Table 2.13: SDS-PAGE gel recipes. Reagents were added to make gels in the order shown. Samples were loaded on to the stacking gel and proteins separated on the running gel. Volumes are for one gel of 1.5 mm thickness.

Samples for SDS-PAGE were prepared by adding Dithiothreitol (20x stock) (Sigma-Aldrich; Cat. No. D9163-5G) and NuPAGE® LDS Sample Buffer (4x stock) (Life Technologies; Cat. No. NP0008) to give a final concentration of 50 mM and 1x, respectively. Samples were heated for 5 mins at 90°C to denature proteins, followed by a brief spin to collect everything at bottom of the tube. The polymerized gels were put in a Mini-PROTEAN® Tetra Cell apparatus (Bio-Rad), which was filled with 1x SDS-PAGE Running Buffer (192 mM Glycine, 24.9 mM Tris base and 3.47 mM SDS). 30 µL of each sample were loaded into each well on the gel. On either side of the sample group, 10 µL of PageRuler™ Prestained Protein Ladder (Life Technologies; Cat. No. 26617) was added, which showed separation of proteins during electrophoresis and allowed to determine approximate size of proteins after Western blotting. When sample loading was complete, 100 V were applied to the gels until sufficient separation between relevant size markers on protein ladder was observed (1.75-2.5 hours). Proteins on gel were then transferred to a membrane via wet transfer.

2.4.6 Wet Transfer

When SDS-PAGE was complete, gel was removed from the running apparatus and the stacking portion of gel discarded. The gel was then placed on to a polyvinylidene difluoride (PVDF) membrane (previously activated via immersion in methanol) and together these were sandwiched between four pieces (two on either side) of blotting paper (3 mm chromatography paper Whatman®; Cat. No. 3030-917) and two sponges, within a gel holder cassette. The assembly was completed while the gel and all other components were submerged in 1x Towbin's Buffer (190 mM Glycine, 24 mM Tris base and 20% methanol). The gel holder cassette was then transferred to a Mini Trans-Blot Module (Bio-Rad) within a gel tank filled with 1x Towbin's Buffer. Proteins from the gel were transferred to the membrane by applying 15 V for 16 – 18 hours, while the tank was placed in ice to minimize heating.

2.4.7 Western Blotting (WB)

When transfer was complete, the PVDF membrane was removed from cassette and blocked with WB blocking buffer [5% skim milk/PBS/0.2% Tween 20 (Sigma-Aldrich; Cat. No. P7949) solution, except for the GFP antibody PBS was replaced by TBS (Tris-Buffered Saline: 50 mM Tris, 150 mM NaCl, pH 7.5)] at room temperature for 1 hour on a rotating platform. Membrane was then immunoblotted with primary antibodies (Table 2.14) for 2 hours on a rolling platform. Membrane was then washed four times using WB blocking buffer with each wash lasting 10 mins, followed by exposure to the appropriate Horse Radish Peroxidase (HRP) conjugated secondary antibody (Table 2.14) for 1 hour. Membrane was then washed four times using 0.2% Tween 20/PBS buffer, with each wash lasting 10 mins.

Primary Antibodies			
Target Protein	**Host Species (class)**	**Company (Cat. No.)**	**Dilution**
V5	Mouse (monoclonal)	Invitrogen (R960-25)	1:3000
GFP	Rabbit (polyclonal)	Cell Signalling (#2555)	1:1500
TBP	Mouse (monoclonal)	Abcam (ab818)	1:2000
LaminB1	Rabbit (polyclonal)	Abcam (ab16048)	1:1500
Secondary Antibodies			
Target Species	**Host Species**	**Company (Cat. No.)**	**Dilution**
Mouse	Rabbit	Invitrogen (61-6520)	1:5000
Rabbit	Donkey	Invitrogen (A16023)	1:5000

Table 2.14: Antibodies used for Western blotting. All antibody dilutions were carried out in 3 mL of WB blocking buffer.

To detect protein bands, membranes were incubated with SuperSignal™ West Pico Chemiluminscent Substrate (~9.26 µL per cm^2 of membrane: Thermo Scientific™; Cat. No. 34079) for 5 mins in a dark environment. The membrane was then exposed to an X-ray film (Amersham Hyperfilm ECL, GE Healthcare Life Sciences; Cat. No. 28-9068-37). Times of exposure varied from 10 secs to overnight, depending on visibility of desired bands. Exposed films were developed by immersion in Multigrade Paper Developer Solution (Ilford) for ~2 mins, washed in water to removed excess solution and fixed by immersion in Rapid Fixer Solution (Ilford). Film was washed again with water and left to air-dry. Developed films were scanned and assembled in Adobe Photoshop CS5.

2.4.8 Luciferase reporter assays

Cells grown on flat bottom cell culture treated 12-well plates (Costar®; Cat. No. CLS3513) were transfected with the relevant combination of constructs. For ZIC transactivation assays a total of 1.6 µg of DNA was added per well: 0.6 µg of the reporter construct and 1.0 µg of the expression construct (V5-ZIC) or the empty expression vector (V5-DEST). For ZIC competition assays a total of 1.5 µg of DNA were added per well: 0.5 µg of the B:*luc*$^+$:Apoe reporter construct,

0.5 µg of the wild-type ZIC expression construct and 0.5 µg of the PTC containing ZIC3 mutants or the V5-DEST vector when required to equalize the amount of transfected DNA. 6 hours post-transfection, cells were dissociated from the growth surface by replacing the culture medium with 0.5 g/L trypsin and incubating at 37°C for 5 mins. The trypsin was inactivated by adding DMEM (with supplements) and cells plated in triplicate on to a solid white tissue-culture treated 96-well plate (Corning®, Cat. no. CLS3917) at ~50% confluency and placed in the incubator. To avoid the potential problem of position bias by the luminometer, sample order on the plate was randomized for each independent experimental repeat. The remaining cells in the 12-well plate were returned to the incubator and harvested for WB analysis at the time of luminescence measurement. 18 hours after re-plating (or 24 hours post-transfection), cells in each well were lysed by addition of 100 µL of a 1:1 dilution of ONE-Glo™ Luciferase Assay System (Promega; Cat. No. E6110) with DMEM and luminescence measured using the GloMax®-96 Microplate Luminometer (Promega) or the TECAN Infinite M1000 Pro.

Luciferase activity was normalized to V5-DEST (to determine relative luciferase activity: RLA) and mean RLA and standard deviation (SD) calculated from three internal repeats (using Microsoft Excel). At least three independent experiments were performed for each assay (unless stated otherwise) with luminescence values from one representative experiment shown. For statistical analysis, GenStat was used to perform an analysis of variance (ANOVA) on raw luminescence data from the three independent repeats and calculate Standard Error of the Mean (SEM). SEM values for RLA were calculated by dividing the SEM of raw luminescence values by the mean luminescence value of V5-DEST from three experimental repeats. A Post Hoc test using the Bonferroni correction method ($\alpha = 0.01$) was performed to identify treatment groups that were significantly different. When the difference between means of two treatment groups was larger than the computed Least Significant Difference (LSD), the treatments were considered significantly different. The software assigned a unique letter (a, b, c, d, e…) to the treatment group that was significantly different from all other treatments. When the difference between

means of two treatment groups was less than the LSD, they were assigned the same letter indicating no significant difference.

2.4.9 ChIP-qPCR

HEK293T cells, grown in 100 mm TC dishes (Sigma; CLS430167) were transfected with 8 µg of the B:*luc2*:Z3M2:β-globin reporter construct and 16 µg of V5-ZIC3-wt or V5-ZIC3-C365S. 6 h post-transfection cells were dissociated from the growth surface using 0.5 g/L trypsin and plated in 150 mm TC dishes (Iwaki; 3030-150). 24 hr post-transfection media was removed and protein-DNA complexes in cells cross-linked with 1.25% formaldehyde (w/v) (Sigma; F8775) at room temperature for 10 min. Cross-linking was terminated by the addition of 125 mM Glycine (Amresco; Cat. No. 0167) and cells washed thrice with ice-cold 1x PBS. Cells were scraped in ice-cold 1x PBS containing 0.02% Tween20 (Sigma; P7949) and pelleted via centrifugation. The pellet was resuspended in 1.8 mL of sonication buffer [50 mM Hepes (pH 7.9), 140 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS and protease inhibitors (AEBSF and PMSF)]. Cells were sonicated to obtain an average chromatin length of 500 bp using the Bioruptor® (Diagenode) at 4°C. To separate cellular debris sonicated samples were spun down at 18,000 g for 5 min at 4°C.

Chromatin was pre-cleared with a 1:1 mixture of protein A (Novex™; 10001D) and protein G (Novex™; 10003D) Dynabeads® for 2 h at 4°C on a rotating platform. Pre-cleared chromatin was incubated with 7 µg of α-V5 antibody (Abcam; ab9116) in sonication buffer (with protease inhibitors) overnight at 4°C on a rotating platform. Following day, beads were washed at 4°C once with ice-cold sonication buffer (with protease inhibitors) and thrice with sonication buffer containing 500 mM NaCl (with protease inhibitors), followed by one wash with ice-cold Lithium Chloride (LiCl) buffer (250 mM LiCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.5% NP-40 and 0.5% Na-deoxycholate) and two washes with TE (10 mM Tris-HCl (pH 8.0), 1 mM EDTA). Beads were resuspended in 91.5 µL of TE and treated with 0.5 µg of RNAse A (Thermo Scientific™; Cat. No. EN0531) at 37°C for 30 min. Cross-links were reversed by adding 5 µL of 10% SDS and 50 µg of

Proteinase K (Thermo Scientific™; Cat. No. EO0491) and incubation at 65°C for 5 h. DNA was extracted using the AMPure purification system (according to manufacturer's protocol).

qPCR was performed using 1 µL of a 1:10 dilution of input (10%) or ChIP-enriched DNA, ImmoMix™ (Bioline; Cat. No. BIO-25020), 0.5 µM Ark1566_F and Ark1567_R primers (Table 2.8), and SYBR® Green dye in a 10 µL volume. PCR reactions were performed in triplicates, and run using the StepOne Real-time PCR machine (Applied Biosystems™) with the following PCR cycling conditions: 95°C for 5 mins, followed by 40 cycles of 95°C for 15 secs, 55°C for 20 secs and 72°C for 30 secs and a melt curve stage that included 95°C for 15 secs, 55°C for 1 min and 0.3°C increase in temperature every 15 secs until it reached 95°C. Data was analysed using the StepOne Software (v2.3). The amount of target-specific DNA precipitated was determined relative to the amount of non-immunoprecipitated (input) DNA, using the percent input method as outlined on: https://www.thermofisher.com/au/en/home/life-science/epigenetics-noncoding-rna-research/chromatin-remodeling/chromatin-immunoprecipitation-chip/chip-analysis.html.

CHAPTER 3

THE KATUN MOUSE STRAIN IS A NULL ALLELE OF *ZIC3*

3.1 INTRODUCTION

3.1.1 *Zic3* mutant mice model Heterotaxy

Mutations of the ZIC3 transcription factor are associated with a congenital defect, called X-linked Heterotaxy (Gebbia *et al.*, 1997, Sutherland *et al.*, 2009). Practical and ethical constraints (imposed by the in utero development of the mammalian embryo) on the study of human embryogenesis has led to the mouse becoming the mammalian model of choice for studying congenital defects (Bedell *et al.*, 1997). Many examples exist wherein a mutation of the orthologous mouse gene recapitulates the human phenotype. This is partly due to significant sequence conservation between the human and mouse orthologues. For example, the human *ZIC3* gene is 1404 bp long with a protein product (ZIC3) of 467 amino acids, while the mouse *Zic3* gene is 1401 bp long with a protein product of 466 amino acids (Figure 3.1). The amino acid sequences between the human and mouse proteins are 98% identical, with 99.4% homology between their ZFDs.

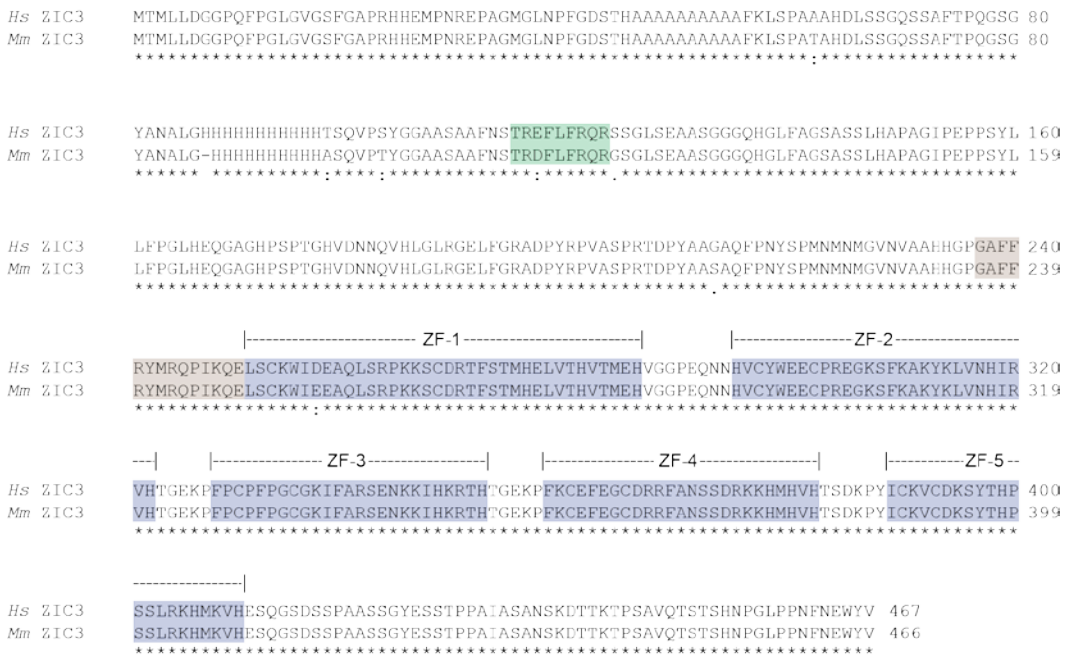


Figure 3.1: Amino acid sequence alignment of human and mouse ZIC3 proteins. Green and brown boxes show the evolutionary conserved ZOC and ZFNC domains, respectively. Blue boxes highlight the five C_2H_2 zinc fingers (ZF). Asterick (*) = identical residues. Colon (:) = functionally conserved residues. Period (.) = weak conservation of residue. *Hs* = *Homo sapiens*; *Mm* = *Mus musculus.*

Mutation of murine *Zic3* also leads to heterotaxy, as revealed by the study of several deletion alleles (Carrel *et al.*, 2000, Purandare *et al.*, 2002). Deletions show gene function but do not reveal fine scale structural information about protein function. A mouse mutant called katun (*Ka*) has been identified (Bogani *et al.*, 2004), which displays phenotypes similar to those seen in *Zic3* null mice and human patients of heterotaxy. Additionally the katun phenotype displays incomplete penetrance and variable expressivity, as is the case for several ZIC3 mutants in humans (Megarbane *et al.*, 2000).

3.1.2 The katun mutation

The kinked tail phenotype (akin to a roller coaster) of carriers formed the basis of naming the mutation katun. Subsequent analysis revealed the mutation occurred in *Zic3* as a result of a spontaneous point mutation prior to ZF1 in *Zic3* (Figure 3.2A-D), which changes the amino acid glutamic acid (E) to a stop codon (X). Due to installation of a premature termination codon (PTC), the katun transcript, if translated, would produce a truncated protein containing only the first 249 amino acids (Figure 3.2E). The mutation has been maintained via a mouse colony and the phenotype and mutation have been found to co-exist in more than 1000 meioses (Ahmed *et al.*, 2013). Additionally other strains of mice (Mus Castaneus, Mus Spretus, C57BL/6J, 129Sv and 101/H) were analyzed and none were found to contain the variant allele, eliminating the possibility that the katun mutation is a naturally occurring polymorphism. Phenotypic comparisons between *Zic3* null mice (Ware *et al.*, 2006c) and katun mutant mice (Ahmed *et al.*, 2013), by analyzing expression of markers for endoderm and mesoderm formation, suggests that the katun allele could behave as a null allele (Figure 3.2F-N). There is, however, no molecular evidence to support this observation. Hence the katun mouse could serve as a useful tool to study the etiology of *ZIC3*-associated heterotaxy, since it will allow determining the molecular causes of the disorder in the presence of a *Zic3* variant (as opposed to its complete absence).

In order to use the katun mutant for this purpose, the type of allele produced as a result of the katun mutation must first be characterized. The katun transcript, in theory, should be degraded via the NMD machinery since the PTC is located more than 50-55 nucleotides upstream of an exon-exon junction (Figure 3.2A) (Nagy *et al.*, 1998). Furthermore the protein product of the katun allele contains only the N-terminal half (that contains some evolutionary conserved sequences) of ZIC3 and lacks the DNA-binding ZFD (that also contains nuclear localization signals) and the remainder of the C-terminus (Figure 3.2E).

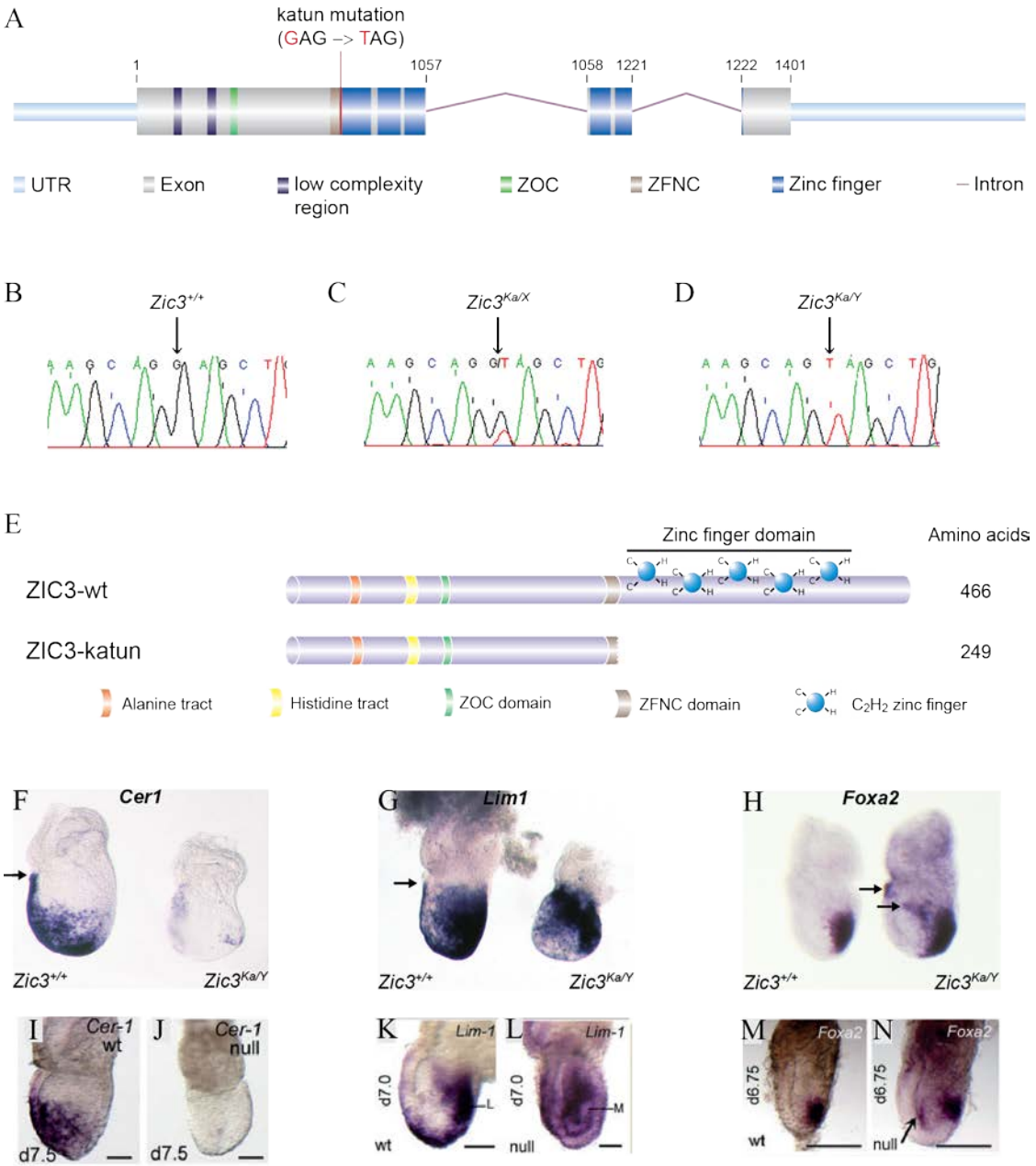


Figure 3.2: The katun mutation. (A) Mutation occurs at nucleotide 745 of *Zic3* CDS (NM_009575.2) converting guanine (G) to thymine (T), changing the amino acid Glutamic acid (GAG) to a stop codon (TAG). Numbers show lengths of the three exons in *Zic3* CDS. **(B-D)** Sequence traces from exon 1 of murine *Zic3* gene of **(B)** Wild-type (*Zic3*$^{+/+}$), **(C)** heterozygous female (*Zic3*$^{Ka/X}$), and **(D)** hemizygous male (*Zic3*$^{Ka/Y}$), with arrows indicating the altered base. **(E)** Protein structure of ZIC3-wt and ZIC3-katun. The katun protein lacks the entire ZFD and C-terminus. **(F-N)** Stage-matched *Zic3*$^{Ka/Y}$ **(F-H)** and *Zic3* null **(I-N)** embryos showing altered primitive streak, mesoderm and endoderm formation. Lateral view of embryos is shown, with anterior to the left, following WMISH to the RNA named on each panel. **(F-L)** *Cer1* and *Lim1* are markers for endoderm and mesoderm formation. In *Zic3*$^{+/+}$ embryos these markers are expressed in the anterior visceral endoderm (black arrow). In *Zic3*$^{Ka/Y}$ and *Zic3* null embryos *Cer1* expression is lacking, while *Lim1* expression is reduced. **(H, M, N)** *Foxa2* shows primitive streak formation. *Zic3*$^{Ka/Y}$ and *Zic3* null embryos display ectopic expression of *Foxa2* going into the amniotic cavity (black arrows). Panels **(B-D)** and **(F-L)** were taken from Ahmed *et al.* (2013). Panels **(I-N)** were taken from Ware *et al.* (2006c).

3.1.3 Aims

To determine the molecular basis of the katun phenotype, two questions were asked: (i) is the katun phenotype seen in mouse embryos a result of the degradation of the katun transcript during gastrulation, and (ii) if the katun protein is expressed, how is the protein function altered? To address these questions the following objectives were set out:

- design a genotyping assay that accurately and reliably distinguishes between wild-type *Zic3* and katun alleles
- determine if the katun transcript is present in embryos during L-R axis development
- verify if the katun protein has stable expression
- examine the subcellular localization of the katun protein
- determine if the katun protein can activate transcription
- investigate if the katun protein can act as a dominant-negative

3.2 RESULTS

3.2.1 Genotyping assay for identifying katun mice

To facilitate the analysis of the katun mouse strain, a reliable and robust genotyping assay capable of detecting a single nucleotide variant (SNV) was required. Two new genotyping assays were developed using Allelic Discrimination and High Resolution Melt Analysis (HRMA).

3.2.1.1 Using Allelic Discrimination

The allelic discrimination assay is a popular PCR-based method used to analyse SNVs (McGuigan *et al.*, 2002, Chen *et al.*, 2003). Reactions are performed using primers that amplify the region of interest and two different fluorescent probes: one that specifically binds the wild-type allele (VIC) and another that binds the mutant allele (FAM). Samples are then identified by measuring the change in fluorescence associated with each probe.

This assay allowed identification of three distinct populations in the katun mouse colony: wild-type (*Zic3*$^{+/+}$), katun heterozygotes (*Zic3*$^{Ka/X}$) and katun hemizygotes (*Zic3*$^{Ka/Y}$) (Figure 3.3A-C). The VIC probe only fluoresced in the presence of the wild-type allele, whereas no fluorescence was recorded from the FAM probe. When katun heterozygous DNA was present, both probes fluoresced at reasonably equal levels. When katun hemizygous DNA was present, only the FAM probe fluoresced while no fluorescence was recorded for the VIC probe. The allelic discrimination plot (Figure 3.3D) shows samples divided into three groups based on the fluorescence being emitted from each probe. Subsequently this assay was routinely used in the laboratory to genotype katun animals and embryos (Ahmed *et al.*, 2013).

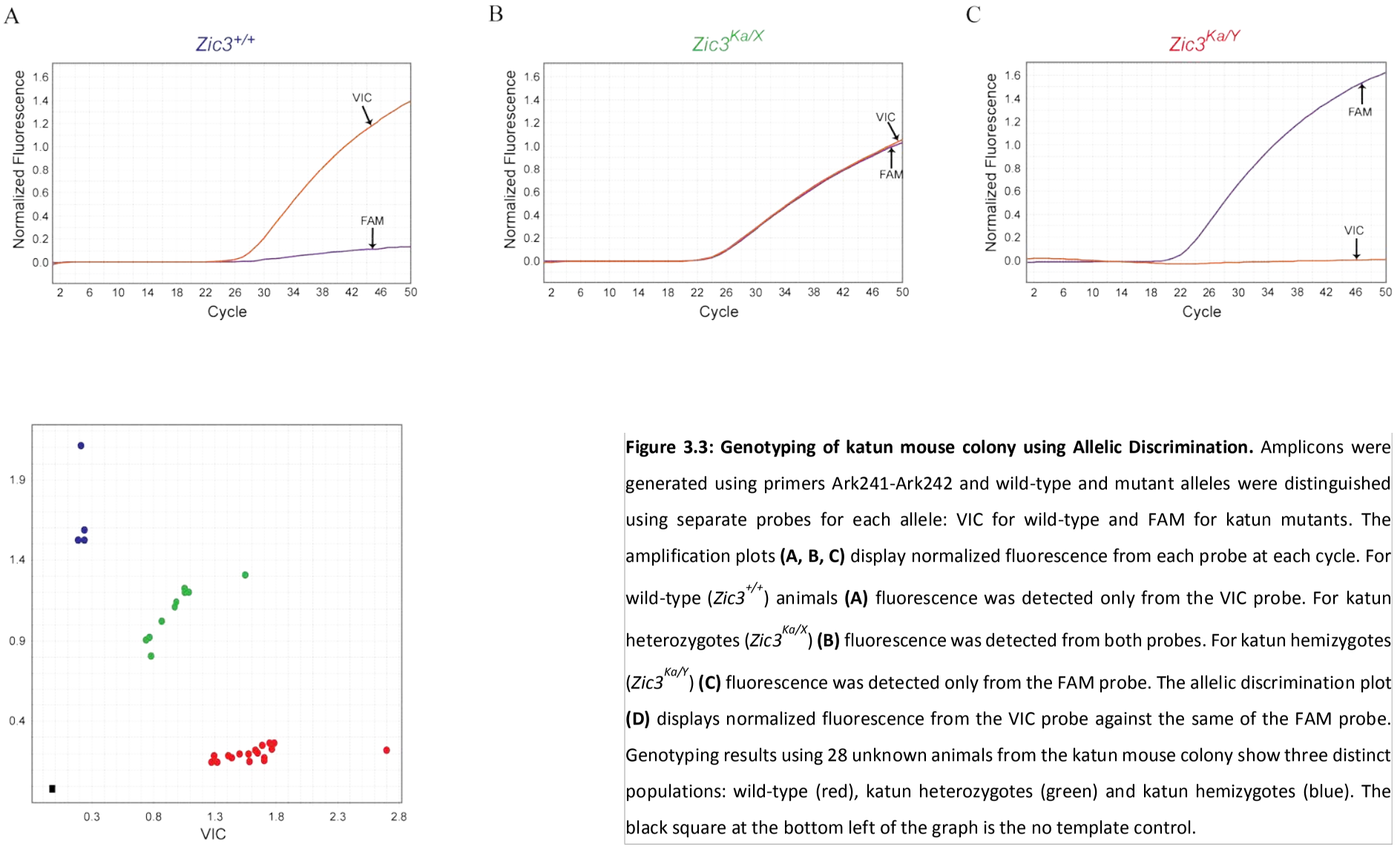


Figure 3.3: Genotyping of katun mouse colony using Allelic Discrimination. Amplicons were generated using primers Ark241-Ark242 and wild-type and mutant alleles were distinguished using separate probes for each allele: VIC for wild-type and FAM for katun mutants. The amplification plots **(A, B, C)** display normalized fluorescence from each probe at each cycle. For wild-type (*Zic3*$^{+/+}$) animals **(A)** fluorescence was detected only from the VIC probe. For katun heterozygotes (*Zic3*$^{Ka/X}$) **(B)** fluorescence was detected from both probes. For katun hemizygotes (*Zic3*$^{Ka/Y}$) **(C)** fluorescence was detected only from the FAM probe. The allelic discrimination plot **(D)** displays normalized fluorescence from the VIC probe against the same of the FAM probe. Genotyping results using 28 unknown animals from the katun mouse colony show three distinct populations: wild-type (red), katun heterozygotes (green) and katun hemizygotes (blue). The black square at the bottom left of the graph is the no template control.

3.2.1.2 Using HRMA

HRMA distinguishes between PCR-amplified DNA fragments based on their melting kinetics by using a DNA-intercalating dye (Wittwer *et al.*, 2003). The first step in designing the assay was identifying a PCR reaction that amplified the region of interest (containing the site of katun mutation) and created amplicons that could reliably be distinguished (see Appendix 2). When optimal reaction conditions were determined, different primer sets were trialled to distinguish between wild-type animals, heterozygous katun females (*Ka/X*) and hemizygous males (*Ka/Y*). Melting curves from each PCR were analysed using two different modes of the LightScanner software: "Expert Scanning" mode that displayed shifted melting curves (bottoms of the curves are superimposed by shifting along the temperature axis to the point where all double-stranded DNA is completely denatured – compensates for well-to-well variations in temperature measurements between samples) (Reed *et al.*, 2007), and "Unlabeled Probe Genotyping" mode that presented data as normalized melting peaks (derivative curves showing the negative derivative of the rate of change of fluorescence (-dF) with respect to changes in temperature (dT) – melt peaks display the approximate melting temperature) (Lay *et al.*, 1997).

Amplicons from primer set 1 were only distinguished as being either wild-type or katun DNA, as shown in the shifted melting curve (Figure 3.4A). The software was unable to differentiate in the melt profile of heterozygous and hemizygous DNA (Figure 3.4A`). Primer set 2 amplicons were only distinguished as being either wild-type or katun DNA (Figure 3.4B). While the shifted melting curves of heterozygous and hemizygous DNA appeared reasonably similar, their normalized melting peaks looked distinct; nonetheless the software was unable to class them as separate samples (Figure 3.4B`). Primer set 3 amplicons had a greater melting temperature range (77-82°C) and based on their shifted melting curves, they were also classified as either being wild-type or katun DNA (Figure 3.4C). The normalized melting peaks of these amplicons clearly displayed three distinct melting profiles with one peak each for wild-type (~81.3°C) and hemizygous DNA (~80.9°C) and two peaks for heterozygous DNA (~79.2°C and ~81°C) (Figure 3.4C`). Hence primer set 4 was chosen to perform all future genotyping of katun animals.

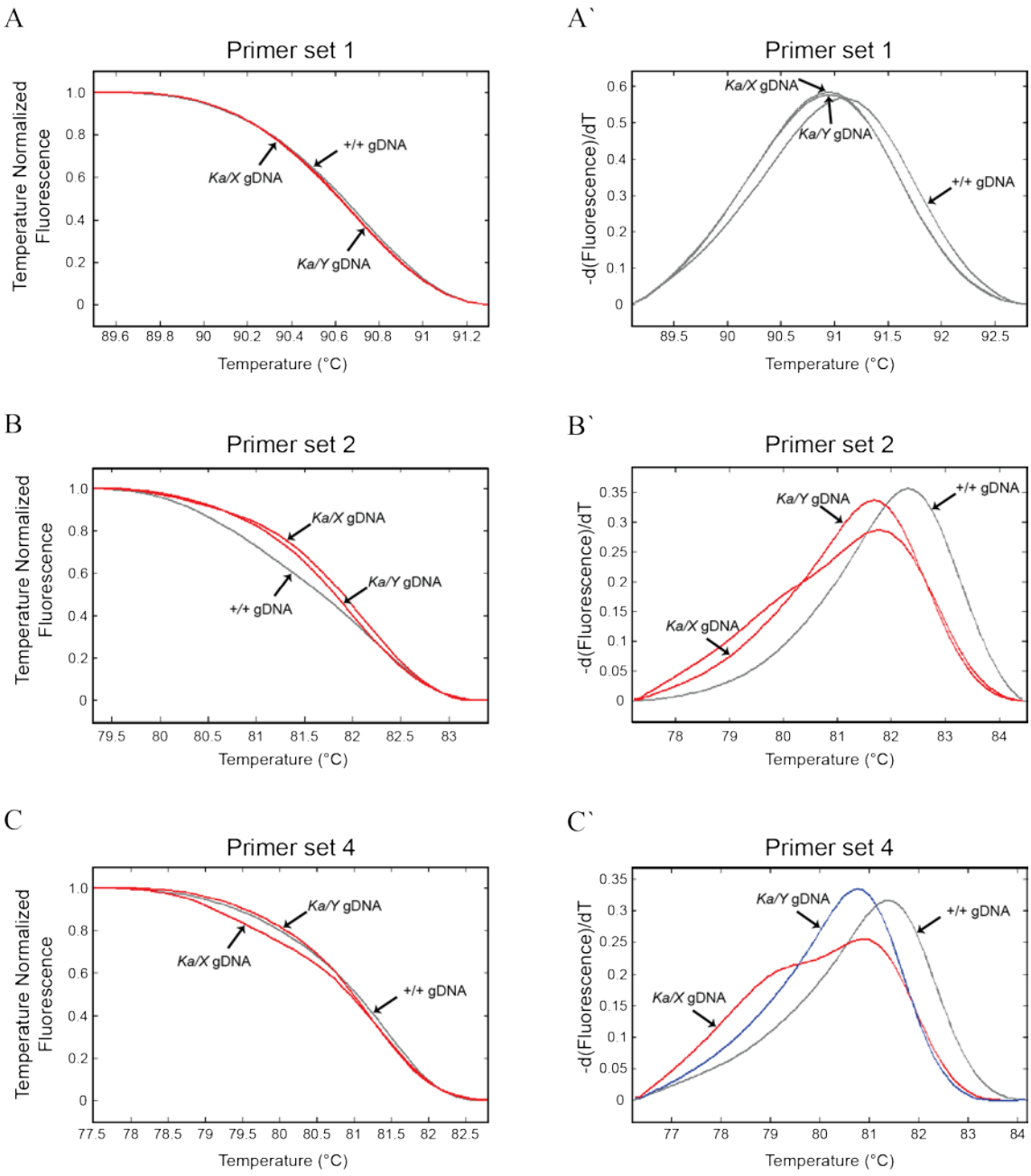


Figure 3.4: Identifying an HRMA assay that can distinguish between all genotypes in the katun mouse colony. Melting profiles of samples were analysed with the LightScanner software using the "Expert Scanning" mode **(A, B, C)** and the "Unlabeled Probe Genotyping" mode **(A`, B`, C`)**. PCRs were performed using genomic DNA (gDNA) from wild-type (+/+), heterozygous (*Ka/X*) and hemizygous (*Ka/Y*) carriers. Each panel represents an assay done using a particular set of primers: **(A and A`)** Primer set 1 = Ark209-Ark242; **(B and B`)** Primer set 2 = Ark241-Ark242; **(C and C`)** Primer set 4 = Ark1085-Ark1086. The colour of the melting curves in each graph were generated automatically by the LightScanner software, based on its ability to distinguish between each genotype.

For routine genotyping (colony maintenance) HRMA was used, since all genotypes within the colony could be identified. However this assay was unable to distinguish between homozygote and hemizygote embryos, thus for embryo genotyping the allelic discrimination assay was used. In addition to a genotyping assay, a sextyping was also developed to correctly identify male and female animals. This assay used a new set of primers (Primer set 5: Ark1002-Ark1003) that bound the *Ube1* allele on both the X and Y chromosomes (Thomsen *et al.*, 2012). Since the allele length varies on both chromosomes, amplicons from the X-chromosome were 132 bp while those from the Y-chromosome were 136 bp (Figure 3.4A). Hence males were expected to show two amplicons while females only one. However for males only the 136 bp was visible on the gel. When the assay was performed using HRMA, two distinct melt profiles were observed (Figure 3.4B). Amplicons from the female DNA displayed a single melting peak at ~81.8°C, while male DNA displayed two peaks at ~81.8°C and ~83.3°C. Sextyping via HRMA proved reliable and robust and all future assays in the laboratory were done using this method.

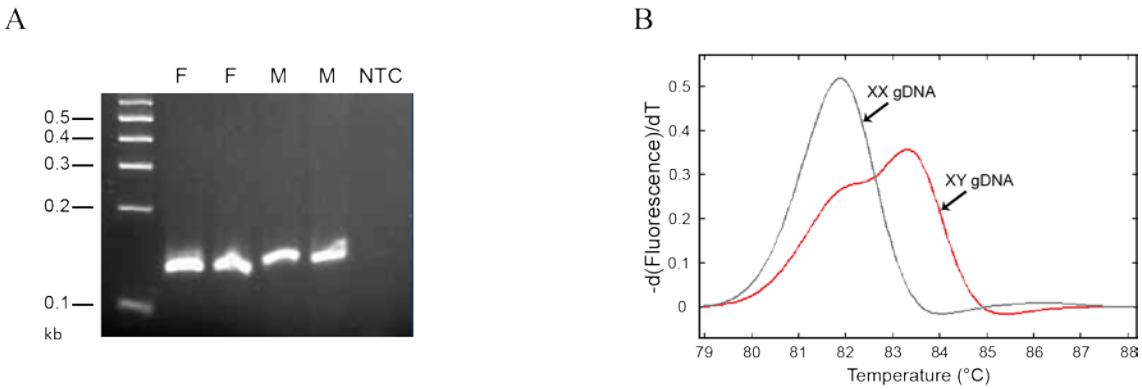


Figure 3.5: HRMA increases reliability of sextyping assay. Genomic DNA (gDNA) extracted from two female (XX) and male (XY) adult mice was used to compare sextyping via agarose gel electrophoresis **(A)** and HRMA **(B)**. Both reactions used primer set 5 (Ark1002-Ark1003) that gave amplicons of 132 bp from the X-chromosome and 136 bp from the Y-chromosome (F, female; M, male; NTC, no template control).

3.2.2 The katun transcript evades NMD

The katun transcript contains a PTC ~313 bases upstream of the exon 1 – exon 2 junction and is expected to be targeted by the NMD machinery (Nagy *et al.*, 1998). Since hemizygous (*Zic3*$^{Ka/Y}$) animals possess only one copy of *Zic3*, which has the katun mutation, katun transcripts should be absent in these animals. To determine whether the katun transcript is subjected to NMD, allele specific RT-PCR was performed using RNA extracted from embryos of each genotype (wild-type, katun heterozygote and katun hemizygote) at 7.5, 8.5 and 9.5 dpc. The absence of genomic DNA in each sample was confirmed by amplification of an intron spanning fragment of the *Zic3* gene (Figure 3.6A-C). The primers correspond to exon 2 (forward primer: Ark364) and exon 3 (reverse primer: Ark311) sequences and the expected product size of cDNA amplicons is 162 bp while for genomic amplicons it is 1006 bp (Figure 3.6D). Once the absence of genomic contamination was confirmed from each sample, allele-specific RT-PCR products were produced (using Primer set 4) and analyzed by HRMA (Figure 3.6A`-C`). In each case, the melt profile of products obtained from cDNA samples were compared with positive control profiles obtained from genomic DNA of animals of known genotype. At each embryonic stage, embryos containing the normal *Zic3* allele exclusively express the wild-type transcript, embryos with one mutant and one wild-type allele express a mixture of the two transcripts, whereas embryos containing only the mutant allele exclusively express the katun transcript. These data, in addition to WMISH analysis showing expression of katun mRNA at gastrulation (Figure 3.7), indicate that the katun transcript is not rapidly subjected to NMD during axis formation (the stage at which *Zic3*-associated heterotaxy phenotype emerges). This implies that a mutant protein could be produced and the heterotaxy phenotype could be due to the aberrant function of this protein.

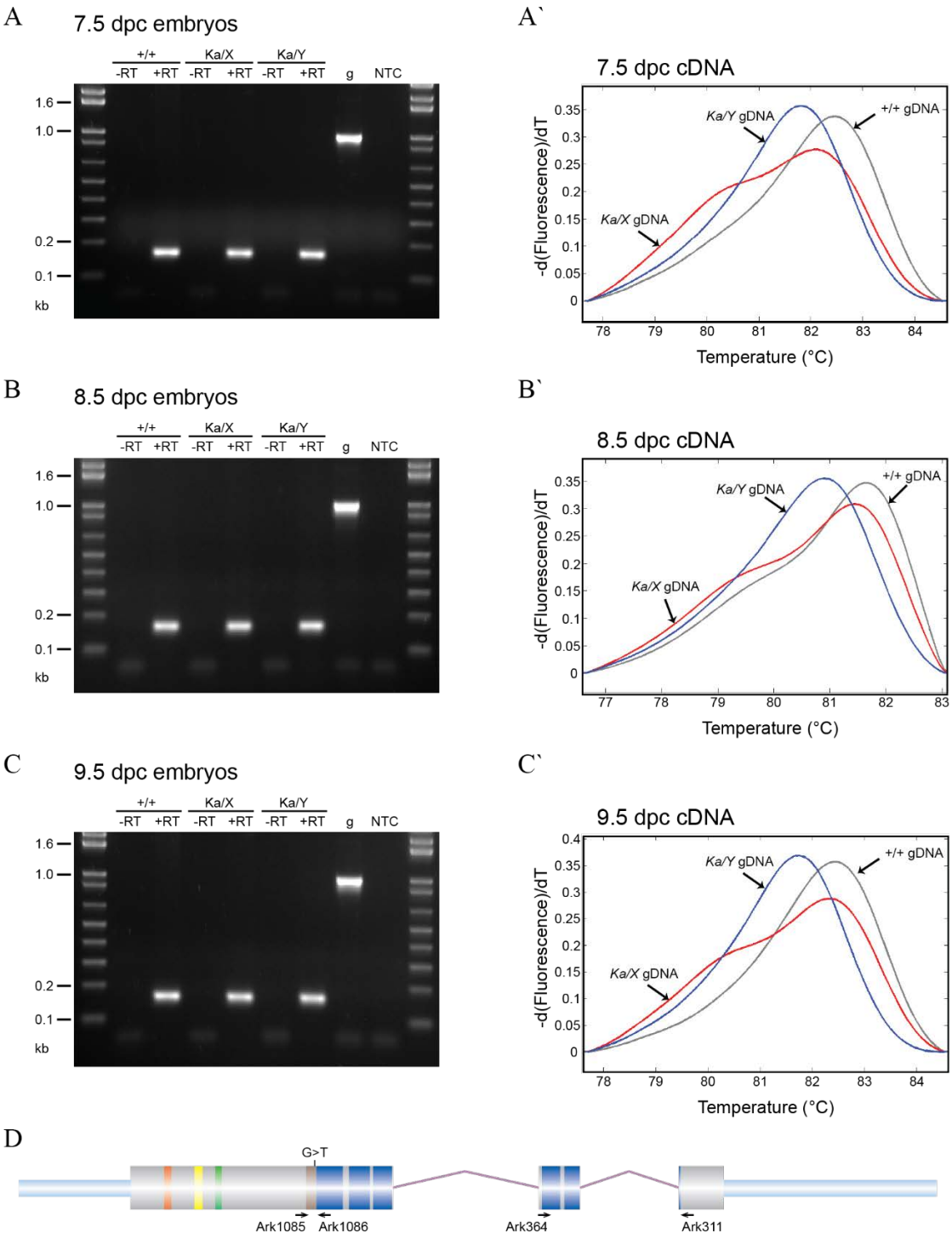


Figure 3.6: The katun mutant transcript is present during gastrulation and organogenesis stage embryos. RNA extracted, from wild-type (+/+), heterozygous (Ka/X) and hemizygous (Ka/Y) embryos at 7.5 dpc **(A)**, 8.5 dpc **(B)** and 9.5 dpc **(C)**, was reverse transcribed followed by PCR amplification of an intron-spanning region (using primers Ark364-Ark311). The expected product size for genomic DNA is 1006 bp and for cDNA is 162 bp (-RT, without reverse transcriptase; +RT, with reverse transcriptase; g, genomic DNA; NTC, no template control). The melting peak plots display DNA melt profiles of Zic3 amplicons after PCR (using primers Ark1085-Ark1086) of cDNA synthesized from 7.5 dpc **(A`)**, 8.5 dpc **(B`)** and 9.5 dpc **(C`)** RNA. **(D)** Diagram of the murine Zic3 genomic locus showing the location of the mutations and the primers used for allele-specific and intron-spanning RT-PCR.

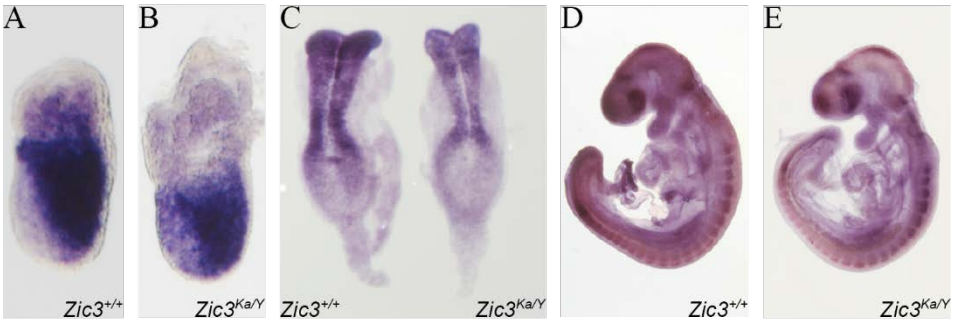


Figure 3.7: The katun transcript accumulates in gastrulation and organogenesis stage mutant embryos. **(A-E)** Lateral views of wild-type (*Zic3*$^{+/+}$) or katun hemizygous (*Zic3*$^{Ka/Y}$) embryos after WMISH to *Zic3* RNA; embryos are shown with anterior to the left **(A-B, D-E)** or top **(C)**. **(A-B)** 7.0 dpc embryos. **(C)** 8.5 dpc embryos. **(D-E)** 9.5 dpc embryos. Pictures taken from Ahmed *et al.* (2013).

3.2.3 Stable expression of katun protein in mammalian cell lines

The finding that the katun transcript escapes NMD during gastrulation raises the possibility that it is translated into a short-lived, truncated protein with some function. Detection of the ZIC3 protein from wild-type and mutant embryos using SDS-PAGE and Western blotting was not successful due to failure of the antibodies to specifically detect endogenous ZIC3 (data not shown). Therefore to determine whether the mutant *ZIC3* transcript can be translated into a stable protein, the katun mutation was introduced in human cDNA using site-directed mutagenesis (Figure 3.8A). HEK293T cells were transfected with either a N-terminal V5-epitope tagged version of ZIC3 (V5-ZIC3-wt) or ZIC3-katun variant (V5-ZIC3-katun). The predicted sizes of the wild-type and katun ZIC3 proteins are 52 kDa and 27 kDa, respectively, with an additional ~4 kDa corresponding to the V5 tag and spacer fragment within the V5-DEST expression vector. Both proteins (V5-ZIC3-wt of ~55 kDa and V5-ZIC3-katun of ~30 kDa) were detected in lysates made 24, 42 and 72 hours post-transfection with Western blotting using an α-V5 antibody (Figure 3.8B), indicating that the katun protein is stable.

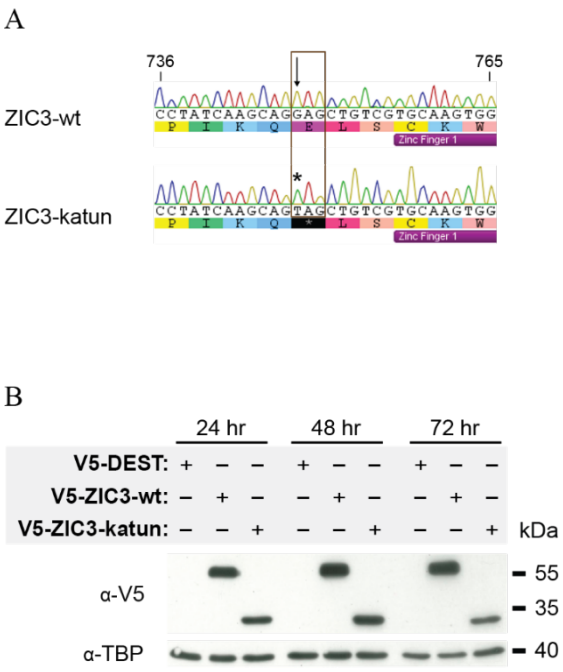


Figure 3.8: Generation of katun mutation on Human *ZIC3* cDNA. Site-directed mutagenesis was carried out to introduce a point mutation on Human *ZIC3* cDNA. **(A)** Sequence read of ZIC3-wt and ZIC3-katun from nucleotide positions 736 to 765 on the Human *ZIC3* cDNA (NM_003413.3), showing conversion of guanine (G) (black arrow) to thymine (T) (asterick) at position 748. This conversion causes creation of a premature stop codon prior to ZF1. Nucleotides are in black font with a clear background. Amino acids are in black font with coloured backgrounds. **(B)** HEK293T cells were transfected with the expression plasmids shown. Cells were harvested 24, 48 and 72 hours post-transfection and subcellular fractionation performed to obtain the nuclear fraction that was used to detect the overexpressed proteins. The α-V5 blots shows level of overexpressed protein and the α-TBP blot serves as nuclear loading control.

3.2.4 Subcellular localization of the katun protein

ZIC proteins are transcriptional regulators and their primary site of function is the nucleus. Previous studies have shown that the ZFD contains a cryptic NLS, required for localization of ZIC proteins to the nucleus (Bedard *et al.*, 2007, Hatayama *et al.*, 2008). Mutations in this domain cause aberrant subcellular distribution when over-expressed in mammalian cell lines (Ware *et al.*, 2004). The katun protein lacks the entire ZFD, hence it should be unable to localize in the nucleus. Following transfection of HEK293T cells with V5-ZIC3-wt or V5-ZIC3-katun, subcellular fractionation of lysates revealed the wild-type protein mostly localized in the nucleus whereas significant proportions of the katun protein were seen in both the nuclear and cytoplasmic compartments (Figure 3.9A). To confirm this result and quantify the extent of katun nuclear accumulation, the subcellular localization of V5-ZIC3-katun was compared with that of V5-ZIC3-wt by immunofluorescent staining following transfection into HEK293T cells (Figure 3.9C). The

transfected protein was detected using an α-V5 antibody, while the nuclear boundary was marked using the α-LaminB1 antibody. Consistent with previous studies on subcellular location of ZIC proteins, 88.5% of the wild-type (V5-ZIC3-wt) protein was found to be within the nucleus. Surprisingly, 61.4% of the mutant (V5-ZIC3-katun) protein accumulated within the nucleus (Figure 3.9E).

This raised the possibility that the N-terminal portion of ZIC3 contains sequences sufficient for directed nuclear transport. It is, however, also possible that the small size of the katun protein allows it to diffuse into the nucleus since proteins smaller than ~40 kDa have been shown to do so (Wei *et al.*, 2003). To distinguish between these possibilities, the size of the katun protein was artificially increased by fusion with enhanced green fluorescent protein (EGFP). The subcellular localization of EGFP-ZIC3-katun was compared with that of EGFP-ZIC3-wt following transfection into HEK293T cells using Western blotting and immunofluorescent staining (Figure 3.9B and D). Western blot analysis detected EGFP-ZIC3-wt in the nuclear fraction at ~85 kDa and EGFP-katun predominantly in the cytoplasmic fraction at ~55 kDa. Immunofluorescent subcellular localization analysis showed that 88.9% of the EGFP-ZIC3-wt protein was within the nucleus, whereas only 10.3% of the mutant EGFP-ZIC3-katun protein accumulated within the nucleus (Figure 3.9E). The difference between the localization patterns of V5-ZIC3-katun protein and the EGFP-ZIC3-katun protein suggests that the katun protein accumulates in the nucleus by passive diffusion.

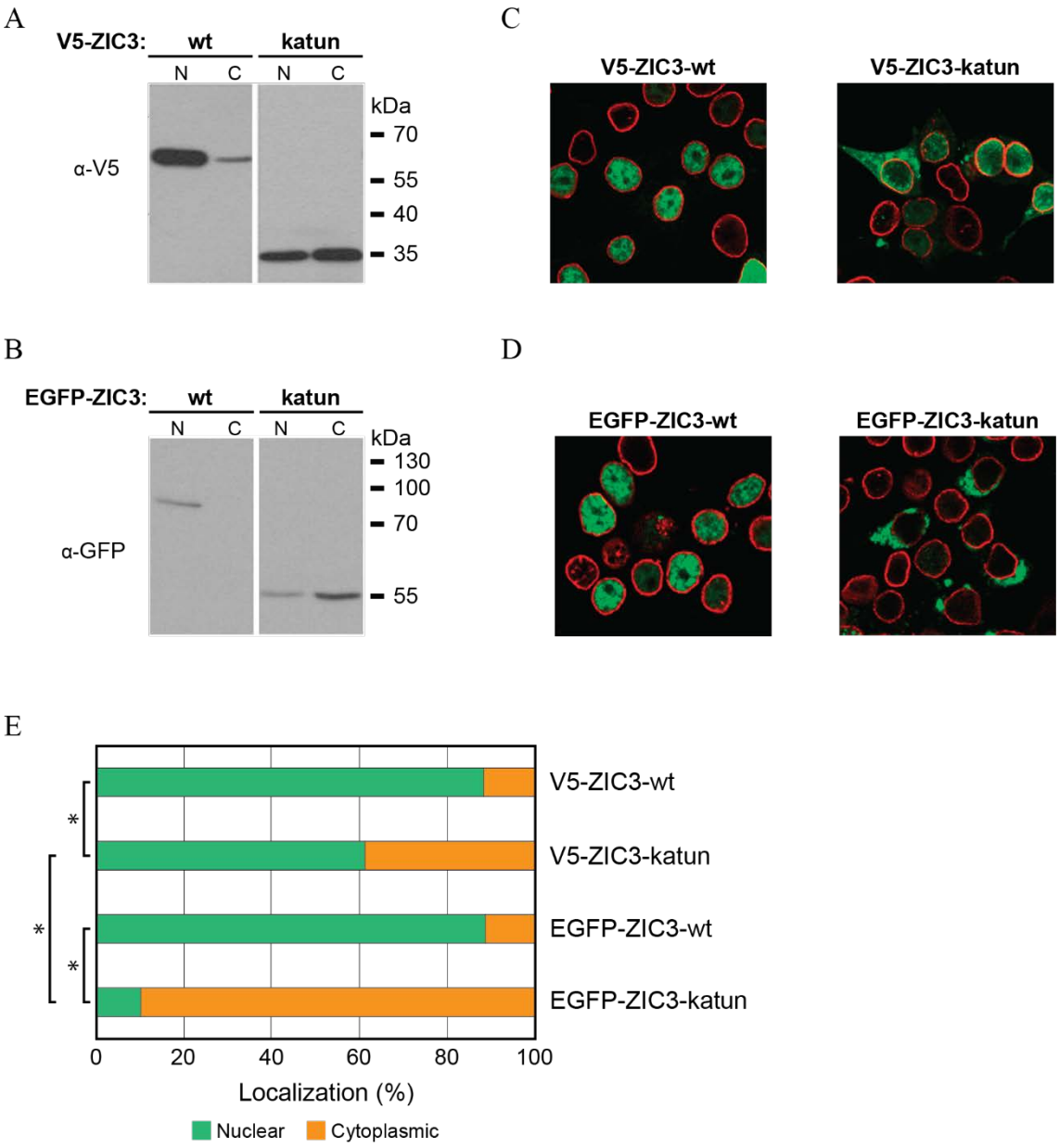


Figure 3.9: Truncated katun protein diffuses into the nucleus. (A, B) HEK293T cells were transfected with the expression plasmids shown and lysed to generate nuclear (N) and cytoplasmic (C) fractions. Lysates were subjected to SDS-PAGE and Western blotting with **(A)** α-V5 or **(B)** α-GFP antibodies. **(C, D)** HEK293T cells transfected with the expression plasmids shown were co-immunostained with α-LaminB1 antibody (to detect nuclear envelope, in red) and either **(C)** α-V5 or **(D)** α-GFP antibodies (to detect overexpressed ZIC3 proteins, in green). Overlaid images are shown here. **(E)** Subcellular localization of V5- or EGFP-tagged wild-type ZIC3 or katun proteins was quantified using ImageJ. Data presented is combined from 3 external repeats, with a 100 cells counted for each construct per experiment. *: $p < 0.01$ ANOVA.

3.2.5 The katun protein is a functional null

3.2.5.1 Katun cannot transactivate

The ZFD is crucial for ZIC protein function, since it is required for DNA-binding, nuclear localization and interacting with protein partner interaction (Matsugi *et al.*, 1990, Gamsjaeger *et al.*, 2007). In addition, mutations in any of the ZFs causes diseased phenotypes in humans (Chhin *et al.*, 2007, D'Alessandro *et al.*, 2013). Ware *et al.* (2004) showed that nonsense, missense or frameshift mutations in the ZFD of ZIC3 prohibit the mutant protein from activating transcription in cell-based reporter assays. Since the katun protein lacks the entire ZFD (Figure 3.2E), it should not be able to bind target DNA sequences and influence transcription. However the katun protein is stable and does localize to the nucleus thus making it possible for the protein to exert some effect.

The ability of V5-ZIC3-katun to stimulate transcription was evaluated using a well-established cell-based luciferase reporter assay, involving the *APOE* promoter (Salero *et al.*, 2001). ZIC proteins have been shown to bind the *APOE* promoter and activate transcription, while DNA-binding ZIC mutants are unable to illicit transcription from this element (Brown *et al.*, 2005). Co-transfection of the *APOE* reporter construct (B:*luc*$^{+}$:APOE) with either V5-ZIC3-wt or V5-ZIC3-katun into HEK293T cells, followed by quantification of luciferase activity demonstrated that wild-type ZIC3 significantly increased relative luciferase activity (RLA) while the katun protein was unable to elicit a transcriptional response (Figure 3.10A).

3.2.5.2 Katun cannot behave in dominant-negative manner

The katun protein lacks the ZFD, but possesses a complete N-terminus (amino acids 1-249) region that contains some evolutionary conserved sequences (ZOC box, alanine and histidine repeat regions) (Figure 3.2E). Kitaguchi *et al.* (2000) observed developmental defects in *Xenopus* blastomeres after injecting a truncated mRNA expressing the N-terminus (amino acids 1-214) of ZIC3. As shown previously the katun protein is unable to stimulate transcription, however it is possible that it could interfere in the function of other ZIC proteins. During murine gastrulation,

expression of *Zic3* overlaps with that of *Zic2* and *Zic5* in the embryonic ectoderm and mesoderm (Houtmeyers *et al.*, 2013). The katun phenotype may be due, at least in part, to the katun protein inhibiting other ZIC proteins and preventing their normal function. The ability of V5-ZIC3-katun to behave in a dominant-negative manner was evaluated using the *APOE* cell-based luciferase reporter assay. The reporter plasmid was co-transfected in HEK293T cells with either wild-type ZIC3, ZIC2 or ZIC5 and V5-DEST or V5-ZIC3-katun and luciferase expression was measured 24 hour post-transfection (Figure 3.10B). In each of the ZIC proteins tested, presence of the katun protein did not significantly alter the ability of the wild-type protein(s) to stimulate transcription.

3.2.5.3 Katun does not inhibit expression of Wnt target genes

The human ZIC2 and *Xenopus* ZIC1-5 proteins have recently been shown to act as co-factors that inhibit Wnt dependent-β-catenin mediated transcription (Pourebrahim *et al.*, 2011, Fujimi *et al.*, 2012). Upon Wnt stimulation, β-catenin enters the nucleus and interacts with the TCF transcription factors to stimulate transcription of target genes (Behrens *et al.*, 1996). A luciferase reporter construct containing consensus TCF binding sites (TOPflash) or mutated TCF sites (FOPflash) is routinely used to assess Wnt dependent transcription (Korinek *et al.*, 1997). Co-transfection of the TOPflash construct with one encoding a stabilized form of β-catenin (β-catenin-ΔN89) into HEK293T cells drives high levels of luciferase activity but this level is not attained in the presence of V5-ZIC3-wt (Figure 3.10C). This indicates that human ZIC3, like ZIC2 is able to inhibit β-catenin mediated transcription of Wnt target genes in cultured HEK293T cells. In addition when ZIC3 is expressed, lower levels of β-catenin ΔN89 are detected consistent with the enhanced β-catenin degradation previously seen with the expression of *Xenopus* ZIC3 (Fujimi *et al.*, 2012). In contrast to wild-type ZIC3 protein, co-transfection with V5-ZIC3-katun does not decrease luciferase activity indicating that the katun protein is unable to inhibit Wnt-dependent β-catenin mediated transcription.

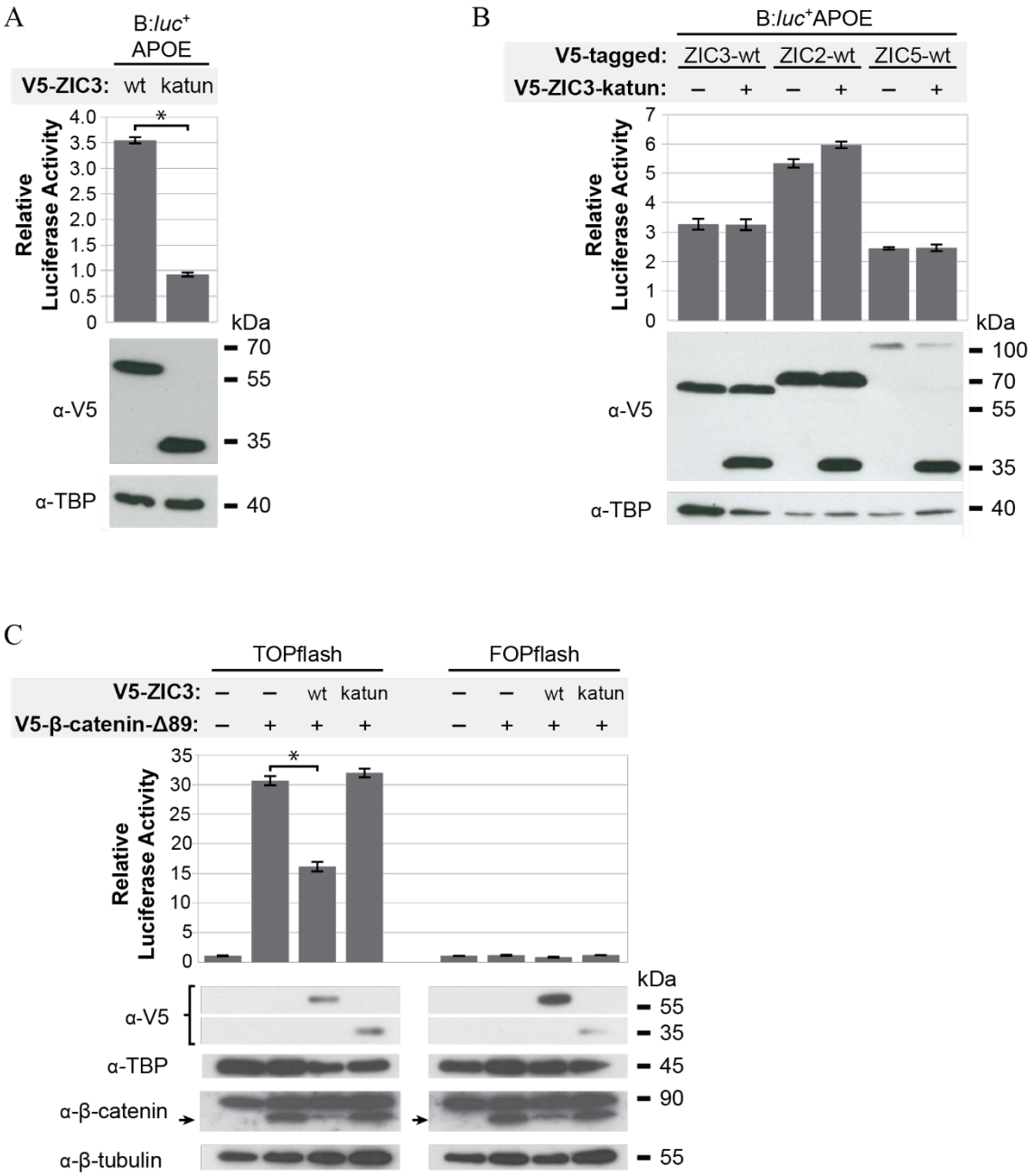


Figure 3.10: The katun protein is functionally inert. HEK293T cells were co-transfected with the reporter and V5-DEST or the expression plasmids shown to check for **(A)** transactivation, **(B)** dominant-negative interference, or **(C)** inhibition of β-catenin mediated transcription via the katun protein. RLA was calculated by dividing luminescence from each expression plasmid by that of V5-DEST. Expression of transfected proteins was confirmed via Western blotting with **(A-C)** α-V5 and **(C)** α-β-catenin in the nuclear and cytoplasmic fractions, respectively. The **(A-C)** α-TBP and **(C)** α-β-tubulin blots served as nuclear and cytoplasmic loading controls, respectively. **(C)** The α-β-catenin antibody detects both endogenous β-catenin and the smaller exogenously expressed β-catenin-Δ89 (marked by arrows). **(A, C)** Error bars represent SEM, N=3. *: $p < 0.01$ ANOVA. **(B)** Error bars represent SD between internal replicates. Experiment was repeated twice to confirm results. Panel **(C)** was taken from Ahmed *et al.* (2013).

3.2.6 PTC-containing *ZIC3* mutant transcripts produce proteins that do not compete with wild-type ZIC3

The magnitude of NMD is dependent upon the strength of the splice donor and acceptor sites within an intron, with stronger sites (i.e. those that more closely match the consensus sequence) subject to increased NMD (Gudikote *et al.*, 2005). To assess whether inefficient NMD of the human *ZIC3* can be predicted, the mouse *Zic3* and human *ZIC3* splice sites were compared. As shown in Table 3.1, the sites are identical (intron 1/2 and 2/3) or nearly so (intron 2/4). Intron 2/3 has the weakest splice signals, as indicated by a lower consensus value score.

Gene	Intron	Donor Site		Acceptor Site	
		Sequence	CV	Sequence	CV
Murine *Zic3* Human *ZIC3*	Intron 1/2	CAGgtaagg CAGgtaagg	98.07	tgccttttgcagGTA tgccttttgcagGTA	92.14
Murine *Zic3* Human *ZIC3*	Intron 2/3	AAGgtaatt AAGgtaatt	86.73	tgttaattttagGTT tgttaattttagGTT	82.17

Table 3.1: *Zic3*/*ZIC3* splice site homology and scores. Splice site consensus values (CV) were calculated using Human Splicing Factor (http://www.umd.be/HSF/4DACTION/input_SSF). Upper case letters: Exon sequence; lower case letters: Intron sequence.

The absolute conservation of *Zic3* and *ZIC3* splice sites prompted the examination of the predicted proteins from each of the *ZIC3*-associated heterotaxy mutations that generate a PTC. Six different *ZIC3*-associated heterotaxy mutations with a PTC have previously been documented and the protein stability, subcellular localization and transcriptional ability of each of these assessed using cell-based assays (Ware *et al.*, 2004). Two of these mutations generate unstable proteins (ZIC3-S43X and ZIC3-Q249X) whereas the remaining mutations generate stable proteins that can either be found exclusively in the cytoplasm (ZIC3-C268X and ZIC3-Q292X) or in both cytoplasmic and nuclear compartments (ZIC3-1507insTT and ZIC3-K408X). To test whether any of these proteins possess dominant-negative properties, the analogous mutant proteins were expressed in HEK293T cells in competition with wild-type ZIC3. Expression of each ZIC3 construct

was verified by SDS-PAGE and Western blotting (Figure 3.11). Each mutant construct was unable to stimulate transcription, except K408X that retained some transactivation ability [the same was noted by Ware *et al.* (2004)]. Furthermore when co-transfected with wild-type ZIC3, the ZIC3-katun, ZIC3-C268X and ZIC3-Q292X proteins did not significantly alter ability of ZIC3 to activate transcription, while ZIC3-1507insTT or ZIC3-K408X showed mild hyper-stimulation of the *APOE* promoter suggesting a synergistic response. Note that the ZIC3-Q249X mutation was excluded from this analysis because it is probably well-modelled by experiments with the V5-ZIC3-katun construct.

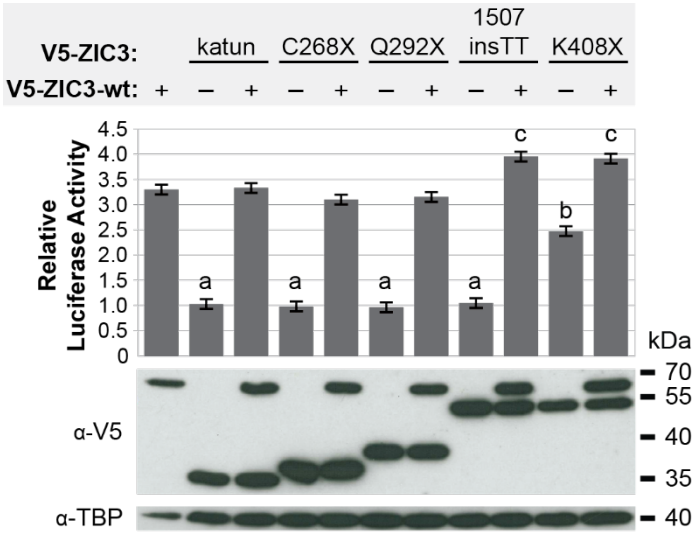


Figure 3.11: ZIC3-heterotaxy associated mutant proteins do not compete with wild-type ZIC3. Competition assay using HEK293T cells co-transfected with the B:*luc*$^{+}$:APOE reporter construct and V5-DEST or the expression plasmids shown, with luminescence quantified 24 hours post-transfection. Graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. a, b, and c: $p < 0.01$ ANOVA. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control.

3.2.7 PTC-containing *ZIC3* mutant transcripts produce proteins that do not compete with wild-type ZIC2 and ZIC5

Given the overlapping expression domains of *ZIC3*, *ZIC2* and *ZIC5*, the ability of each of the PTC-containing ZIC3 mutants to dominantly interfere with ZIC2 and ZIC5 was also tested in HEK293T cells and expression of each ZIC construct verified by SDS-PAGE and Western blotting. As previously shown (Brown *et al.*, 2005), ZIC2 strongly stimulates the *APOE* promoter element (Figure 3.12A). When co-transfected with wild-type ZIC2, ZIC3-katun and ZIC3-1507insTT proteins did not significantly alter ability of ZIC2 to activate transcription, while ZIC3-C268X, ZIC3-Q292X and ZIC3-K408X showed a small, but significant, inhibitory effect. The ability of ZIC5 to transactivate the *APOE* promoter has not previously been tested, but as shown in Figure 3.12B the ZIC5 proteins has a small stimulatory effect on this promoter element. In the presence of V5-ZIC3-katun, V5-ZIC3-C268X, V5-ZIC3-Q292X and V5-ZIC3-1507insTT no significant change was observed in RLA by wild-type ZIC5. Evidently the PTC mutants have differing effects on wild-type ZIC3, ZIC2 and ZIC5.

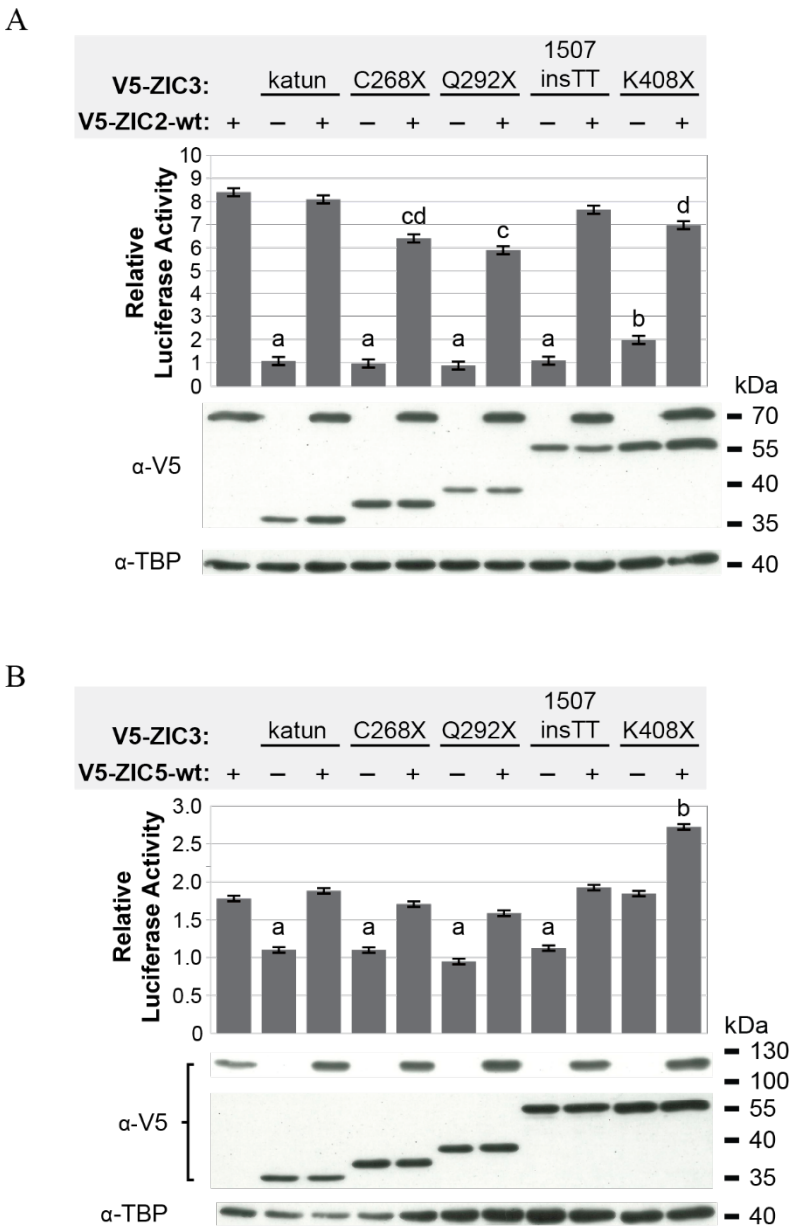


Figure 3.12: Dominant-negative behaviour of ZIC3-heterotaxy associated mutant proteins in the presence of wild-type ZIC2 and ZIC5. Competition assay using HEK293T cells co-transfected with the B:*luc*$^{+}$:APOE reporter construct and V5-DEST or the expression plasmids shown, with luminescence quantified 24 hours post-transfection. Graphs show mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. **(A)** a, b, c and d: $p < 0.01$ ANOVA. **(B)** a and b: $p < 0.01$ ANOVA. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control.

3.3 DISCUSSION

Zic genes are vertebrate homologues of *Drosophila odd-paired* (*opa*) gene, which is involved in the body-plan formation of the *Drosophila* embryo (Aruga *et al.*, 1996). Previous studies in human, mouse, frog and ascidian *Zic* homologues demonstrated that the *Zic* gene family encodes transcription factors that play critical roles in variety of developmental processes such as neurogenesis, myogenesis, skeletal patterning and L-R axis development (Aruga, 2004, Houtmeyers *et al.*, 2013). Nonetheless, little is known about the biochemical and molecular properties of these ZF proteins.

The work presented in this chapter focused on understanding the molecular mechanism(s) underlying the phenotype of the katun mouse mutant. This mutation arose spontaneously during an ENU-mutagenesis experiment (Bogani *et al.*, 2004). Functional assays reported herein demonstrate that katun is a null allele of *Zic3*. Importantly these finding uncovered the molecular mechanism of the katun phenotype and provided new insights into the functional significance of the different structural domains in ZIC3, namely, the N-terminal region and the ZFD.

3.3.1 The katun mouse strain carries a null allele of *Zic3*

The katun mutation introduces a nonsense codon into the *Zic3* transcript that conforms to the rule for PTC recognition in mammalian cells. Allele-specific RT-PCR of 7.5, 8.5 and 9.5 dpc embryos showed that the only transcript expressed in $Zic3^{Ka/Y}$ embryos carries the nonsense mutation, while both wild-type and mutant transcripts are present in $Zic3^{Ka/X}$ embryos (Figure 3.6) (Ahmed *et al.*, 2013). In addition expression of the katun protein was observed up to 72 hours post-transfection in HEK293T cells (Figure 3.8B). Unfortunately a positive control was not available to determine whether NMD is at play in HEK293T cells and that they adhere to the rule of degrading transcripts with a PTC 50-55 nucleotides prior to an exon-exon junction. This test needs to be completed to conclusively ascertain that the katun transcript is a weak target for

NMD. Moreover in vitro expression of the katun protein should be checked in other cell lines for up to 72 hours post-transfection (Festing, 2001). RT-PCR data and WMISH on katun embryos does show that the katun transcript is not degraded (Ahmed *et al.*, 2013), indicating that the katun transcript is present at the time of L-R axis formation and raises the possibility that the katun mutation does not represent a null allele.

Investigations into the molecular properties of the katun mutant protein, however, confirmed that it lacks activities associated with wild-type ZIC3. The katun protein is truncated just upstream of the ZFD, which is required by the wild-type protein for nuclear localization (Bedard *et al.*, 2007, Hatayama *et al.*, 2008) and transactivation of target gene expression (Koyabu *et al.*, 2001, Mizugishi *et al.*, 2001, Ware *et al.*, 2004). Although the katun protein can accumulate in the nucleus to appreciable levels (due to passive diffusion rather than active transport) (Figure 3.9), it is transcriptionally inert (Figure 3.10A) and does not inhibit β-catenin mediated transcription (Figure 3.10C) (Ahmed *et al.*, 2013).

These data establish that the katun protein is null for known ZIC3 molecular activities, but *Xenopus* experiments have implied that a ZIC3 protein similar to katun interferes with the function of wild-type ZIC3 during L-R axis formation (Kitaguchi *et al.*, 2000). In addition other studies have shown that truncated mutants have the potential to display dominant-negative behaviour (Miyamoto *et al.*, 2002, Kiefer *et al.*, 2003). This possibility has not previously been biochemically tested for the katun mutant. When the katun protein is placed in competition with either wild-type ZIC3 or other ZIC proteins co-expressed at the time of L-R axis formation (ZIC2 and ZIC5), it does not interfere with their ability to activate transcription of a target promoter in HEK293T cells (Figure 3.10B). It remains possible that the katun protein displays dominant-negative behaviour against other currently unknown protein activities, however, the mutation phenocopies other deletion alleles of *Zic3* (Purandare *et al.*, 2002, Ware *et al.*, 2006c) indicating that dominant-negative behaviour of the katun protein at gastrulation is inconsequential. Taken together these data demonstrate that *Ka* encodes a null allele of *Zic3* and that the N-terminal portion of mammalian ZIC3 does not encode a dominant-negative molecule (Ahmed *et al.*,

2013).

3.3.2 Incomplete NMD can allow ZIC3 PTC-inducing mutations to have some function

The finding that murine *Zic3* transcript is a poor substrate for NMD at axis formation suggests that this is also the case for the human *ZIC3* transcript. Direct assessment of the fate of human *ZIC3* PTC-containing transcripts during axis formation is not possible and predictions regarding the likely NMD behavior based on the murine transcript provide one alternative (Ahmed *et al.*, 2013). The features that render a transcript as a poor target for NMD are not fully characterized. There is, however, evidence that position of the PTC within the transcript and RNA splicing influences NMD amplitude (Nagy *et al.*, 1998, Gudikote *et al.*, 2005). The genomic arrangement of the murine (*Zic3*) and human (*ZIC3*) genes is nearly identical and the splice donor and acceptor sites are completely conserved. It is likely that the human *ZIC3* transcript is similarly able to avoid mRNA surveillance mechanisms, which needs to be considered when assessing the probable effect of *ZIC3* PTC-inducing mutations (Ahmed *et al.*, 2013). If PTC-containing transcripts are translated, the two most likely effects are: (i) proteins that truncate downstream of crucial domains might be hypomorphic; and (ii) proteins that do not produce crucial domains might encode dominant-negative molecules.

Six mutations that introduce a PTC into the human *ZIC3* transcript have been associated with congenital defects (Gebbia *et al.*, 1997, Megarbane *et al.*, 2000, Ware *et al.*, 2004). Four of these adhere to the position rule for NMD (i.e. the PTC is sited more than about 50-55 nucleotides upstream of an exon-exon junction) (Nagy *et al.*, 1998). The most 5'-ward of the mutations (C633A) encodes a severely truncated molecule (S43X) if not degraded. Previous examination of this putative protein has shown that it is not stably produced in cell lines (Ware *et al.*, 2004) and is therefore likely to generate a null allele regardless of NMD amplitude. The C1250T mutation encodes the ZIC3-Q249X protein. An expression construct incorporating this mutation into the *ZIC3* CDS has previously been reported to produce no protein (Ware *et al.*, 2004). This PTC lies

very close to the site of the katun mutation that produces the ZIC3-E250X protein, however the katun protein shows stable expression. While there could be several reasons for the discrepancy between these results, experiments on the katun protein suggest that even if the ZIC3-Q249X protein is generated in vivo it would encode a protein that is unable to behave in dominant-negative manner (Ahmed *et al.*, 2013). The remaining two mutations (C1338A and C1408T) that conform to the NMD PTC position rule would generate proteins that contain part of the ZFD (ZIC3-C268X and ZIC3-Q292X, respectively). Previous studies indicate that these proteins are transcriptionally inactive, which has also been confirmed by data shown here. In addition these proteins fail to dominantly interfere with the transactivation ability of ZIC3 (Figure 3.11) and ZIC5 (Figure 3.12B). Interestingly these proteins are able to significantly reduce transactivation via ZIC2 (Figure 3.12A). The molecular nature of this needs to be examined further, however it implicates that the mutant phenotype is due to a combination of inactivity of ZIC3 and limited transactivation via ZIC2. The remaining two mutations (1507insTT and A1741T) introduce a PTC close to the last intron and regardless of *ZIC3* transcript sensitivity to NMD, are translated into stable proteins. Data shown here confirms that ZIC3-1507insTT is transcriptionally inert whereas the ZIC3-K408X protein (corresponding to the A1741T mutation) retains some transactivation ability. Both proteins cause an increase in transactivation via ZIC3 when co-expressed with the wild-type protein. In case of ZIC3-K408X the increase is perhaps due to a synergistic effect of both wild-type ZIC3 and the ZIC3-K408X protein on the promoter, however, since ZIC3-1507insTT is transcriptionally inert the increase in transactivation via ZIC3 is likely due to a protein-protein interaction. The ZIC3-1507insTT protein has an intact ZIC3 amino acid sequence until after ZF2, thus perhaps it is able to mobilize the transcriptional machinery which then allows wild-type ZIC3 to simply bind DNA and activate transcription. Nonetheless, in reality, due to the location of *Zic3* on the X-chromosome, *Zic3* mutant alleles cannot act in a dominant-negative manner against wild-type *Zic3* alleles. This is, (i) due to presence of only one copy of the *Zic3* gene in males as they possess only one X-chromosome, and (ii) due to X-inactivation in females that causes silencing of gene expression from one of the X-chromosomes during

embryonic development (Heard *et al.*, 1997). Therefore to conclusively determine the dominant-negative activity of these mutants similar experiments were conducted with wild-type ZIC2 and ZIC5. When the ZIC3-1507insTT protein was co-expressed with wild-type ZIC2 or ZIC5, no change in transactivation was observed. This was surprising, especially in the case of ZIC2 since two other PTC mutants (ZIC3-C268X and ZIC3-Q292X) can limit transactivation via wild-type ZIC2. However the difference may be the presence of the ZF2 in the ZIC3-1507insTT. When ZIC3-K408X was co-expressed with wild-type ZIC2 it also caused a reduction in transactivation via wild-type ZIC2 albeit not as much as that caused by the C268X and Q292X mutants. In conjunction with wild-type ZIC5, ZIC3-K408X caused an increase in transactivation. While further work needs to be done to assess the authenticity of these data, it can be inferred that the mutant phenotype of the ZIC3-K408X variant is due to reduced expression of ZIC3 and ZIC2 target genes and increased expression of wild-type ZIC5 target genes.

CHAPTER 4

THE ZICFLASH ASSAY

4.1 INTRODUCTION

4.1.1 Transactivation assays for ZIC proteins

Transactivation assays monitor the effect of a transcription factor on a response element. They are an invaluable tool for dissecting transcription factor function since they enable rapid and highly sensitive comparisons of variant forms of the response element and/or the transcription factor itself (Rosenthal, 1987, Jiang *et al.*, 2008). Typically these assays rely on the use of a non-mammalian reporter gene, such as luciferase, GFP or *lacZ*, whose expression is regulated by a response element, such as a heterologous promoter and/or enhancer sequence (Houck *et al.*, 2006). The response element is cloned upstream of the reporter gene (Figure 4.1A) and the recombinant plasmid transfected into cultured mammalian cells. The transcriptional response is quantified based on expression of the reporter gene and is representative of the interaction between transcription factor and response element. Since the response element is removed from its native genomic context, such that long-range regulatory elements are mostly missing, only the cloned portion of the promoter or enhancer is assayed (Landolin *et al.*, 2010). Nonetheless such assays are uniquely suited to studying transcription factors since the activation of gene expression is directly linked to their function (Paguio *et al.*, 2010).

Commonly used reporter genes generally express two types of proteins: (i) with intrinsic fluorescence (such as GFP), or (ii) enzymes (such as luciferase or β-galactosidase) (Houck *et al.*, 2006). Hallmark of a good reporter gene is that its protein product is easily assayable and has low endogenous background activity (Ma, 2014). The luciferase gene used in this study meets this criterion and is one of most commonly used reporters for studying transcription factor function (Trinklein *et al.*, 2003, Cooper *et al.*, 2006, Chabot *et al.*, 2007). The luciferase protein, a monomeric enzyme of 61 kDa, was initially cloned from the firefly (*Photinus pyralis*) (de Wet *et al.*, 1985, de Wet *et al.*, 1987). Use of the luciferase protein in transactivation assays is based on the natural phenomenon of bioluminescence, which involves enzyme-catalysed oxidation of substrates resulting in emission of photons (Figure 4.1B) (Bronstein *et al.*, 1994, Fan *et al.*, 2007,

Fraga, 2008, Thorne *et al.*, 2010). In general the intensity of light released as a result of this chemical reaction is correlated to the chemical concentrations of the reaction components, such as the amount of luciferase enzyme produced or the amount of substrate present (Fan *et al.*, 2007).

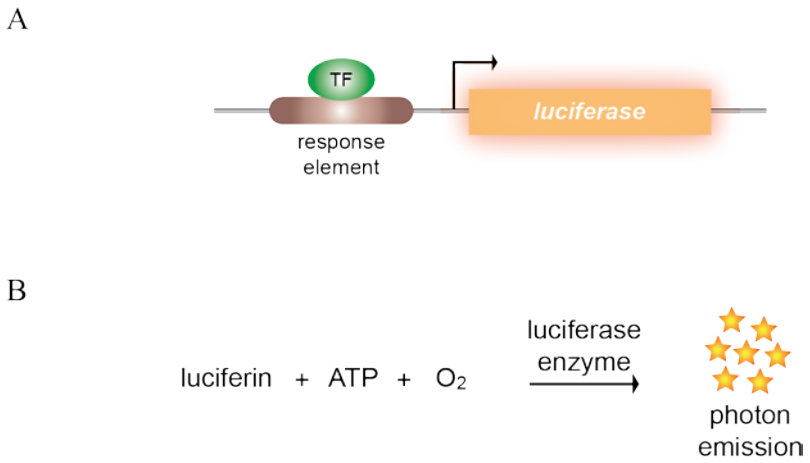


Figure 4.1: Luciferase reporter gene assay. (A) Schematic representation of a typical reporter gene construct where a response element is cloned upstream of the luciferase cDNA. Transcription factors (TF) bind to specific DNA sequences on the response element and activate transcription of luciferase. **(B)** Chemical reaction catalyzed by the luciferase enzyme. The substrate luciferin, in the presence of adenosine triphosphate (ATP) and oxygen (O_2), is oxidized by the luciferase enzyme, which leads to emission of photons.

The physics of bioluminescence provides inherent advantages for cell-based assays since the emission of light is due to an innate exothermic chemical reaction (Watson, 2011). In contrast, fluorescence based methods require an external excitation source that can cause off target fluorescence from cellular components resulting in a low signal-to-background ratio, which poses difficulties in detecting small changes. Since mammalian systems lack bioluminescence, background is basically non-existent allowing detection of as low as 0.05 attomole (10^{-18}) of the enzyme (Pazzagli *et al.*, 1992). Additionally the luciferase protein is primed for function immediately upon translation without the need for any post-translational processing (de Wet *et al.*, 1987), while data generated from luciferase assays are relatively straightforward to analyse (Yun *et al.*, 2014). The primary disadvantage is reagent expense and the requirement of a

luminometer to take measurements. Nonetheless, despite providing a significantly weaker signal intensity than fluorescence, luciferase reporter assays provide exceptional sensitivity and a broad dynamic range with a rapid and quantitative platform for studying transcription factors regulating mammalian gene expression in vitro (Siebring-van Olst *et al.*, 2013).

In order to study the transcription factor function of ZIC proteins (and mutants) a reliable and robust transactivation assay is needed, which provides a potent signal via wild-type ZIC proteins and allows accurate distinction of mutants with hyper- or hypo-morphic activities. These facets rely on the specificity of the response element used to stimulate transcription, type of luciferase cDNA, vector backbone, and the cell line used for transfections.

Initial transactivation assays for ZIC proteins used basal viral promoters such as the herpes simplex virus thymidine kinase (TK) promoter (Koyabu *et al.*, 2001), or the simian virus 40 (SV40) promoter (Ware *et al.*, 2004, D'Alessandro *et al.*, 2013, Cowan *et al.*, 2014). Both promoters display enhanced transactivation in the presence of ZIC1-3 (Mizugishi *et al.*, 2001). This interaction must, however, be nonspecific since these viral promoters do not contain known ZIC DNA-binding sequences, rendering this system sub-optimal for dissection of ZIC protein function. Moreover, as demonstrated by Brown *et al.* (2005), several ZIC2 variants have been found to stimulate the TK promoter in an unpredictable manner when used as a transfection control, thus providing erroneous results.

Additionally Salero *et al.* (2001) used a yeast one-hybrid system followed by electrophoretic mobility shift assays to show that ZIC1 and ZIC2 can bind the human *APOE* promoter region. Subsequently this human *APOE* promoter has been used to assay transactivation via ZIC proteins (Salero *et al.*, 2001, Brown *et al.*, 2005, Ahmed *et al.*, 2013). However the *APOE* promoter is not likely to be a real in vivo target of ZIC proteins as *Apoe* (Harrison *et al.*, 1995) and *Zic* genes (Elms *et al.*, 2004) are not co-expressed during gastrulation (Figure 4.2).

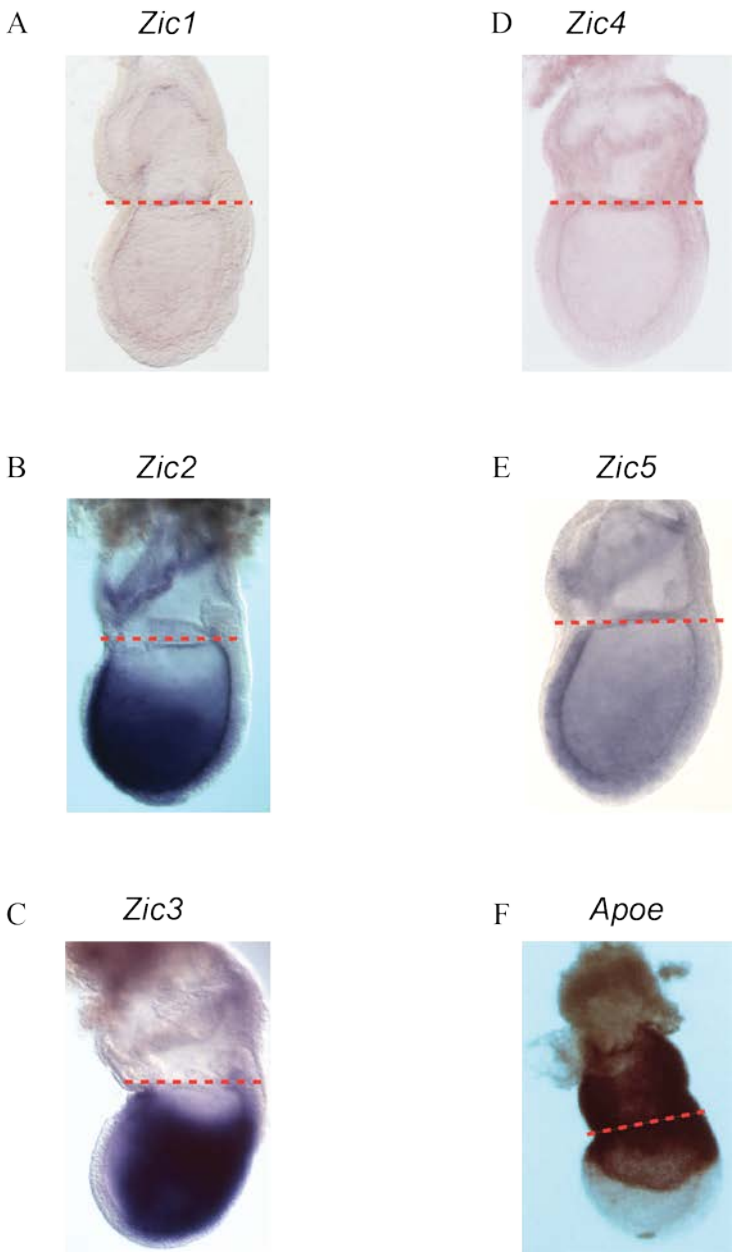


Figure 4.2: *Zic* genes are not co-expressed with *Apoe* during gastrulation. Lateral view of 7.5 dpc mouse embryos following whole-mount in situ hybridization (WMISH) to the genes shown. The (orange) dotted line divides the extra-embryonic (above) and embryonic (below) regions. *Zic1* **(A)** and *Zic4* **(D)** are not expressed during gastrulation. *Zic2* **(B)** and *Zic5* **(E)** are expressed in the extra-embryonic ectoderm and embryonic ectoderm and mesoderm. *Zic3* **(C)** is not expressed in the extra-embryonic region but is expressed in the embryonic ectoderm and mesoderm. *Apoe* **(F)** is expressed only in the extra-embryonic endoderm. WMISH images of *Zic2* and *Zic3* were taken from Elms *et al.* (2004). *Zic1,* Zic4, and *Zic5* WMISH images were taken by Ruth Arkell. *Apoe* WMISH image was taken from Harrison *et al.* (1995).

One difficulty in producing a faithful transactivation assays for ZIC proteins is that very little is known about bona fide DNA-binding sites for ZIC proteins. In case of ZIC3, recent progress has begun to resolve this issue. A universal PBM assay revealed an optimal ZIC3 DNA-binding sequence (Badis *et al.*, 2009). In addition, Lim *et al.* (2010) performed a ChIP-chip experiment on mouse ESCs followed by a de novo motif search and identified similar binding sites for ZIC3 (Figure 4.3). The latter study also showed that ZIC3 is able to bind the *Nanog* promoter region to activate transcription.

A

B

Figure 4.3: 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 ESCs.

4.1.2 Aims

Using these data a new transactivation assay (ZICflash) for ZIC3 was devised. Design of the new transactivation assay for ZIC3 involved:

- an overhaul of the transactivation assay protocol
- using a new generation luciferase reporter
- testing the suitability of synthetic enhancers
- selecting appropriate negative controls
- comparing genomic promoters versus synthetic enhancers
- choosing an appropriate minimal promoter
- finding a combination of components that leads to a robust assay for analyzing ZIC-dependent transactivation.

4.2 RESULTS

4.2.1 An improved transactivation assay protocol

A robust protocol is essential for any experiment since it enables correct inference of the underlying biological mechanism from the experimental data. All experiments require repeat sampling to increase confidence that the correct value has been obtained (Vaux *et al.*, 2012), however, large variation between repeats can obscure small differences between treatment groups. Typically two types of repeats are employed within experimental protocols: (i) internal replicates (or technical repeats) measure experimental error, such as differences in pipetting and; (ii) external replicates (or biological repeats) measure variability between biological samples (Blainey *et al.*, 2014). Transactivation assays performed in the laboratory exhibited high internal variability and efforts were therefore made to improve experimental design and reduce variation between technical replicates.

The transfection protocol used initially (Protocol 1) involved growing cells, prior to transfection, in a 96-well plate. A transfection mix was then made for each treatment group and the mix split between each of four wells to produce four internal replicates with luminescence measured from each well 24 hours post-transfection (Figure 4.4A). A new transactivation assay protocol (Protocol 2) was designed (Figure 4.4B), which involved growing cells in a 12-well plate and the addition of the entire transfection mix to one well of cells. 6 hours post-transfection, these cells were split and three aliquots from each treatment group transferred to three wells of a 96-well plate to produce three internal replicates with luminescence from each well measured 18 hours post-splitting. Additionally, media was added to extra wells surrounding the transfected cells to prevent evaporation in the edge wells from affecting cell growth and viability. Moreover, the remaining cells in the 12-well plate were incubated in parallel with those split for luminescence assay and harvested for Western blot analysis.

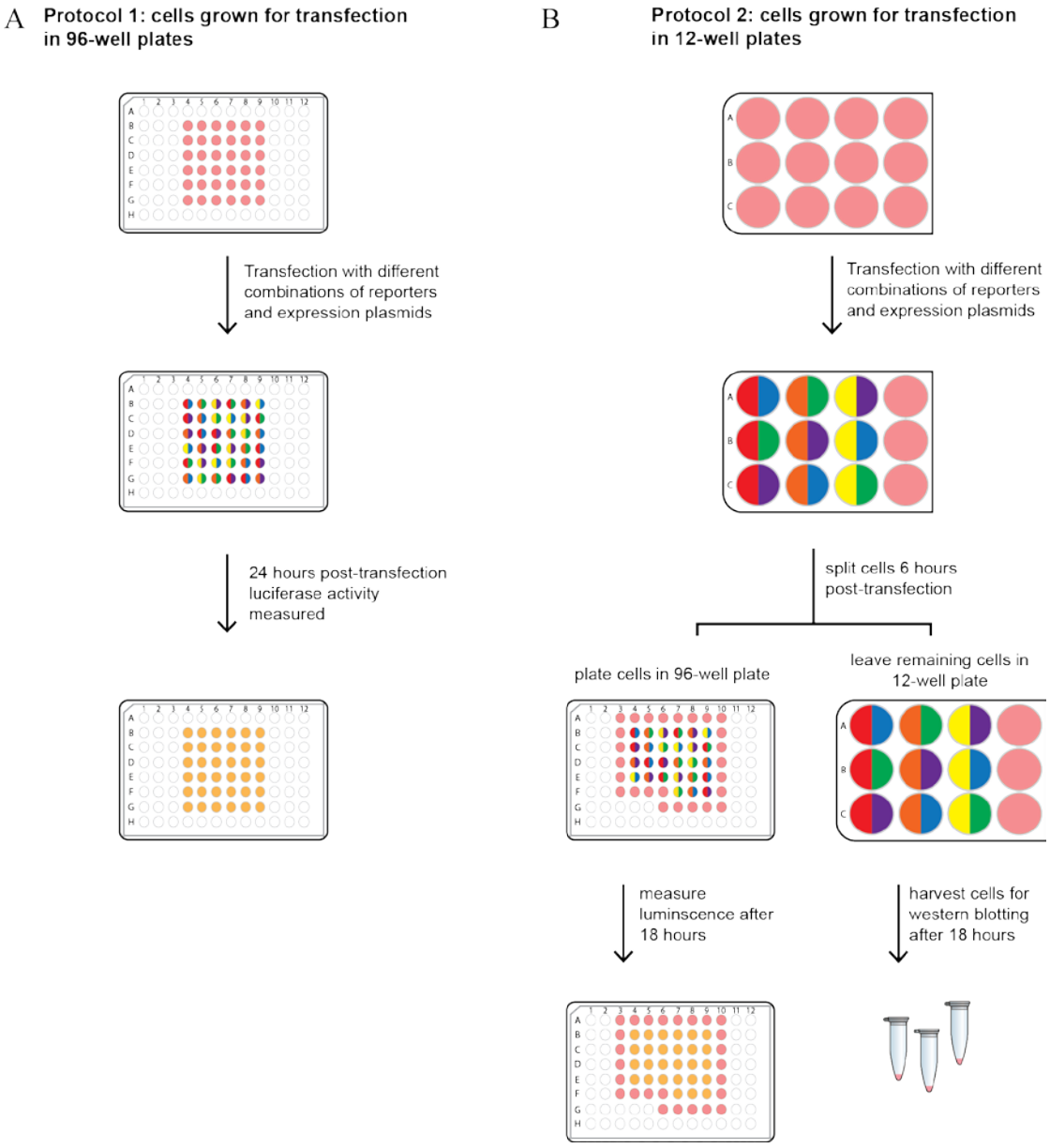


Figure 4.4: Modifications in transactivation assay protocol. Illustration of a typical transactivation assay testing the transactivation ability of three expression plasmids (blue, green and purple semi-circles) from three reporter plasmids (red, orange and yellow semi-circles). **(A)** For protocol 1, cells were grown in a 96-well plate for 24 hours such that at time of transfection each well was 60-70% confluent. Each transfection combination had four internal replicates on the plate. 24 hours after transfection luciferase substrate was added to each well (yellow wells) and luminescence measured using a luminometer. **(B)** For protocol 2, cells were grown in a 12-well plate for 24 hours such that at time of transfection each well was 80-90% confluent. 6 hours after transfection cells were split from 12-well plate and transferred to a 96-well plate such that the following day each well of the 96-well plate was 90-100% confluent. Each transfection combination had three internal replicates on the plate. Media was added to empty wells (red wells) surrounding the wells containing transfected cells. The remaining cells after splitting were incubated overnight in the 12-well plate. 18 hours after splitting (or 24 hours after transfection) luciferase substrate was added to each well (yellow wells) and luminescence measured using a luminometer. Cells from the 12-well plate were harvested, pelleted and snap-frozen to process for Western blotting.

Protocol 1 and 2 were compared using the assay shown in Chapter 3, which compared the transactivation abilities of wild-type and katun ZIC3 at B:*luc*$^{+}$:APOE (Figure 4.5). Protocol 2 reduced variation between the internal replicates (in each experimental repeat; N=3) and reduced the amount of luciferase substrate used.

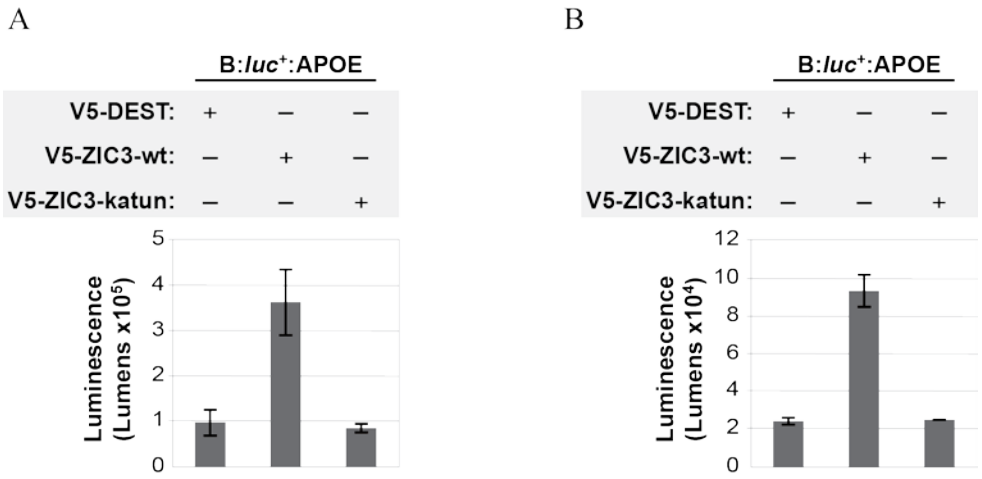


Figure 4.5: A more robust transactivation assay protocol. HEK293T cells were transfected in a 96-well plate **(A)** as in Protocol 1, or a 12-well plate **(B)** as in Protocol 2 with the B:*luc*$^{+}$:APOE reporter and the expression plasmid shown. 24 hours post-transfection luciferase substrate was added and luminescence measured using a luminometer. Errors bars represent standard deviation between three internal replicates. Data presented is from one representative experiment.

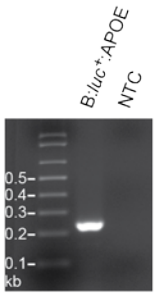
All subsequent transactivation assays were performed using Protocol 2. In all cases the V5-DEST parental vector for all expression constructs was used as one treatment group. This provided a background luminescence measurement of transactivation by endogenous factors (Doppler *et al.*, 2014). Generally throughout subsequent experiments data is presented from one external repeat in lumens, directly reporting the raw luminescence scores from the luminometer and showing the standard deviation (SD) between three internal repeats. A corresponding Western blot shows the presence of transfected proteins. A minimum of three experimental repeats were performed for each treatment group and the raw data analyzed using a two-way ANOVA. Following statistical analysis, the raw luciferase values were converted to relative luciferase activity (RLA) calculated by dividing the raw luminescence value from an expression construct

(e.g. V5-ZIC3-wt) by the background luminescence (measured by V5-DEST) and data presented as the mean RLA from three external replicate experiments, with error bars indicating the standard error of the mean (SEM) and relevant statistics from ANOVA shown on this graph. This method provided a direct indication of the effect ZIC proteins had on a particular response element, and allowed determination of the ramifications of a particular mutation on the transactivational ability of ZIC proteins.

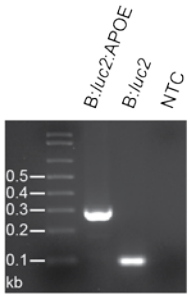
4.2.2 A new generation reporter plasmid (when old is *not* gold)

The *APOE* reporter construct used in Chapter 3 and Figure 4.5 has pXP2 as the backbone. Several new generation vectors are now available which use a codon optimized luciferase CDS (called *luc2*) and engineered to decrease cryptic transactivation from backbone sequences. Promega's pGL4.20 (referred to as B:*luc2* in this thesis) was selected to determine whether these improvements would benefit a ZIC transactivation assay. The *APOE* promoter from the B:*luc*$^{+}$:APOE vector was therefore cloned into B:*luc2* (Figure 4.6A-C) and checked for integrity by sequencing. The new reporter plasmid (B:*luc2*:APOE) was compared to B:*luc*$^{+}$:APOE following co-transfection with V5-DEST or V5-ZIC3-wt (or V5-ZIC3-C365S) in HEK293T cells (Figure 4.6D). The B:*luc2*:APOE reporter exhibited both increased raw luminescence via V5-ZIC3-wt and decreased background luminescence (i.e. with V5-DEST). Consequently, significantly more transactivation was measured at the B:*luc2*:APOE reporter than at the B:*luc*$^{+}$:APOE reporter, despite both constructs containing the same response element (see RLA plot in Figure 4.6D). For both reporter vectors ZIC3-C365S (DNA-binding mutant: see section 4.2.5) was unable to activate transcription. The data indicate the B:*luc2* vector is a better choice for the new ZIC transactivation assay.

A



B



C

Template	Insert expected size (bp)	Restriction digest	Colony PCR expected size (bp)
B:*luc*⁺:APOE	242[a]	*Kpn*I and *Bgl*II	N/A
B:*luc2*	N/A	*Kpn*I and *Bgl*II	95[b]
B:*luc2*:APOE	N/A	*Kpn*I and *Bgl*II	289[b]

a: PCR from B:*luc*⁺:APOE with primers Ark1281-Ark1282.

b: PCR with primers Ark1566-Ark1567 to amplify MCS of B:*luc2* vector.

D

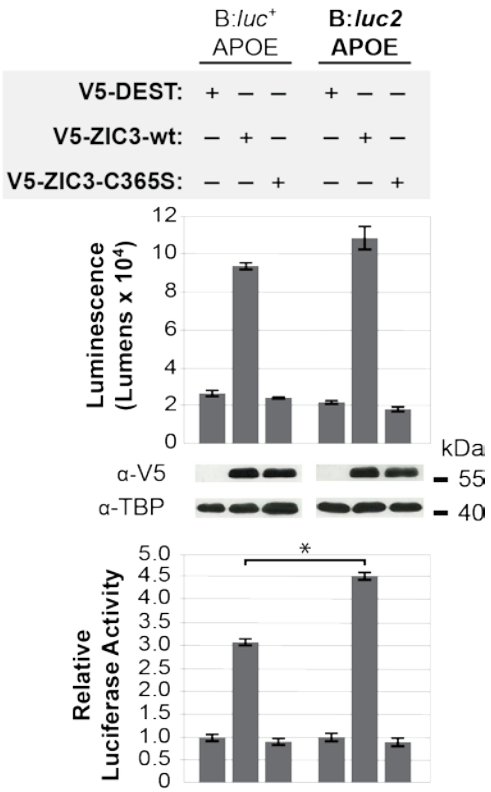


Figure 4.6: Vector backbone influences transactivation assay performance. (A) The *APOE* promoter region (-189/+1) was PCR amplified from the B:*luc*⁺:APOE plasmid and analysed by agarose gel electrophoresis, NTC: no template control . **(B)** The *APOE* promoter was cloned into the B:*luc2* vector and correctly recombined colonies identified via colony PCR and agarose gel electrophoresis. NTC: no template control. **(C)** The expected size of products and other cloning details are shown in the table. **(D)** HEK293T cells were co-transfected with the reporter and expression plasmids shown, and luminescence quantified 24 hours post-transfection. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. *: $p < 0.01$ ANOVA.

4.2.3 ZIC-specific transactivation via synthetic enhancers

The *APOE* promoter used in the experiments above contains three ZIC protein binding sites, in addition to binding sites for SP-1 and AP-2 (Figure 4.7A) (Salero *et al.*, 2001). It is possible that some background activity of this element occurs via endogenous SP-1 and AP-2 transcription factors. The publication of two studies that each identified the same, new ZIC3 binding site (Badis *et al.*, 2009, Lim *et al.*, 2010) in addition to the known ZIC binding site in the *APOE* promoter (Salero *et al.*, 2001) provided the opportunity to generate a synthetic enhancer composed only of ZIC binding sites, such as in the widely used TOPflash reporter plasmid (Korinek *et al.*, 1997). In total, six synthetic ZIC enhancer constructs were designed such that there were six copies of a particular enhancer separated by a five nucleotide spacer fragment (Figure 4.7C). Three constructs contained multimerized copies of the ZIC-protein binding sites from the *APOE* promoter (APOE-M1, APOE-M2 and APOE-M3; Figure 4.7A) and the other three contained sites computed by Lim et al., (Z3M1, Z3M2 and Z3M8; Figure 4.7B). Each construct was cloned into the B:*luc2* vector (Figure 4.8) and the integrity of new plasmids confirmed by sequencing.

The new reporter constructs were tested for transactivation via ZIC3. Amongst the *APOE* enhancers, highest level of luminescence was observed via the APOE-M2 enhancer (Figure 4.9A). However calculation of RLA showed a weak transcriptional response from ZIC3 via APOE-M2 (~3.2 fold), in comparison to stronger activation seen via APOE-M1 (~5 fold) and APOE-M3 (~4.7 fold). Amongst the Z3M enhancers, high levels of luminescence were observed from the Z3M2 and Z3M8 enhancers (Figure 4.9B). Calculation of RLA showed that strongest transcriptional activity of ZIC3 was via Z3M8 (~6.3 fold), followed by Z3M2 (~4.5 fold) and Z3M1 (~1.6 fold). With all reporters, the transcriptional response via ZIC3-C365S was similar to as observed with V5-DEST. Hence four of six synthetic enhancers displayed a greater transcriptional response via ZIC3, in comparison to the genomic *APOE* promoter.

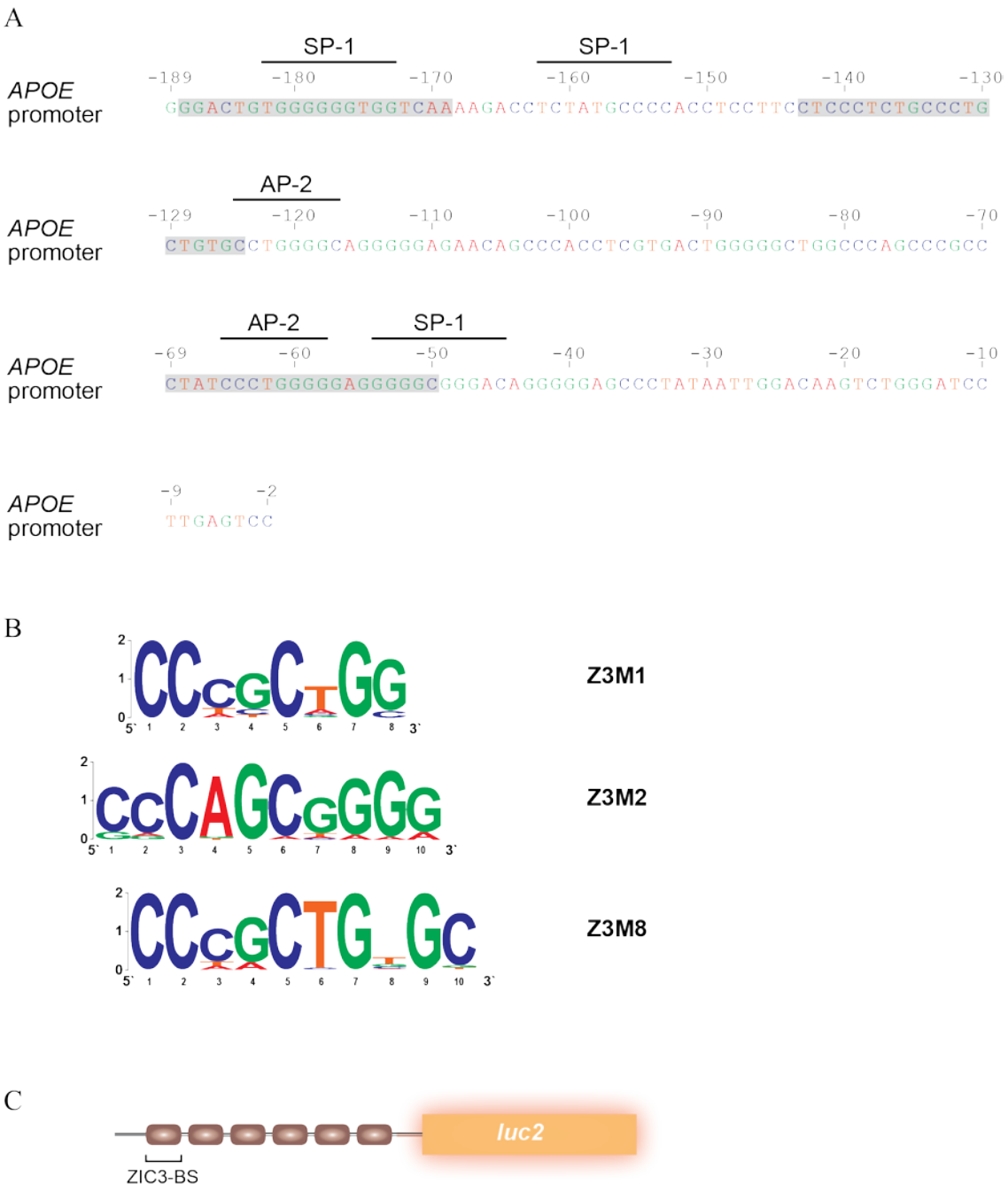
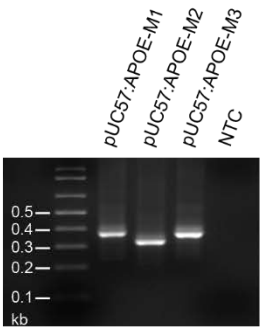
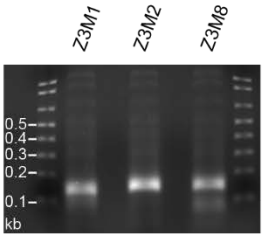


Figure 4.7: Known DNA-binding sites of ZIC proteins. (A) Nucleotide sequence of the human *APOE* promoter. Numbers above the sequence represent distance (in bp) from the major transcription start site. Bases are coloured: adenine (A) in red; cytosine (C) in blue; guanine (G) in green and thymine (T) in orange. Grey boxes highlight binding sites for ZIC proteins (APOE-M1: -188/-169; APOE-M2: -143/-124; APOE-M3: -69/-50). Location of binding sites for transcription factors SP-1 and AP-2 are included. **(B)** Weeder program read-out of ZIC3 DNA-binding sites provided by Lim *et al.* (2010). Data was obtained by performing ChIP-chip on mouse embryonic stem cells. **(C)** The synthetic enhancer was designed to contain six copies of a particular ZIC3-binding site (ZIC3-BS) with each separated by a 5-nucleotide spacer fragment, cloned upstream of the luciferase (*luc2*) CDS.

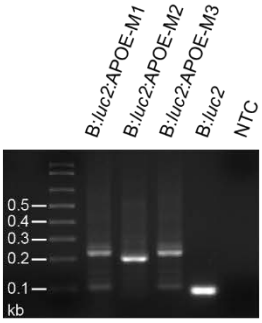
A



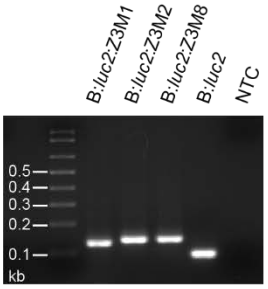
B



C



D



E

Template	Insert expected size (bp)	Restriction digest	Colony PCR expected size (bp)
B:*luc2*	N/A	*Kpn*I and *Bgl*II *Kpn*I and *Hind*III	95[c]
B:*luc2*:APOE-M1	363[a]	*Kpn*I and *Bgl*II	231[c]
B:*luc2*:APOE-M2	314[a]	*Kpn*I and *Hind*III	206[c]
B:*luc2*:APOE-M3	363[a]	*Kpn*I and *Bgl*II	231[c]
B:*luc2*:Z3M1	125[b]	*Kpn*I and *Hind*III	134[c]
B:*luc2*:Z3M2	137[b]	*Kpn*I and *Hind*III	146[c]
B:*luc2*:Z3M8	137[b]	*Kpn*I and *Hind*III	146[c]

a: PCR from pUC57:APOE-M1, pUC57:APOE-M3 pUC57:APOE-M3 with primers Ark106-Ark107.
b: Annealed single stranded oligonucleotides.
c: PCR with primers Ark1566-1567 to amplify MCS of B:*luc2* vector.

Figure 4.8: Construction of synthetic enhancer reporter constructs. (A) The synthetic *APOE* enhancers (APOE-M1, APOE-M2 and APOE-M3) were purchased as cloned fragments in pUC57 vector (Genscript) and were PCR amplified from these vectors using primers Ark106-Ark107 and products analysed by agarose gel electrophoresis. **(B)** The synthetic Z3M enhancers (ordered as single-stranded oligos) were annealed and analysed by agarose gel electrophoresis. **(C, D)** Each synthetic enhancer was cloned into the B:*luc2* vector and success of cloning was determined via PCR of the B:*luc2* MCS using primers Ark1566-Ark1567. Amplicons were resolved on an agarose gel. **(E)** The expected size of products and other cloning details are shown in the table.

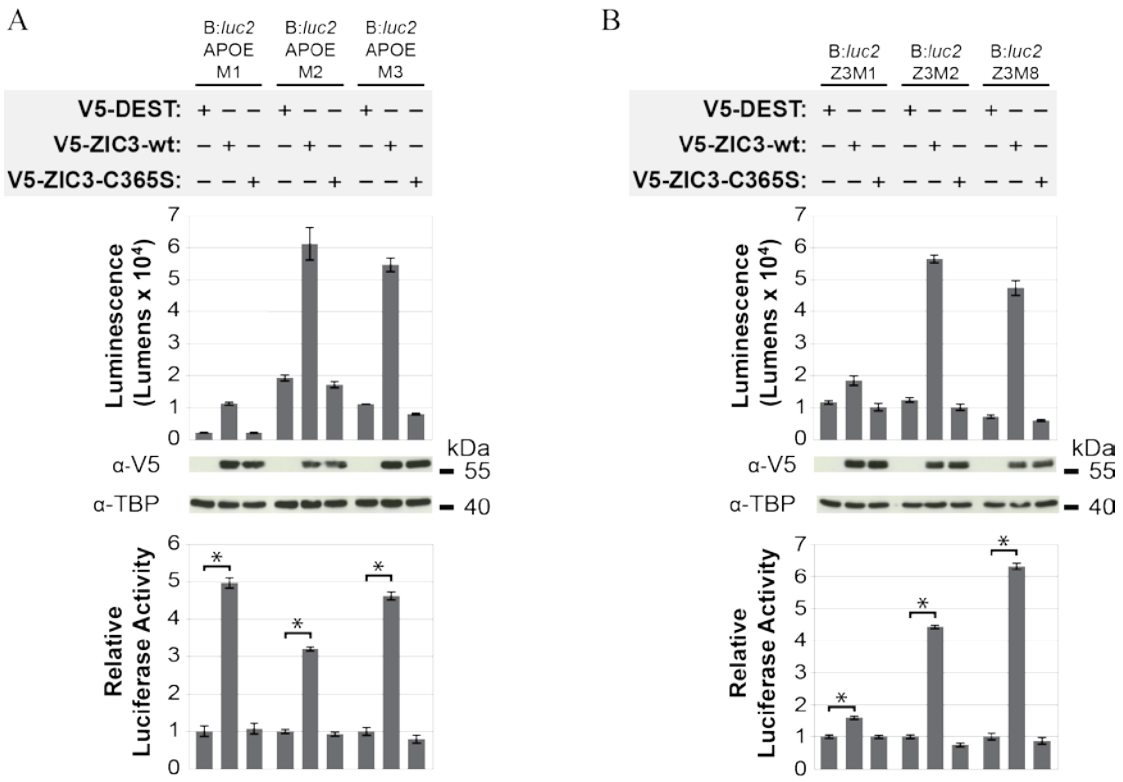
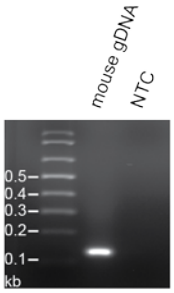


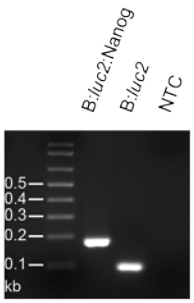
Figure 4.9: Transactivation via reporter vectors containing synthetic enhancers. HEK293T cells were co-transfected with the reporter **(A)** B:*luc2*:APOE-M1, or B:*luc2*:APOE-M2, or B:*luc2*:APOE-M3, or **(B)** B:*luc2*:Z3M1, or B:*luc2*:Z3M2, or B:*luc2*:Z3M8, and the expression plasmids shown. 24 hours post-transfection luciferase expression was measured. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. *: $p < 0.01$ ANOVA.

The choice of response element plays a critical role in the eventual transactivation, since this is the sequence that draws transcription factors to activate transcription of target genes. Thus far ZIC3 seemed to provide a better transcriptional response via the synthetic *APOE* and Z3M motifs. However to conclusively decide on the choice of response element to use for ZIC3-mediate transcription, another genomic promoter region was tested. Lim *et al.* (2010) demonstrated that wild-type ZIC3 binds the *Nanog* promoter region and activates transcription. This *Nanog* promoter region was amplified from mouse genomic DNA, cloned into B:*luc2* (Figure 4.10A-C) and checked for integrity by sequencing. Transactivation via the *Nanog* reporter was compared with the genomic *APOE* promoter and Z3M2 synthetic enhancer, following co-transfection with V5-DEST or V5-ZIC3-wt (or V5-ZIC3-C365S) in HEK293T cells (Figure 4.10D). In comparison to the *APOE* promoter and Z3M2 motif, the *Nanog* promoter region provided the lowest level of luminescence with V5-ZIC3-wt. Highest level of transactivation was measured at the Z3M2 enhancer (~5.6 fold), followed by the *APOE* promoter (~3 fold) and the *Nanog* promoter (~2.5 fold) (see RLA plot in Figure 4.10D). For all reporter vectors ZIC3-C365S (DNA-binding mutant: see section 4.2.5) was unable to activate transcription. The data confirm that the synthetic Z3M2 enhancer is more potent in promoting transactivation via ZIC3 than the genomic promoters.

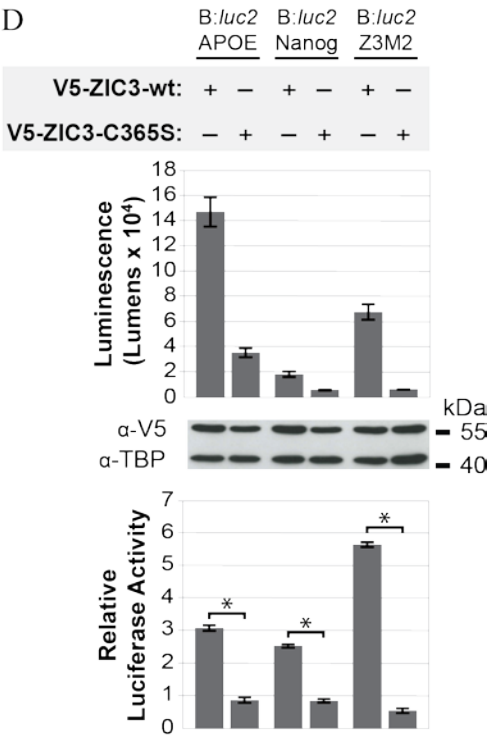
A



B



D



C

Template	Insert expected size (bp)	Restriction digest	Colony PCR expected size (bp)
Mouse gDNA	136[a]	*Kpn*I and *Bgl*II	N/A
B:*luc2*	N/A	*Kpn*I and *Bgl*II	95[b]
B:*luc2*:Nanog	N/A	*Kpn*I and *Bgl*II	190[b]

a: PCR from mouse gDNA with primers Ark1543-Ark1544.
b: PCR with primers Ark1566-Ark1567 to amplify MCS of B:*luc2* vector.

Figure 4.10: Finding the best response element for transactivation via ZIC3. (A) The *Nanog* promoter region was PCR amplified from mouse genomic DNA (gDNA) and analysed by agarose gel electrophoresis, NTC: no template control **(B)** The *Nanog* promoter was cloned into B:*luc2* and correctly recombined colonies identified via colony PCR and agarose gel electrophoresis. NTC: no template control. **(C)** The expected size of products and other cloning details are shown in the table **(D)** HEK293T cells were co-transfected with the reporter and the expression plasmids shown. V5-DEST was also co-transfected with each reporter to measure background (not shown). 24 hours post-transfection luciferase expression was measured. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. *: $p < 0.01$ ANOVA.

4.2.4 Choice of negative control

In order to make judgments regarding the response of an effector in the presence of a stimulus, the two variables must be separated and individually assessed on how they are contributing to the eventual experimental findings. The aim of this study was to determine the efficacy of transactivation via wild-type ZIC3 using different enhancer sequences. This raised two questions: 1) how much luminescence from transfected cells is due to the reporter plasmid alone, and 2) is the change in luminescence, in the presence of wild-type ZIC3, specifically due to DNA-binding? The first question was answered by including the empty expression plasmid (V5-DEST), which allowed measurement of luminescence purely due to the presence of reporter vector in cells. Hence data could be analyzed by comparing luminescence via V5-ZIC3-wt to V5-DEST, or by calculating RLA. To answer the second question, a transcription factor was required that was unable to recognize and bind ZIC3-DNA target sites, thus limiting its ability to activate transcription. The transcription factors used in this study (Figure 4.11A) included: i) ZIC3-katun that has previously been shown to fail to activate transcription (Ahmed *et al.*, 2013), presumably due to lack of the ZFD required for DNA-binding; ii) ZIC3-C365S that contains a missense mutation at the site of the second cysteine in ZF4 (Figure 4.11B). A similar mutation in ZIC2 has been shown to abolish DNA-binding activity (Brown *et al.*, 2005); and iii) CDX2, an unrelated transcription factor that is known to bind to AT-rich DNA sequences (Freund *et al.*, 1998), whereas ZIC proteins generally prefer GC-rich sequences (Aruga, 2004). These transcription factors were tested against the reporters B:*luc2*:Z3M1, B:*luc2*:Z3M2 and B:*luc2*:Z3M8 (Figure 4.11C). Wild-type ZIC3 stimulated transcription from all reporters (as seen before). Surprisingly ZIC3-katun was able to activate transcription significantly with all reporters, albeit not at the same levels as wild-type ZIC3 with B:*luc2*-Z3M2 and B:*luc2*-Z3M8. ZIC3-C365S and CDX2 were unable to activate transcription from all reporters, thus both fulfilled the desired solution to question 2. Nonetheless ZIC3-C365S was selected as the most appropriate negative control, since this variant of ZIC3 differs from wild-type ZIC3 by one amino acid, where CDX2 belongs to a completely different family of transcription factors.

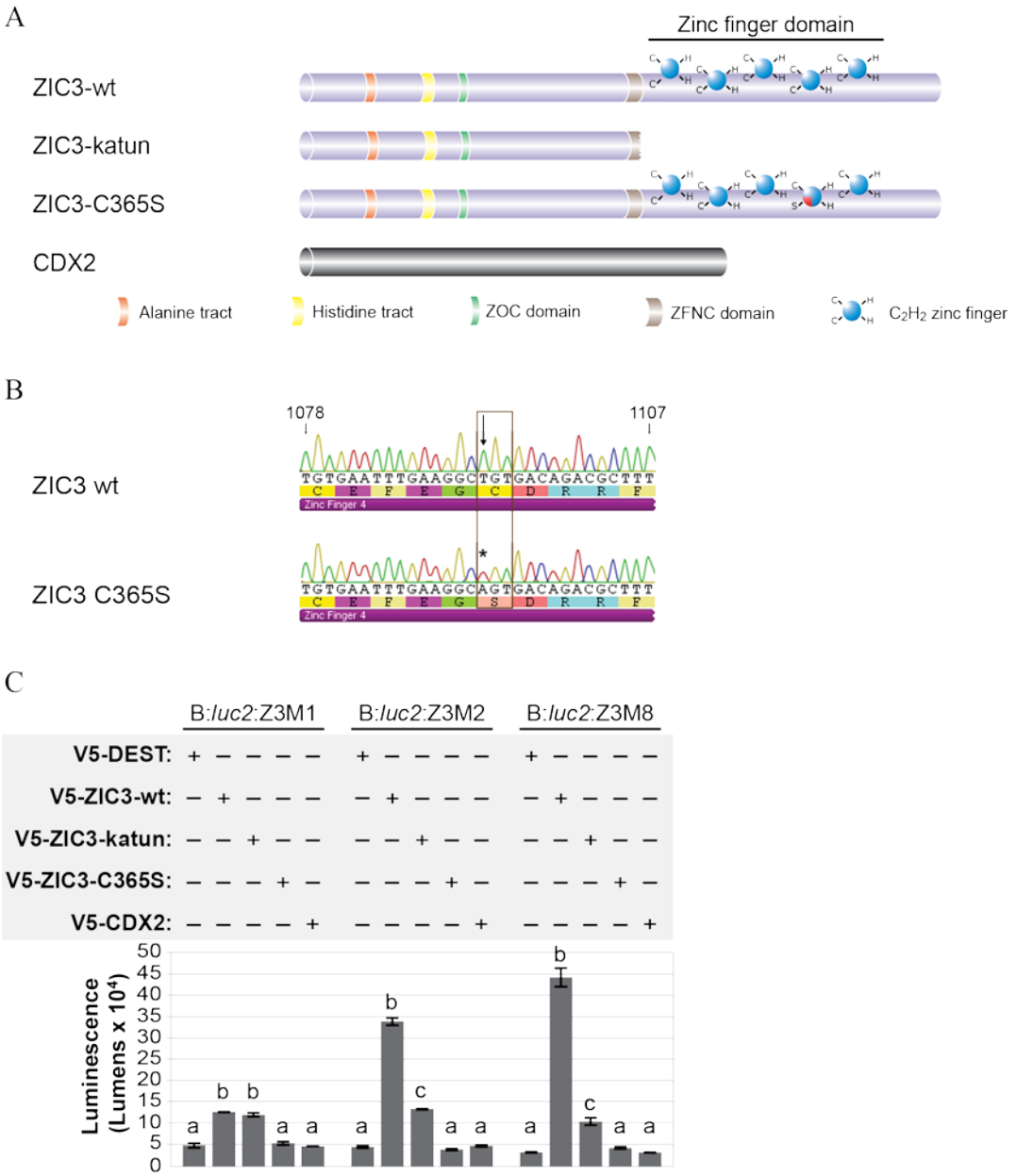


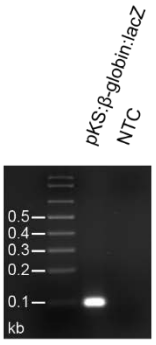
Figure 4.11: Choosing an appropriate negative control. (A) Schematic representation of the proteins used to determine an appropriate negative control. Full length ZIC3-wt contains 467 amino acids. ZIC3-katun cDNA contains a premature termination codon and expresses a truncated protein containing only the first 250 amino acids. ZIC3-C365S is a full length variant containing a point mutation that converts the second canonical cysteine to a serine residue on ZF4. CDX2 is an unrelated transcription factor that contains 313 amino acids. **(B)** Sequence read of ZIC3-wt and ZIC3-C365S from nucleotide positions 1078 to 1107 on the Human *ZIC3* cDNA (NM_003413.3), showing conversion of thymine (T) (black arrow) to adenine (A) (asterick) at position 1093. This conversion changes the triplet codon for cysteine to serine. Nucleotides are in black font with a clear background. Amino acids are in black font with coloured backgrounds. **(C)** HEK293T cells were co-transfected with the reporter and expression plasmids shown, and luminescence quantified 24 hours post-transfection. The graph shows one representative experiment. Error bars represent SD between three internal repeats. a, b and c: $p < 0.01$ ANOVA (performed on data from each reporter). Experiment was repeated twice to confirm results.

4.2.5 Choosing a minimal promoter

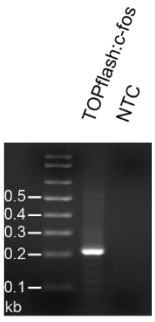
A minimal promoter is a sequence of DNA located ~35 bp upstream and/or downstream of the transcription site that interacts with the basal transcriptional machinery (RNA polymerase II) to initiate transcription (Smale *et al.*, 2003). Presence of other DNA elements recognized by specific transcription factors then allows to increase the rate of transcription. While the Z3M2 synthetic enhancer in comparison to previously used promoters provided reasonably higher levels of transcription via ZIC3, perhaps a further boost could be provided by adding a minimal promoter. The minimal promoters trialed included the human β-globin promoter (amplified from pKS:β-globin:lacZ), the c-fos promoter (amplified from TOPflash:c-fos) and the viral TK promoter (amplified from TOPflash:TK) (Figure 4.12A-C). Amplicons were digested with *Hind*III, cloned into B:*luc2* (Figure 4.12D) and checked for integrity by sequencing.

When tested for transactivation via ZIC3, B:*luc2*:c-fos provided the highest level of luminescence, followed by B:*luc2*:β-globin and B:*luc2*:TK (Figure 4.13). However both wild-type ZIC3 and ZIC3-C365S repressed transcription via the c-fos promoter. Wild-type ZIC3 was significantly able to enhance luciferase expression above background levels via all other reporters, albeit RLA was considerably low. ZIC3-C365S was unable to enhance luciferase expression beyond background levels (V5-DEST).

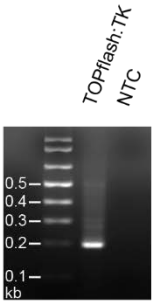
A



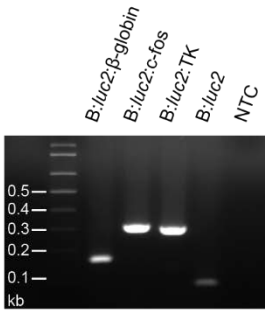
B



C



D



E

Template	Insert expected size (bp)	Restriction digest	Colony PCR expected size (bp)
pKS:β-globin:lacZ	93[a]	*Hind*III	N/A
TOPflash:c-fos	210[b]	*Hind*III	N/A
TOPflash:TK	198[c]	*Hind*III	N/A
B:*luc2*	N/A	*Hind*III	95[d]
B:*luc2*:β-globin	N/A	*Hind*III	179[d]
B:*luc2*:c-fos	N/A	*Hind*III	296[d]
B:*luc2*:TK	N/A	*Hind*III	288[d]

a: PCR from pKS:β-globin:lacZ with primers Ark1510-Ark1507.
b: PCR from TOPflash:c-fos with primers Ark1515-Ark1516.
b: PCR from TOPflash:TK with primers Ark1513-Ark1514.
d: PCR with primers Ark1566-Ark1567 to amplify MCS of B:*luc2* vector.

Figure 4.12: Inclusion of minimal promoters in the B:*luc2* reporter. Minimal promoters **(A)** β-globin, **(B)** c-fos, and **(C)** TK were amplified from parent vectors for cloning and observed on agarose gels, NTC: no template control. **(D)** The minimal promoters were cloned into the B:*luc2* vector and correctly recombined colonies identified via colony PCR and agarose gel electrophoresis, NTC: no template control. **(E)** The expected size of products and other cloning details are shown in the table.

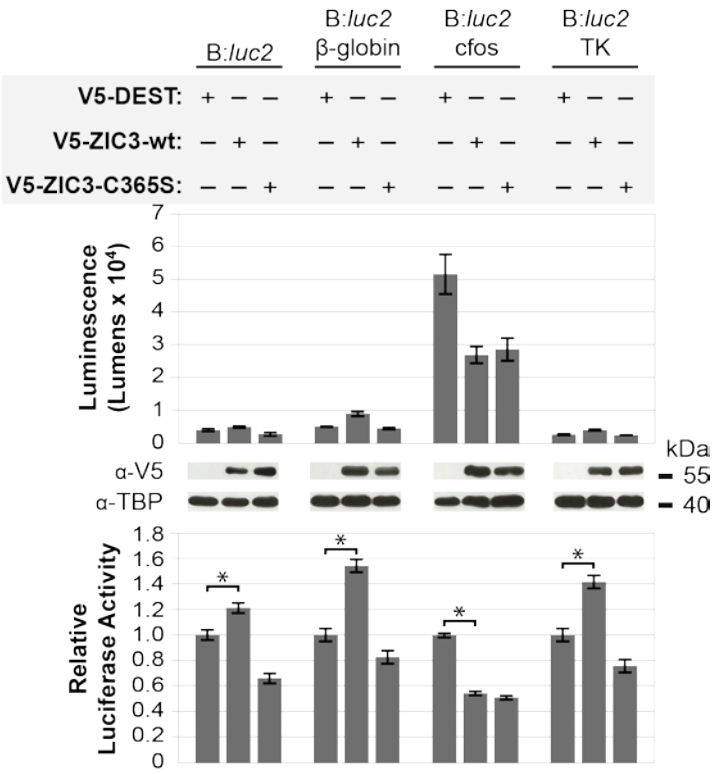
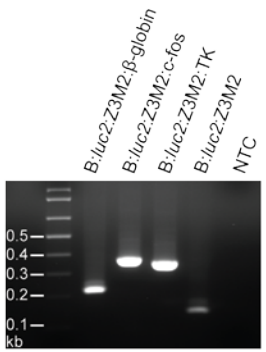


Figure 4.13: Choosing an appropriate minimal promoter. HEK293T cells were co-transfected with the reporter and expression plasmids shown and luminescence quantified 24 hours post-transfection. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. *: $p < 0.01$ ANOVA.

To investigate if these minimal promoters influence transactivation via the Z3M2 motif, they were separately cloned into B:*luc2*:Z3M2. Cloned plasmids were identified via colony PCR (Figure 4.14A-B) and checked for integrity by sequencing. When tested for transactivation via ZIC3 (Figure 4.14C-E), the β-globin promoter in combination with the Z3M2 motif (B:*luc2*:Z3M2:β-globin) showed a remarkable increase in transactivation. While both c-fos and TK promoters were unable to enhance luciferase expression in comparison to when Z3M2 was present alone.

A

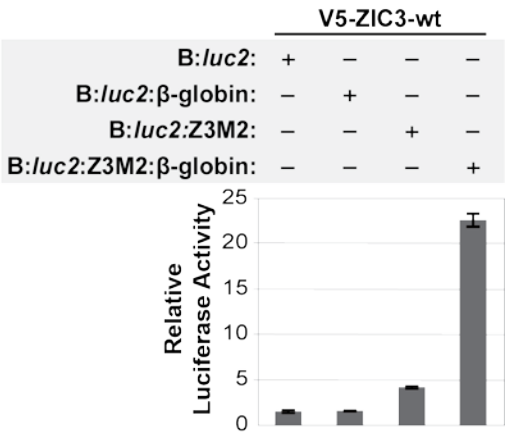


B

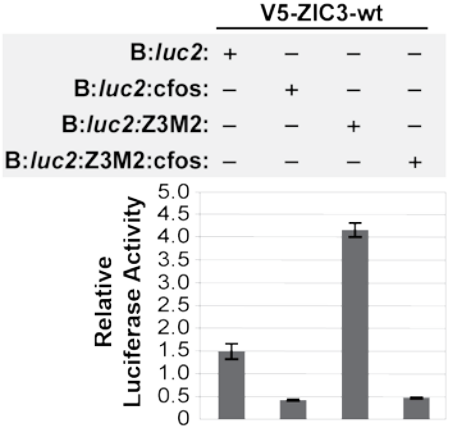
Construct	Restriction digest	Colony PCR expected size (bp)
B:*luc2*:Z3M2	*Hind*III	146[a]
B:*luc2*:Z3M2:β-globin	*Hind*III	228[a]
B:*luc2*:Z3M2:c-fos	*Hind*III	346[a]
B:*luc2*:Z3M2:TK	*Hind*III	334[a]

a: PCR with primers Ark1566-Ark1567 to amplify MCS of B:*luc2* vector.

C



D



E

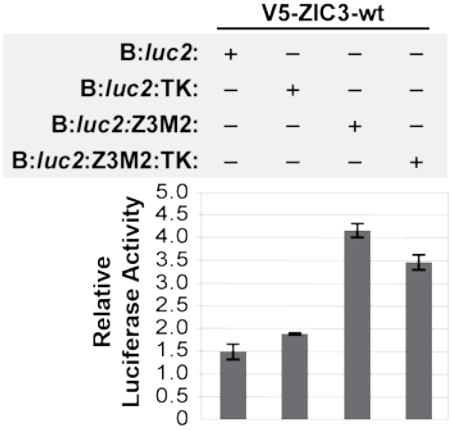
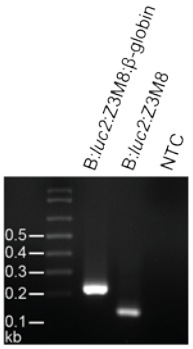


Figure 4.14: Combining a minimal promoter with the Z3M2 synthetic enhancer. (A) The β-globin, c-fos and TK minimal promoters were cloned into B:*luc2*:Z3M2 and correctly recombined colonies identified via colony PCR and agarose gel electrophoresis. **(B)** The expected size of products and other cloning details are shown in the table. **(C, D, E)** HEK293T cells were transfected with the reporter plasmids shown and V5-DEST or V5-ZIC3-wt. Data shown represents RLA via V5-ZIC3-wt for each reporter (with reference to V5-DEST). Errors bars represent SD between internal repeats from one representative experiment. Experiment was repeated three times to confirm results.

Since the β-globin promoter significantly enhanced transactivation in combination with Z3M2, it was also cloned downstream of Z3M8 to investigate if it will have the same effect. The cloned plasmid was identified via colony PCR and checked for integrity by sequencing (Figure 4.15A-B). When tested for transactivation via ZIC3, the combination of β-globin promoter with Z3M8 did not improve transactivation (Figure 4.15C). As a result the B:*luc2*:Z3M2:β-globin was selected as the best reporter to use for transactivation via ZIC3.

A



B

Construct	Restriction digest	Colony PCR expected size (bp)
B:*luc2*:Z3M8	*Hind*III	146[a]
B:*luc2*:Z3M8:β-globin	*Hind*III	229[a]

a: PCR with primers Ark1566-Ark1567 to amplify MCS of B:*luc2* vector.

C

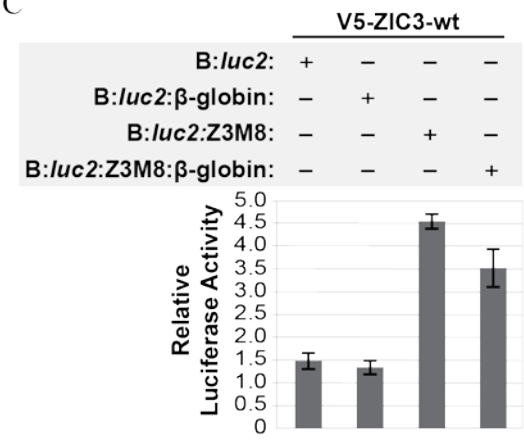
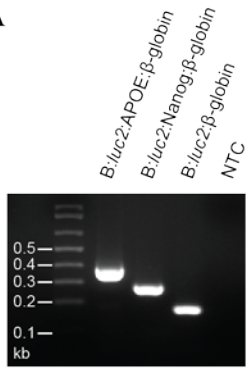


Figure 4.15: Inclusion of the β-globin minimal promoter with Z3M8 does not improve transactivation. (A) The β-globin promoter was cloned into B:*luc2*:Z3M8 and correctly recombined colonies identified via colony PCR and agarose gel electrophoresis. **(B)** The expected size of products and other cloning details are shown in the table. **(C)** HEK293T cells were co-transfected with the reporter plasmids shown and V5-DEST or V5-ZIC3-wt. Data shown represents RLA via V5-ZIC3-wt for each reporter (with reference to V5-DEST). Errors bars represent SD between three internal repeats from one representative experiment. Experiment was repeated three times to confirm results.

The β-globin promoter was also combined with the *APOE* and *Nanog* promoters to investigate if transactivation via these genomic promoters can be improved. Cloned plasmids were identified via colony PCR (Figure 4.16A-B) and checked for integrity by sequencing. When tested for transactivation via ZIC3, B:*luc2*:Nanog:β-globin displayed the lowest luminescence in comparison to B:*luc2*:APOE:β-globin and B:*luc2*:Z3M2:β-globin, while addition of the β-globin promoter did not improve transactivation via the *APOE* and *Nanog* promoters (Figure 4.16C). As a result, B:*luc2*:Z3M2:β-globin was chosen for all further transactivation assays involving ZIC3. To confirm transactivation via B:*luc2*:Z3M2:β-globin was specifically due to the DNA-binding of ZIC3, ChIP-qPCR was performed on HEK293T cells transfected with the reporter plasmid and ZIC3-wt or ZIC3-C365S (Figure 4.16D). Preliminary data showed that chromatin extracted from cells transfected with ZIC3-wt displayed significantly higher enrichment of the Z3M2:β-globin region in comparison to ZIC3-C365S, indicating that interaction of ZIC3 with the Z3M2:β-globin region is required for enhancing luciferase expression.

A

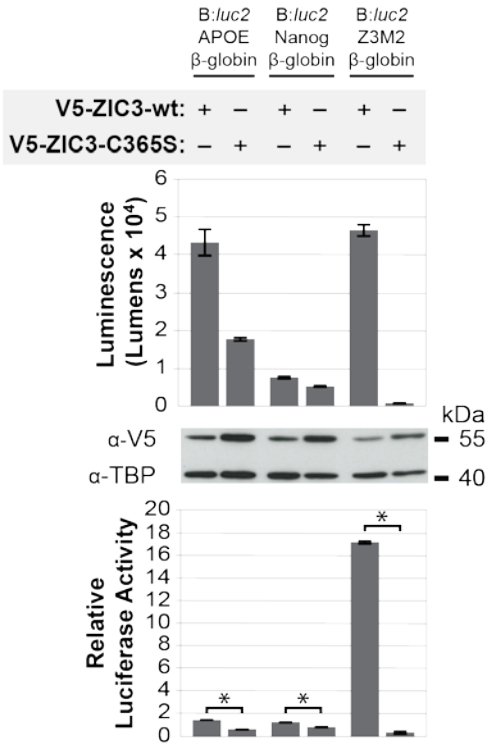


B

Construct	Restriction digest	Colony PCR expected size (bp)
B:*luc2*:β-globin	*Kpn*I and *Bgl*II	179[a]
B:*luc2*:APOE:β-globin	*Kpn*I and *Bgl*II	373[a]
B:*luc2*:Nanog:β-globin	*Kpn*I and *Bgl*II	273[a]

a: PCR with primers Ark1566-Ark1567 to amplify MCS of B:*luc2* vector.

C



D

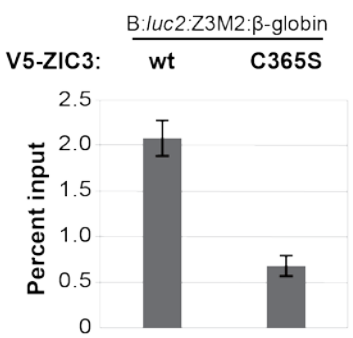


Figure 4.16: Z3M2 synthetic enhancer combined with β-globin minimal promoter boosts transactivation via ZIC3. (A) The *APOE* and *Nanog* promoters were cloned into B:*luc2*:β-globin and correctly recombined colonies identified via colony PCR and agarose gel electrophoresis. **(B)** The expected size of products and other cloning details are shown in the table. **(C)** HEK293T cells were co-transfected with the reporter and the expression plasmids shown. V5-DEST was also co-transfected with each reporter to measure background (not shown). 24 hours post-transfection luciferase expression was measured. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. *: $p < 0.01$ ANOVA. **(D)** qPCR output following ChIP from HEK293T cells transfected with B:*luc2*:Z3M2:β-globin and V5-ZIC3-wt or V5-ZIC3-C365S. Errors bars represent SD within three internal repeats.

4.3 DISCUSSION

4.3.1 Good experimental design decreases variability

A PubMed search (conducted at the time of writing this thesis) using the terms "luciferase reporter assay" retrieved more than 13,000 publications (since 1987). Despite their popularity, reporter gene assays are plagued by external and internal variability that can be introduced due to several non-biological factors (Table 4.1). While designing the new transactivation assay for ZIC3, the effects of these factors were considered and efforts made to minimise their influence on the experimental data.

Factor	Type of variation effected	Reference
Cell Culture		
Passage number	External	Dalby *et al.*, 2004 Yun *et al.*, 2014
Cell number/density	External	Dalby *et al.*, 2004 Yun *et al.*, 2014
Edge effects	Internal	Houck *et al.*, 2006 Maddox *et al.*, 2008
Incubator effects	External	Maddox *et al.*, 2008
Pipetting inconsistencies	Internal	Schagat *et al.*, 2007 Yun *et al.*, 2014
Transfection factors		
Cytotoxicity	External	Felgner *et al.*, 1987 Zhi *et al.*, 2010
Plasmid DNA concentration	External	Felgner *et al.*, 1987 Romoren *et al.*, 2004
Luminescence assay		
Plate position effects	Internal	Malo *et al.*, 2006
Reagent stability	External	Lundin, 2014
Luminometer	External	Stanley, 1992

Table 4.1: Non-biological factors that can distort the outcome of reporter gene assays. Each of these variables can affect the external and/or internal variability, as documented in the listed references.

4.3.1.1 Practices used to minimize external variability

Cell passage number: increasing passage number causes genetic drift (Freshney, 2010), reduces transfection efficiency and/or loss of reporter activity (Miraglia *et al.*, 2011). All cell lines used in this study were expanded and frozen at an early passage (p = 1-2) and all experiments performed within passage 4-15.

Cell density: plating density can impact cell growth (Dalby *et al.*, 2004) and molecular function of the protein of interest (Zhang *et al.*, 2001, Tsuji *et al.*, 2005). The appropriate plating density was determined based on cell growth and the same plating procedure used for each experiment.

Cytotoxicity: transfection reagents are known to be cytotoxic (Felgner *et al.*, 1987, Chesnoy *et al.*, 2000, Hawley-Nelson *et al.*, 2008). Appropriate concentration of Lipofectamine for high transfection efficiency was determined and the Lipofectamine:cell number ratio kept constant for each experiment.

Plasmid DNA concentration: high plasmid DNA concentrations can adversely influence transfection efficiency (Felgner *et al.*, 1987, Romoren *et al.*, 2004). Plasmid DNA amounts were optimised by observing the desired expression of transfected constructs. In addition fresh DNA preparations, quantified via spectrophotometry and agarose gel electrophoresis, were used for each experiment.

Incubator effects: changing plate position within the incubator can affect cell growth due to inconsistencies in temperature regulation and CO_2 presence in different areas of the incubator (Maddox *et al.*, 2008). All incubations were performed by keeping plates in the same position within the incubator.

Luciferase substrate stability: luciferin is highly sensitive to light and oxygen, and is rapidly degraded when exposed to these agents (Lundin, 2014). Small aliquots of the substrate were stored and reasonably fresh preparations were used for each experiment.

Type of luminometer: luminometers vary in their dynamic ranges (Stanley, 1992), which can affect their ability to correctly measure luminescence from certain treatment groups where changes are too high or too low. Thus the assay was optimised using a particular instrument and

the same equipment consistently used for each experiment.

4.3.1.2 Practices used to minimize internal variability

Experiments performed in Chapter 3 showed that Protocol 1 exhibited minimal external variation. However a significant level of internal variation was prevalent, which led to high SD resulting in a large SEM; since SEM is directly proportional to SD (Altman *et al.*, 2005). This prompted a redesign of the experimental setup to allow comparisons between treatment groups with small but meaningful differences.

Pipetting inconsistencies: pipetting errors introduced during plating or transfection are exacerbated over the duration of the experiment and can greatly influence the eventual result (Schagat *et al.*, 2007, Yun *et al.*, 2014). Within internal replicates, pipetting inconsistencies can affect cell numbers, DNA and transfectant concentrations and the transfection efficiency. Protocol 2 minimised pipetting steps prior to transfection and utilised master mixes to reduce internal variability.

Edge effects: Routine experimentation using Protocol 1 showed wells at the perimeter of the 96-well plate exhibited significantly different luminescence relative to those on the interior. This phenomenon, known as “edge effects”, markedly increases internal variation for transiently transfected cells (Houck *et al.*, 2006, Maddox *et al.*, 2008). Reasons for edge effects include thermal and gas gradients (Burt *et al.*, 1979, Oliver *et al.*, 1981) and differential adsorption characteristics across the plate (Kricka *et al.*, 1980). In this study edge effects were circumvented by addition of culture medium in wells surrounding the perimeter sample wells. This approach limited use to 60 of 96 plate wells but significantly reduced variation between internal replicates.

Positional effects in luminometer: luminometers vary in their ability to measure luminescence from individual wells (Malo *et al.*, 2006), a phenomenon known as positional effect, and randomising the position of samples on the 96-well plate greatly decreased internal variability.

These methods combined to decrease internal variability and ensure reproducibility of experimental data. Moreover Protocol 2 both saved on luciferase substrate (as the total

numbers of assayed wells is reduced) and enabled post-read processing of transfected cells for Western analysis.

4.3.1.3 Other experimental considerations

A common method to circumvent sample-to-sample variability is the use of an internal control plasmid, encoding a different reporter gene (for example, the *Renilla* luciferase gene), in each treatment. The rationale being that this reporter will be constitutively expressed in transfected cells and not altered by transfected proteins (Sherf *et al.*, 1996). As such expression from this plasmid should correlate to the transfection efficiency and the cells' general ability to express proteins (Schagat *et al.*, 2007). This method is, however, prone to error for a variety of reasons. For example, particular treatment groups can influence the expression of the internal control plasmid (Everett *et al.*, 1999, Ibrahim *et al.*, 2000, Zhang *et al.*, 2003, Ghazawi *et al.*, 2005) and the presence of the internal control plasmid can increase competition for plasmid entry into cells. Since this method can erroneously alter experimental results it was not used throughout these experiments. Instead, well-to-well variation was detected by analysis of levels of the overexpressed proteins by Western blot and exclusion of entire experiments in cases where unequal transfection could have skewed measurements. The small internal and external variation seen in the data validates this approach.

Another means of minimizing erroneous conclusions, due to biological and experimental variation in the system, is the use of statistical tests (Plant *et al.*, 2000). T-tests, although commonly used, suffer from the disadvantage of accruing multiple type 2 errors (McKillup, 2011). For all transactivation assays reported in this thesis, a two-way analysis of variance (ANOVA) was used to assess the effect of two independent variables (or factors) on one measurable variable and examine the contribution of individual factors in the total variability of the data (Wabed *et al.*, 2010). This type of statistical testing has extensively been used for reporter gene assays (Sharma *et al.*, 2003, Berwick *et al.*, 2010, Vaeth *et al.*, 2014). The two independent variables in the transactivation assays reported here were: (i) technical repeats

(number of experiments), and (ii) transfection treatment groups. The measurable variable was the response of cells to the transfection treatment across each technical repeat. Hence the two-way ANOVA tested three null hypotheses: 1) there is no difference in the response of cells between technical repeats, 2) there is no difference in the response of cells to individual transfection treatments, and 3) there is no interaction between technical repeats and transfection treatments and does that influence the response of cells? For all transactivation assays reported when differences were statistically significant, null hypothesis 2 was rejected, while null hypothesis 1 and 3 failed to be rejected.

4.3.2 ZICflash reporter boosts ZIC3-dependent transactivation

4.3.2.1 A suitable reporter vector

The reporter vector can significantly influence experimental outcome of transactivation assays. The ideal reporter vector should express uniformly and optimally in host cells, have minimal off-target responses and respond sensitively to transcriptional dynamics (Paguio *et al.*, 2005). At the start of this project, transactivation assays for ZIC proteins utilised a known ZIC-responsive reporter construct containing a fragment of the human *APOE* promoter (Salero *et al.*, 2001). The reporter backbone, pXP2, was constructed more than 25 years ago (Nordeen, 1988). There were two problems associated with this vector. Firstly the *APOE* promoter is most likely a non-specific target of ZIC proteins, since the *Apoe* and *Zic* genes are not co-expressed during gastrulation (Figure 4.2). Secondly background luminescence via the plasmid was unsuitably high, potentially due to binding sites for several transcription factors such as CCAAT/enhancer binding protein (C/EBPα) (Li *et al.*, 1994) and activator protein-1 (AP-1) (Grimm *et al.*, 1999), within the pXP2 backbone. Reporter plasmids, such as the pGL4 vectors, have since been modified by removal of cryptic regulatory sequences (Paguio *et al.*, 2005) and codon optimisation of the luciferase cDNA (called *luc2*) for expression in mammalian cells (Almond *et al.*, 2003). Comparison of *luc*$^{+}$ and *luc2* cDNAs showed that luciferase expression via *luc2* was 4- to 11-fold higher than *luc*$^{+}$ across different mammalian cell lines (Paguio *et al.*, 2005). The pGL4.20 vector (referred to as

B:*luc2* in this thesis) was purchased from Promega and exhibited an increased signal:noise ratio when the *APOE* promoter was used to drive luciferase expression via ZIC3 (Figure 4.6). Furthermore analysis of the transcriptional landscape from a variety of currently available reporter vectors showed that the pGL4 backbone had the lowest level of spurious transcription, indicating that this reporter will have minimal background in most cell lines (Nejepinska *et al.*, 2012).

4.3.2.2 Genomic promoters vs synthetic enhancers

Expression of reporter genes in transactivation assays is generally driven by two types of response elements: (i) genomic promoters, or (ii) artificial/synthetic enhancers, either of which contain specific binding sites for the transcription factor being studied. Genomic promoters have the advantage of providing a native biological context, since they contain regions of chromosomal DNA known to carry sequences required for the controlled expression of a particular gene (Miraglia *et al.*, 2011). However signal induction from such elements can be low (Paguio *et al.*, 2010). In addition genomic promoters innately contain binding sites for other transcription factors which can influence reporter gene expression, complicating data interpretation. Synthetic enhancers are sequences of DNA that do not exist in nature and have been designed to control gene expression based on knowledge of natural DNA-binding sites of the transcription factor (Roberts, 2011). Since transcriptional enhancement from a single transcription factor binding site may be too low to be reliably distinguished from background, synthetic enhancers are designed using a building block approach such that multiple copies of the transcription factor binding sites are placed upstream of the reporter gene to initiate RNA polymerase II-mediated transcription (Schlabach *et al.*, 2010). This approach has successfully been used to produce synthetic elements that display significantly higher reporter gene expression, in comparison to naturally occurring promoters (Jacquet *et al.*, 1989, Mader *et al.*, 1993). Although synthetic enhancers do not necessarily represent the true nature of transcriptional control via the transcription factor within the genome, they provide a useful tool

to examine the potencies of transcription factors (Korinek *et al.*, 1997, Li *et al.*, 1999).

To improve transactivation via ZIC3, synthetic enhancers were created using the known ZIC-binding sites in the *APOE* promoter (Salero *et al.*, 2001) and the ChIP-chip identified Zic3-binding sites on the *Nanog* promoter (Lim *et al.*, 2010) (Figure 4.7). While the luminescence generated from each enhancer was lower than the genomic *APOE* promoter, four (APOE-M1, APOE-M3, Z3M2 and Z3M8) of the six enhancers provided better transactivation (Figure 4.9). APOE-M1 and APOE-M3 motifs also contain known binding sites for SP-1 and AP-2, thus transactivation via these elements cannot be conclusively attributed to the activity of ZIC3. In contrast, Z3M2 and Z3M8 solely represent ZIC3-binding sites and thus provide a better alternative for assaying transactivation via ZIC3 and its variants. To find the most potent response element for transactivation via ZIC3, the *APOE* and *Nanog* genomic promoter regions were tested in addition to the Z3M2 synthetic enhancer. The genomic promoters displayed weak transactivation in comparison to the Z3M2 synthetic enhancer (Figure 4.10C), providing further evidence that the latter is a more suitable choice for the new transactivation assay. Using the synthetic enhancer increased sensitivity of the assay (Yun *et al.*, 2014), eliminated issues regarding repressive elements within genomic promoters and provided a "clean" readout of transcriptional activity mediated via the DNA-binding of ZIC3.

4.3.2.3 DNA-binding mutant serves as the best negative control

The most fundamental negative control involves co-transfection of an 'empty' expression plasmid with the test reporter construct in order to measure background luminescence (Salero *et al.*, 2001, Mizugishi *et al.*, 2004, Glait *et al.*, 2006, Lim *et al.*, 2010, Posokhova *et al.*, 2015). This control measures transactivation via endogenous factors (i.e. at cryptic binding sites on the reporter plasmid) and was routinely used herein. It does not, however, reveal the mode of transactivation of the transcription factor of interest. Ideally a second type of control that distinguishes whether the observed transactivation is due to DNA binding of the test transcription factor or to a secondary activation (i.e. where the transfected protein binds an

endogenous protein which contacts the DNA) should also be employed. For this purpose, three controls were considered (Figure 4.11):

(i) An unrelated transcription factor (Cdx2) that cannot transactivate at ZIC3 DNA-binding sites.

(ii) A non DNA-binding structural mutant, (the ZIC3 katun protein that lacks the DNA-binding domain (Ahmed *et al.*, 2013).

(iii) A non DNA-binding point mutant that contains other functionally important structural units.

A variant of murine ZIC2 has been identified that meets this criterion. The ZIC2-C370S variant (with a cysteine to serine substitution in ZF4) is unable to bind DNA or activate transcription (Brown *et al.*, 2005) and causes developmental defects in mouse embryos (Nolan *et al.*, 2000, Elms *et al.*, 2003b). The equivalent mutation was introduced into the ZIC3 cDNA (to produce ZIC3-C365S) and defective DNA-binding confirmed via ChIP-qPCR (Figure 4.16D).

Controls (i) and (iii) failed to activate transcription at the Z3M synthetic enhancers, but the katun mutant displayed some transactivation ability, perhaps because truncated transcription factor mutants can behave unpredictably by interacting with co-regulators (Kiefer *et al.*, 2003, Rudd *et al.*, 2007, Zheng *et al.*, 2010). Overall, the availability of a full length ZIC3 variant (Figure 4.11A) that specifically blocked DNA-binding was judged the best control for specificity of the ZIC3-response element interaction (Figure 4.11C) and ZIC3-C365S selected for use in subsequent experiments.

4.3.2.4 Inclusion of minimal promoter alters signal

Synthetic enhancers may not efficiently initiate transcription, thus inclusion of a minimal promoter can facilitate assembly of basal transcriptional machinery on the minimal promoter and increase presence of transcription factor on the synthetic enhancer (Butler *et al.*, 2001). This strategy has successfully been used in reporter gene studies to enhance reporter expression via mammalian expression vectors (Foecking *et al.*, 1986, Czarnecka *et al.*, 2012, Redden *et al.*,

2015). Key features of a suitable minimal promoter are that it is unresponsive to the transcription factor being studied and does not display high background activity due to endogenous factors. Previous investigations into transactivation via ZIC proteins used the TK (Koyabu *et al.*, 2001) or SV40 (Ware *et al.*, 2004) promoters for driving reporter gene expression. Since the transcription assay reported here employs HEK293T cells, the SV40 promoter was not useful as several reports suggest that the host cell machinery keeps it constitutively active (Rio *et al.*, 1985, Bullock, 1997). The TK promoter has low background activity in HEK293T cells (Chen *et al.*, 2011) and can drive low levels of reporter gene expression in the presence of ZIC3 (Figure 4.13). The human β-globin promoter showed similar trends as TK, while the c-fos promoter displayed high background and reduced transactivation in the presence of ZIC3. When combined with the Z3M2 synthetic enhancer (Figure 4.14C-E), the human β-globin promoter displayed transactivation enhancement to a level never observed before for any ZIC transactivation assay. Furthermore it produces low background in HEK293T cells (Figure 4.13) (Soni *et al.*, 2014) and is not known to be a real target of ZIC proteins in vivo, since during embryonic development expression of β-globin is relatively low and the promoter is most active in the adult bone marrow (Mantovani *et al.*, 1988) where ZIC proteins are not known to function.

CHAPTER 5

LINKING STRUCTURE TO FUNCTION

5.1 INTRODUCTION

ZIC proteins can directly influence gene expression by binding to DNA and activating transcription (Salero *et al.*, 2001, Brown *et al.*, 2005, Lim *et al.*, 2010), or indirectly by binding to TCF and inhibiting transcription of Wnt target genes (Pourebrahim *et al.*, 2011, Fujimi *et al.*, 2012). These functions require (amongst others) a DNA-binding domain, a transactivation domain and a protein-protein interaction domain. The DNA-binding domain recognizes specific DNA sites and determines the transcriptional target genes (Latchman, 2008, Zhu *et al.*, 2011). When bound to target DNA, the transactivation domain interacts with the transcriptional machinery to enhance the rate of gene expression. The protein-protein interaction domain binds to specific domains of partner proteins to increase or repress transcription (Wolberger, 1999). For ZIC proteins we lack knowledge regarding the location of these and other functional domains.

5.1.1 C2H2 type zinc fingers

The ZIC ZFD is comprised of five tandem C_2H_2 ZFs. This type of ZF was first identified in the *Xenopus* transcription factor IIIA (TFIIIA) (Miller *et al.*, 1985) and ~700 human proteins contain this motif (Weirauch *et al.*, 2011). The C_2H_2 consensus sequence is (F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)-X_5-ψ-X_2-H-X_{3-5}-H, where X represents any amino acid and ψ is a hydrophobic residue (Klug *et al.*, 1995, Wolfe *et al.*, 2000). This motif is able to tetrahedrally bind a zinc ion (Zn^{2+}) via the conserved cysteine (C) and histidine (H) residues (Freedman *et al.*, 1988). Initial work on TFIIIA revealed the cysteine loop [(F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)] forms two antiparallel β-sheets (Figure 5.1) while the histidine loop (X_5-ψ-X_2-H-X_{3-5}-H) forms an α-helix, and these structural units are held together by the Zn^{2+} ion (Berg, 1988).

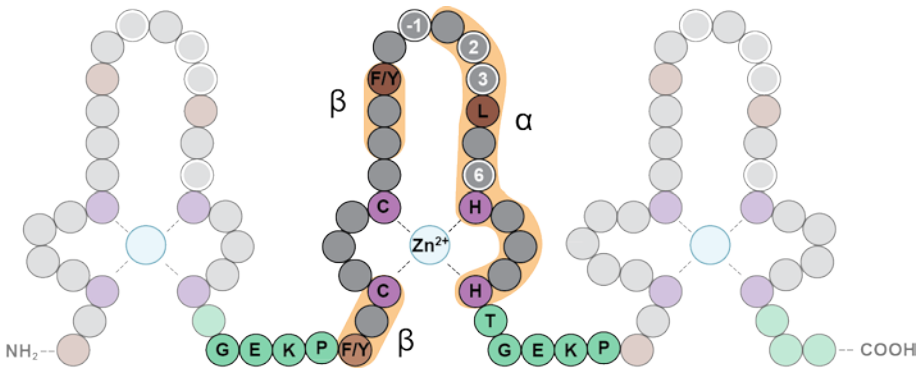


Figure 5.1: Secondary structure of C_2H_2 zinc fingers. Schematic representation of three tandem C_2H_2 zinc fingers. Each zinc finger consists of two β-sheets and one α-helix. Individual zinc ions (Zn^{2+}) interact with paired cysteine (C) and histidine (H) residues to stabilize protein folding. Hydrophobic residues phenylalanine (F) or tyrosine (Y) and leucine (L) that form part of the β-sheets and α-helix respectively, are represented in brown circles. Amino acid residues of any type are represented by grey circles. The highly conserved linker sequence is represented in green circles. Numbers -1, 2, 3 and 6 represent positions (relative to the α-helix) of residues involved in making contact with DNA. This figure was adapted from Stubbs *et al.* (2011).

In each ZIC protein four of the five ZFs (ZF2-ZF5) are well conserved and conform to the C_2H_2 consensus sequence (Figure 5.2). Conversely, ZF1 is significantly longer than other C_2H_2 ZFs (for all ZIC proteins) and shows high divergence across several species (Aruga *et al.*, 2006). In addition all ZIC proteins contain the tandem $CWCH_2$ motif within their ZF1 and ZF2. The function of this motif is unknown, however, based on the position of the tryptophan residues relative to cysteine and histidine in the crystal structure of ZIC3 ZF1-ZF2, the tryptophans likely play a role in stabilisation of their own ZFs and may be involved in the formation of a hydrophobic core between ZF1-ZF2 (Hatayama *et al.*, 2010).

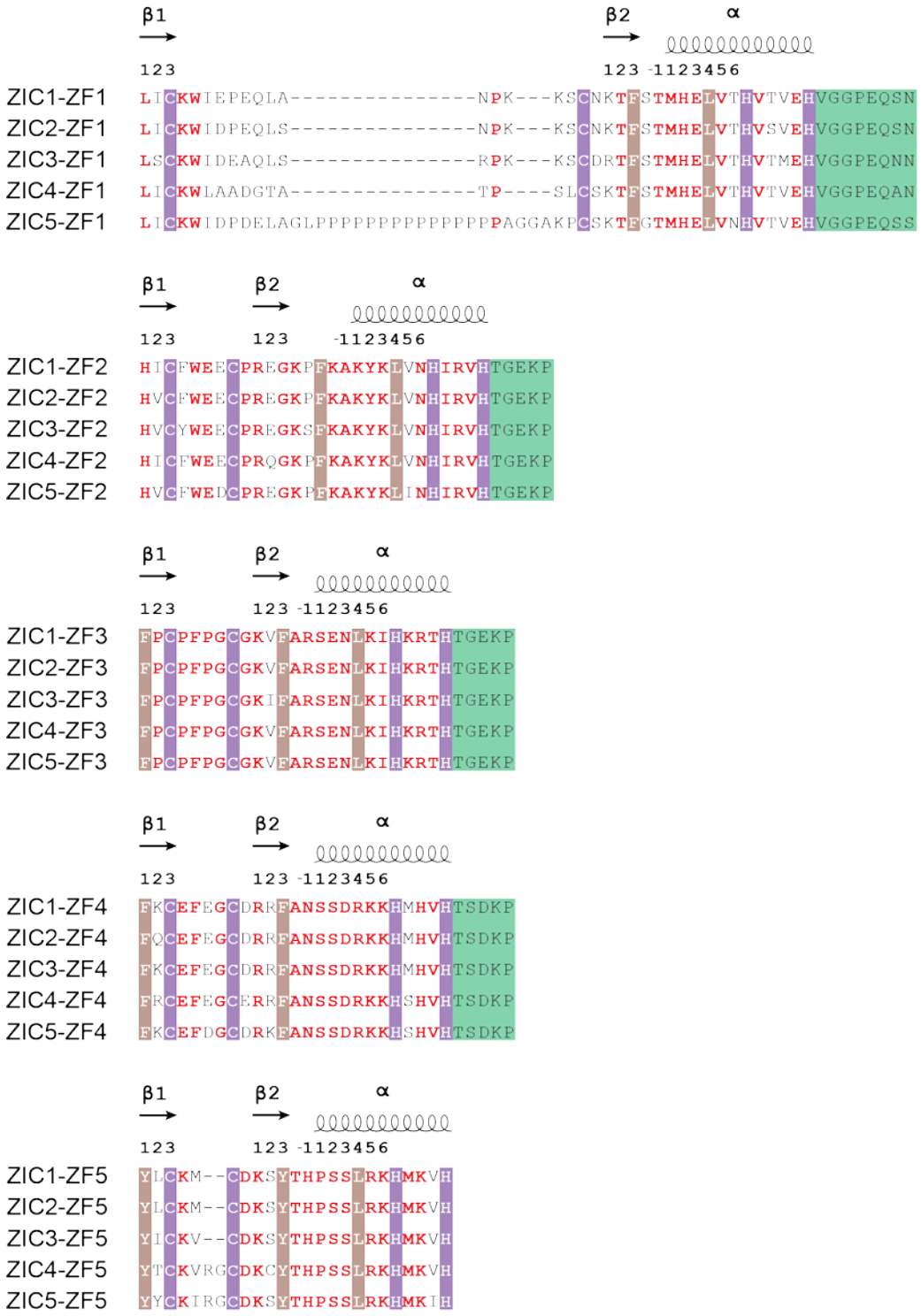


Figure 5.2: ZF motif is highly conserved across ZIC family members. Putative amino acids forming secondary structures (α and β side-chains) are labelled based on their position within the ZFs. Conserved cysteine (C) and histidine (H) residues are in white with a purple background. Residues forming the hydrophobic core are in white with a brown background. Other conserved residues, including the tryptophan (W) residues of ZF1 and ZF2, are in red. Non-conserved residues are in black. Linker sequences are highlighted with a green background.

One common function of C_2H_2 ZFs is DNA-binding with an affinity determined by certain amino acids in the α-helix (Iuchi, 2001, Polozov *et al.*, 2015). In particular, amino acid residues on positions -1, 2, 3 and 6 (relative to the α-helix) confer binding specificities on each C_2H_2 ZF (Pavletich *et al.*, 1991, Suzuki *et al.*, 1994). There are examples of proteins that do not follow this pattern (Fairall *et al.*, 1993), nonetheless mutating amino acids at these key positions changes the ability of ZFs to recognize specific DNA sequences (Choo *et al.*, 1994). Like other proteins containing multiple ZFs (Jacobs, 1992), ZIC proteins have dissimilar residues at positions -1, 2, 3 and 6 between ZFs, however these residues are highly conserved between corresponding fingers across orthologues and paralogues (Aruga *et al.*, 2006). Recognition of target nucleotides occurs in the major groove of DNA, where one ZF recognizes a tri- or tetra-nucleotide sequence. Contacts between protein and DNA are established via sequence specific hydrogen bonds between amino acids side chains and target nucleotides (Michalek *et al.*, 2011). This interaction is stabilized by wrapping of the fingers around the major groove of DNA and C-capping of protein-DNA complex (see below) (Iuchi, 2001). The tandem arrangement of ZFs allows interaction between adjacent fingers and stabilization of the protein-DNA complex (Stubbs *et al.*, 2011).

The Zn^{2+} ion binding ability of C_2H_2 ZFs is also critical for DNA interaction, as in the absence of a Zn^{2+} ion or if amino acids in the conserved C_2H_2 structure are mutated, ZFs lose their ability to form the ββα secondary structure and bind DNA (Miller *et al.*, 1985, Lee *et al.*, 1989, Pavletich *et al.*, 1991, Patel *et al.*, 2007). In addition to the canonical C_2H_2 residues there are three other conserved amino acids: a phenyalanine (F) or tyrosine (Y) residue prior to the first cysteine, a phenyalanine (F) or tyrosine (Y) residue after the second cysteine, and a leucine (L) residue prior to the first histidine. Collectively these form a hydrophobic core and in conjunction with the cysteines and histidines they provide the structural framework of tertiary folding, while other variable residues determine the DNA-binding specificity of each domain (Klug, 2010).

A highly conserved inter-finger “spacer” or H/C link sequence (linker) connects adjacent ZFs and is a vital structural element that contributes to DNA binding by controlling the spacing of fingers

along the DNA site (Wolfe *et al.*, 2000, Stubbs *et al.*, 2011). The linker is usually a five amino acid segment (consensus sequence: TGEKP) between the final histidine (H) of one finger and the first conserved aromatic residue (F/Y) of the next finger. Mutations in the TGEKP linker in transcription factor TFIIIA reduce binding affinity by 10-100 fold in vitro, while in vivo these mutations result in a loss of function (Choo *et al.*, 1993, Clemens *et al.*, 1994, Ryan *et al.*, 1998). Structural analysis has revealed that the lysine (K) residue in the linker makes hydrogen bonds with the DNA-phosphate backbone (Elrod-Erickson *et al.*, 1996). In addition, chemical shift analysis of ZFs has shown that the linker stabilizes the protein-DNA complex by forming hydrogen bonds with the C-terminus (C-capping) of the adjacent α-helix (Laity *et al.*, 2000). ZIC proteins possess four linkers, two of which (between ZF2-ZF3 and ZF3-ZF4) conform to the consensus sequence. The linker between ZF1-ZF2 is significantly longer (8 amino acids in length) than others, with the first 6 amino acids conserved across all ZIC proteins. The linker between ZF4-ZF5 consist of the sequence TSDKP and presumably possesses the ability to establish connections with the phosphate backbone upon formation of ZIC protein-target DNA complex.

5.1.2 Other evolutionary conserved domains

Outside of the ZFD, ZIC proteins contain other evolutionary conserved domains, namely the ZFNC and the ZOC domains (Aruga *et al.*, 2006). The ZFNC domain [consensus sequence: GAF(F/L)RYMRQP-(X_{0-7})-IKQE] is present in all mammalian ZIC proteins and immediately precedes the first ZF. It contains a target site for sumoylation (IKQE), a post-translational modification that influences protein localization, transactivational ability or protein-protein interactions (Rodriguez *et al.*, 2001). The ZOC domain [consensus sequence: (S/T)RDFL-X_3-R] is present only in ZICs 1-3 and is located within the N-terminus of the protein. It has been shown to be involved in protein-protein interactions that eventually influence the transcriptional output of ZIC proteins. Each ZIC protein containing this domain binds to MDFI (I-mfa), which inhibits their translocation to the nucleus (Mizugishi *et al.*, 2004).

5.1.3 Aims

This knowledge regarding putative functional domains of the ZIC proteins can now be examined via the new ZIC-specific transactivation assays developed in Chapter 4. The following questions will be investigated:

- Which regions of the protein are required for transactivation?
- Do individual ZFs contribute differently to the DNA binding and protein-protein binding activities of this domain?

5.2 RESULTS

5.2.1 ZIC3 variant proteins

Domain mapping relies on variant forms of the protein of interest, such as deletions and/or missense variants. Changes to protein sequence can, however, alter fundamental characteristics such as stability or subcellular localization and this information is fundamental to interpretation of functional studies. Each expression construct used in this chapter was therefore introduced into HEK293T cells and the stability and subcellular localization of the encoded ZIC3 variant protein assessed (Figure 5.3) using the same procedures shown in Chapter 3. This is preliminary data and the work was shared amongst laboratory members often assisted by undergraduate students. Each protein was detectable following transfection into HEK293T cells and the majority of the protein was always found in the nucleus, with three nonsense (ZIC3-E155X, ZIC3-C268X and Q292X), two missense (ZIC3- W255G and ZIC3-H286R) mutations and deletion of the N-terminus (ZIC3-Ndel) having the greatest effect on subcellular localization.

Figure 5.3: Subcellular distribution of ZIC3 variants. V5-tagged ZIC3 expression constructs containing the relevant mutations were transfected into HEK293T cells. 24 hours post-transfection cells were prepared for immunofluorescence microscopy and stained with α-V5 (to detect the transfected protein) and α-LaminB1 (to mark the nuclear boundary). Images of transfected cells were captured and subcellular localization of transfected protein quantified using ImageJ. Green and orange regions represent the nuclear and cytoplasmic compartments, respectively. Experiments and data analysis were performed by K. Diamand.

5.2.2 The ZFD and C-terminal regions of ZIC3 are required for transactivation

To determine which regions of the ZIC3 protein contribute to the transactivation ability of the protein, a series of truncated proteins were assayed. Transactivation assays were performed using either the B:*luc2*:Nanog or B:*luc2*:Z3M2:β-globin reporters, since both contain target ZIC DNA-binding sites (see sections 4.2.3 and 4.2.5). The truncated proteins (Figure 5.4A) were obtained based on PTC-containing *ZIC3* variants associated with Heterotaxy in humans (Gebbia *et al.*, 1997, Megarbane *et al.*, 2000) and each was missing the C-terminus and either the whole ZFD (ZIC3-E155X and ZIC3-katun) or parts of it (ZIC3-C268X, ZIC3-Q292X, ZIC3-1507insTT, ZIC3-K408X). Each protein was stably expressed and able to accumulate within the nucleus in HEK293T cells, although not always to the same degree as the wild-type protein (Figure 5.3). The transactivation ability of these mutants (except ZIC3-E155X) was previously assessed (see sections 3.2.7 and 3.2.8) using the B:*luc*$^{+}$:APOE reporter, where ZIC3-C268X, ZIC3-Q292X and ZIC3-1507insTT failed to activate transcription, while ZIC3-K408X appeared to be a hypomorph (Ahmed *et al.*, 2013). When tested using the B:*luc2*:Nanog or B:*luc2*:Z3M2:β-globin reporters, none of the truncated proteins were able to transactivate (Figure 5.4B-C), further confirming the importance of the ZFD and C-terminus.

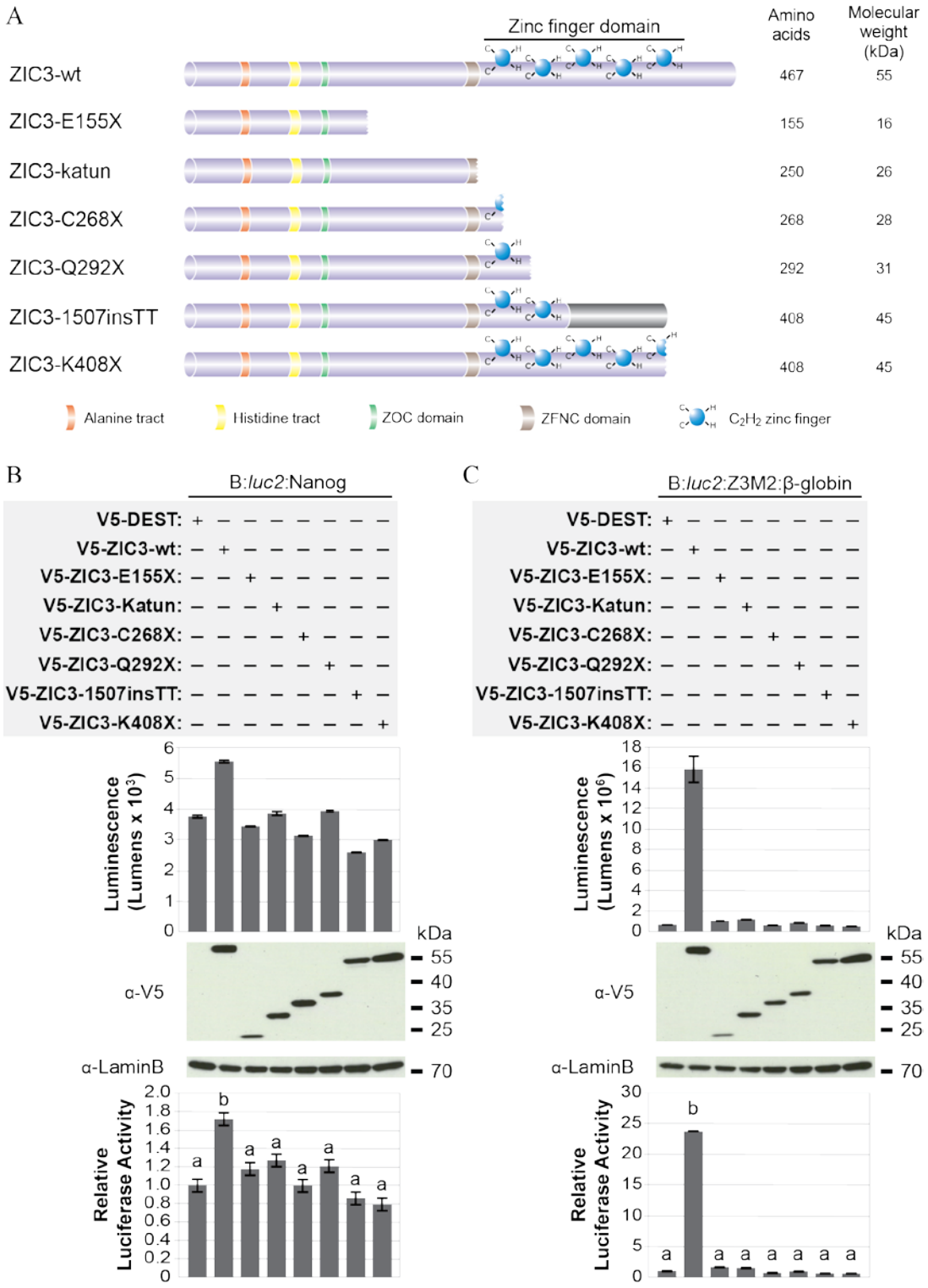


Figure 5.4: ZIC3 PTC-containing mutants are unable to activate transcription. (A) Schematic representation of wild-type ZIC3 and PTC-containing mutants. HEK293T cells were co-transfected with the reporter B:*luc2*:Nanog **(B)** or B:*luc2*:Z3M2:β-globin **(C)** and the expression plasmids shown, with luminescence measured 24 hours post-transfection. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. a and b: $p < 0.01$ ANOVA.

The most complete of all PTC-mutants tested (ZIC3-K408X) truncates at amino acid 408 which lies between the two canonical histidines in ZF5 of ZIC3. To distinguish if the inability of ZIC3-K408X to transactivate reporter sequences is due to a defect in ZF5 or to lack of the C-terminal domain, ZIC3 proteins with a mutation at the canonical cysteine in each ZF were produced and tested in the same assays. Each of these proteins (ZIC3-C268S, ZIC3-C302S, ZIC3-C335S, ZIC3-C365S and ZIC3-C393S) was stably expressed in HEK293T cells and localised to the nucleus (Figure 5.3). Mutation of ZF 1-4 fully ablates transactivation ability, whereas mutation of ZF5 only partially ablates transactivation ability (Figure 5.5). The contrast between the partial loss of transactivation by the ZF5 mutant and the total loss shown by ZIC3-K408X indicates that sequences critical for transactivation are located downstream of the ZFD.

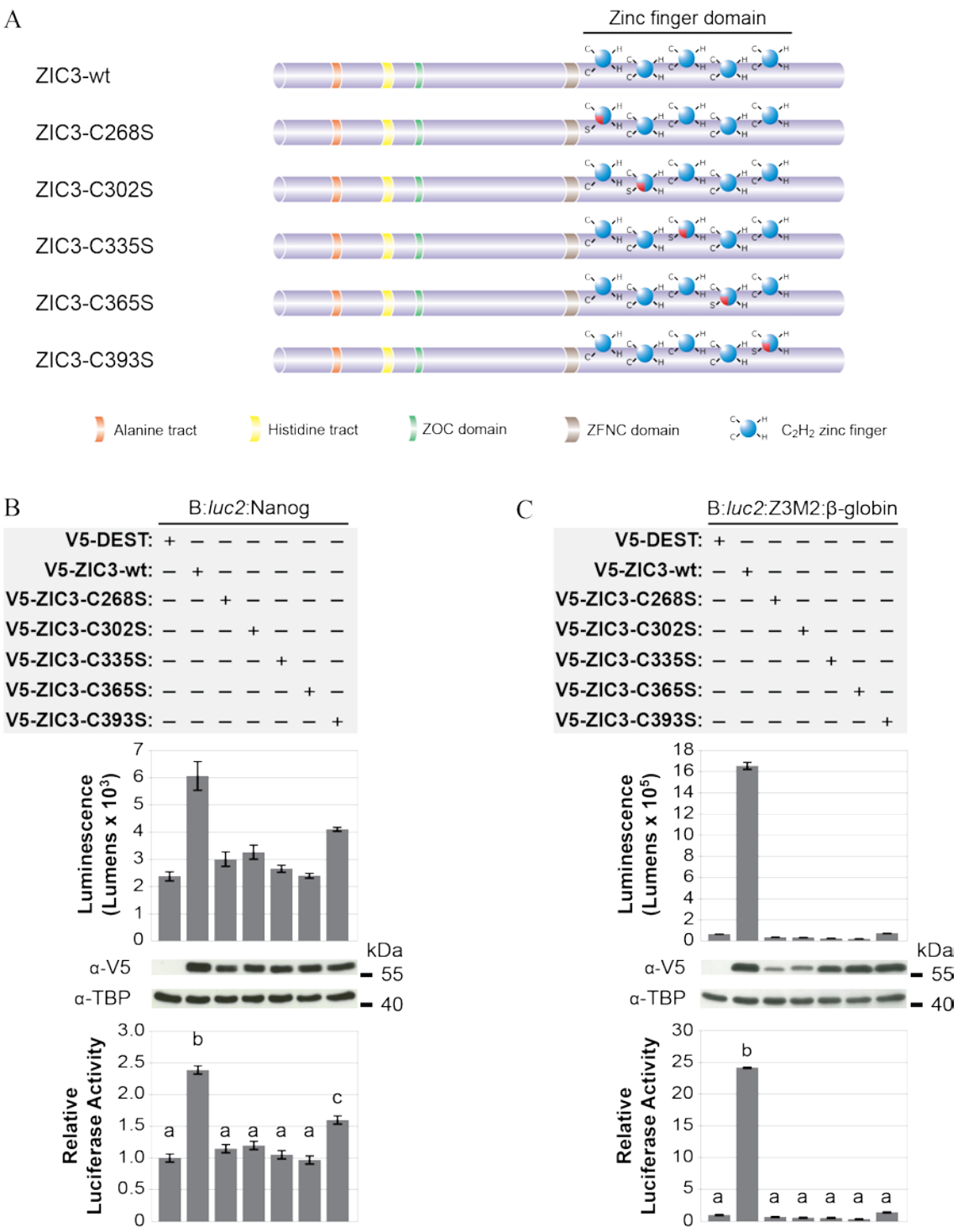


Figure 5.5: Mutations in ZFs of ZIC3 protein abolish the ability to activate transcription. (A) Position of mutations [cysteine (C) converted to serine (S)] on each ZF mutant are shown (in red). HEK293T cells were co-transfected with the reporter B:*luc2*:Nanog **(B)** or B:*luc2*:Z3M2:β-globin **(C)** and the expression plasmids shown, with luminescence measured 24 hours post-transfection. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. a, b and c: $p < 0.01$ ANOVA.

5.2.3 Mutations within ZF-1 abate transactivational ability of ZIC3

As described in section 5.1.1, ZF1 shows highest divergence amongst ZIC proteins, has a significantly longer consensus C_2H_2 ZF sequence and (along with ZF2) contains the $CWCH_2$ motif (Aruga *et al.*, 2006). It has therefore been hypothesized that ZF1 may not be directly involved in DNA-binding at ZIC-responsive elements (ZREs) but may facilitate the DNA binding of the remaining ZFs (Hatayama *et al.*, 2010) and/or participate in protein-protein interactions (Houtmeyers *et al.*, 2013). To determine if this is the case, a series of missense proteins with mutations in the first ZF were assayed. The specific substitutions were again selected based on information from human genetics with each substitution (ZIC3-C253S, ZIC3-W255G and ZIC3-H286R) corresponding to a predicted protein in Heterotaxy patients (Gebbia *et al.*, 1997, Ware *et al.*, 2004, Chhin *et al.*, 2007). Additionally, each of these residues are evolutionary conserved in ZF1 of all ZIC proteins (Figure 5.2; ZIC3-C253 and ZIC3-H286 are part of the C_2H_2 domain, while ZIC3-W255 is the defining feature of the $CWCH_2$ motif) predicting functional significance. Each variant protein was stably expressed in HEK293T cells and localised to the nucleus (Figure 5.3). When tested for transactivation at the *Nanog* promoter, in comparison to ZIC3-wt, ZIC3-C253S exhibited significantly reduced RLA, while ZIC3-W255G and ZIC3-H286R were unable to stimulate transcription from the *Nanog* promoter (Figure 5.6B). To corroborate these findings the mutants were also tested with B:*luc2*:Z3M2:β-globin (Figure 5.6C). Here, all mutants were judged to ablate ZIC3-dependent transactivation. The Western blot analysis used to confirm expression of exogenous proteins in the pool of cells used for the luciferase assay consistently showed that each variant was apparently expressed at lower levels in comparison to ZIC3-wt. These results are in agreement with what has previously been observed in other transactivation assays (Ware *et al.*, 2004, Chhin *et al.*, 2007, D'Alessandro *et al.*, 2013) and overall suggest that ZF1 is essential for transactivation at ZREs.

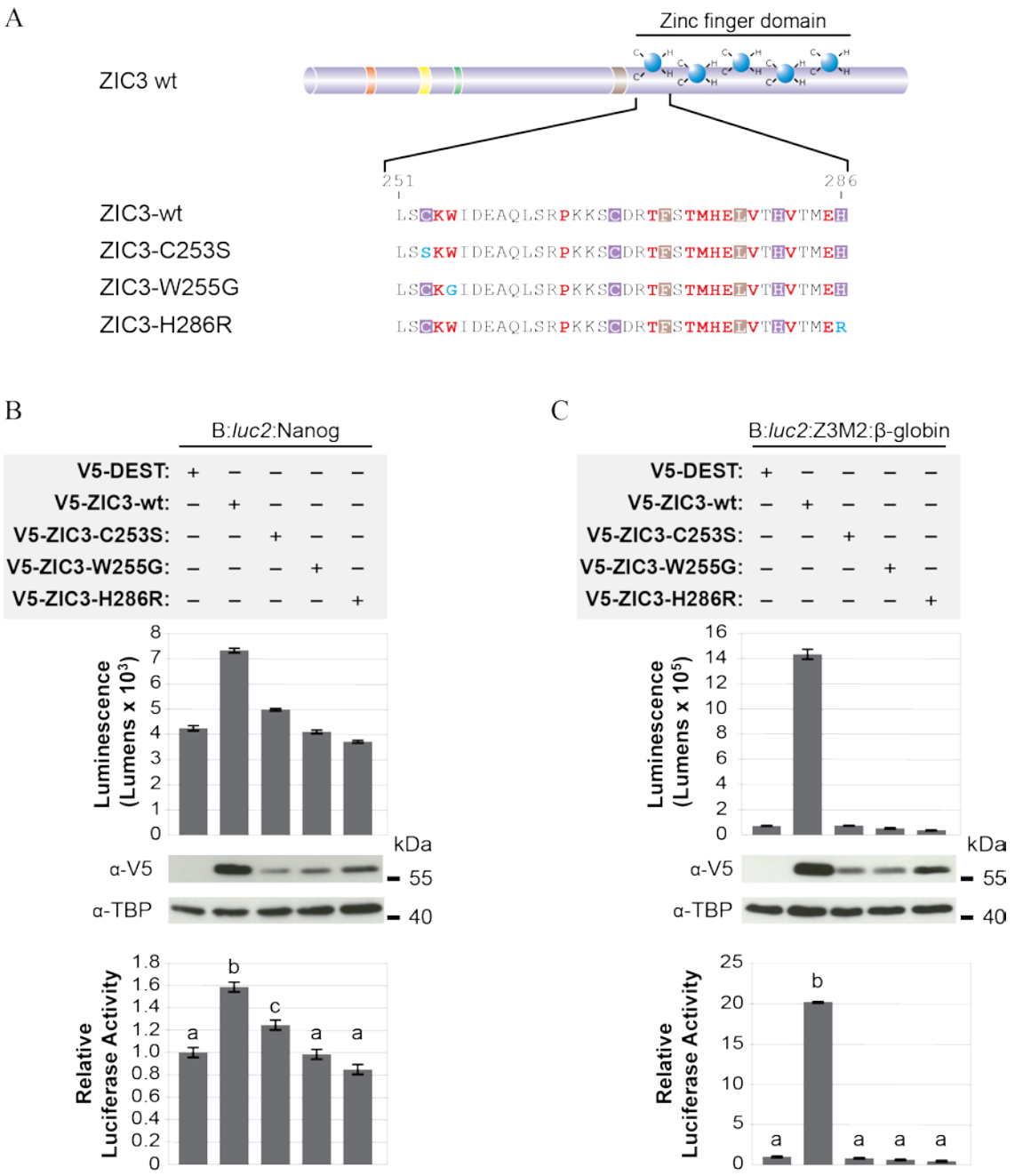


Figure 5.6: Conserved amino acids in ZF1 of ZIC3 are required for transactivation. (A) Amino acid sequences of ZF-1 (position 251-286) of ZIC3-wt, ZIC3-C253S, ZIC3-W255G and ZIC3-H286R are shown. Conserved cysteine (C) and histidine (H) residues are in white with a purple background. Residues forming the hydrophobic core are in white with a brown background. Other conserved residues are in red. Non-conserved residues are in black. Each mutated residue is in light blue. HEK293T cells were co-transfected with the reporter B:*luc2*:Nanog **(B)** or B:*luc2*:Z3M2:β-globin **(C)** and the expression plasmids shown, with luminescence measured 24 hours post-transfection. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. a, b and c: $p < 0.01$ ANOVA.

5.2.4 Multiple domains contribute to an overall repressive effect of the ZIC3 N-terminus.

Phylogenetic analysis of ZIC proteins has revealed two conserved domains outside the ZFD (Aruga *et al.*, 2006): (i) ZFNC domain found in all ZIC proteins, and (ii) ZOC domain present only in ZICs 1-3. While the functional significance of the ZFD for transactivation is clear, the requirement of the ZFNC and ZOC domains is unknown. To assess the importance of these and other N-terminal sequences, one N-terminal deletion and two interstitial deletion mutants were constructed (Figure 5.7A). Each of these proteins was stably expressed in HEK293T cells and localised to the nucleus (Figure 5.3). When the N-terminal deletion (ZIC3-Ndel) was assayed, it was found incapable of eliciting transcription from the Z3M2:β-globin synthetic enhancer, but drove vast over stimulation of the *Nanog* promoter relative to wild-type ZIC3 protein (Figure 5.7B-C). This suggests interaction with other DNA bound proteins at the *Nanog* promoter enable transactivation via ZIC3-Ndel and in this context the N-terminus has a strong transcriptional repression activity. Specific deletion of the N-terminus conserved domains (ZOC and ZFNC) also gave contrasting results at the Z3M2:β-globin synthetic enhancer and genomic *Nanog* reporter constructs (Figure 5.7B-C). In both cases deletion of these domains led to hyper-stimulation of the Z3M2:β-globin synthetic enhancer but were unable to stimulate the *Nanog* promoter to the same extent as wild-type ZIC3 protein. Overall this indicates that a strong repressive effect of these domains at an isolated ZIC3 binding site is converted to a mild stimulatory role at the *Nanog* promoter. These data suggest that regions within the N-terminus that elicit a strong repressor effect lie outside of the ZOC and ZFNC domains.

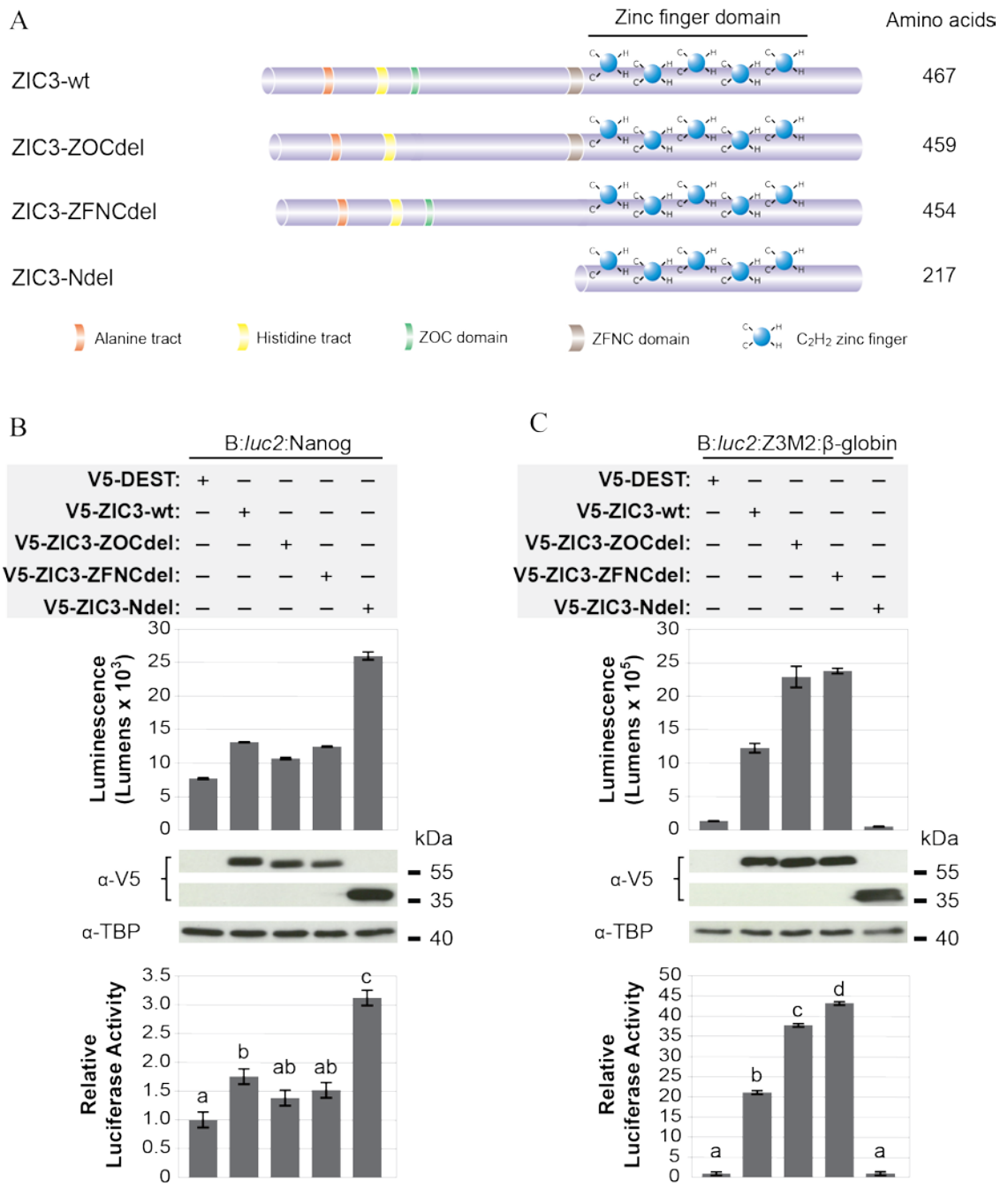


Figure 5.7: N-terminal of ZIC3 is required for transactivation. (A) Schematic representation of the ZIC3 deletion mutants used. ZIC3-ZOCdel is missing the ZOC domain (green band); ZIC3-ZFNCdel is missing the ZFNC domain (brown band); and ZIC3-Ndel is missing the entire N-terminal (amino acids preceding ZF1). HEK293T cells were transfected with the reporter B:*luc2*:Nanog **(B)** or B:*luc2*:Z3M2:β-globin **(C)** and the expression plasmids shown, with luminescence measured 24 hours post-transfection. The top graph shows one representative experiment. Error bars represent SD between three internal repeats. Expression of the transfected proteins was confirmed via Western blotting with α-V5 and the α-TBP blot served as nuclear fraction loading control. The bottom graph shows mean RLA (with reference to V5-DEST), N=3. Error bars represent SEM. a, b, c and d: $p < 0.01$ ANOVA.

5.3 DISCUSSION

As transcriptional regulators, ZIC proteins should be modular in nature. The ZFD is typically responsible for binding to DNA, RNA or proteins (Iuchi, 2001, Gamsjaeger *et al.*, 2007). In the case of ZIC proteins the ZFD has also been found to be required for localization to the nucleus (Bedard *et al.*, 2007, Hatayama *et al.*, 2008). However, for ZIC proteins alone to directly influence transcription, they must have a transcriptional regulatory domain that interacts with or recruits the basal transcriptional machinery to enhance (or reduce) the rate of transcription (Latchman, 1997). In contrast to DNA-binding domains, transcriptional regulatory domains are relatively less conserved, thus are difficult to locate based on sequence analysis.

5.3.1 Transactivation domain of ZIC3 resides in the C-terminus

To identify regions within ZIC3 required for transactivation a classic deletion series experiment was performed using human ZIC3 PTC-containing variants. Their protein products had C-terminal truncations that were either missing the whole ZFD (ZIC3-E155X and ZIC3-katun) or part of it (ZIC3-C268X, ZIC3-Q292X, ZIC3-1507insTT and ZIC3-K408X). None of these variants activated transcription at either the B:*luc2*:Nanog or B:*luc2*:Z3M2:β-globin reporters (Figure 5.4B-C) despite accumulating in the nucleus upon overexpression in HEK293T cells (Figure 5.3). Presumably, the lack of some or all ZFs in these mutants would change the overall tertiary structure of the protein and prohibit DNA-binding as observed in functional assays of other ZF proteins (Gebelein *et al.*, 2001, Kim *et al.*, 2003, Vassen *et al.*, 2005). This assumption could be tested by DNA binding assays with the variant proteins and validated ZREs.

The ZIC3-K408X variant contains most of the ZFD and lacks only the last three residues of ZF-5 and the adjoining C-terminus (Figure 5.8). Previous transactivation studies with this protein categorised it as hypomorphic at either the viral SV40 promoter (Ware *et al.*, 2004) or human *APOE* promoter (Ahmed *et al.*, 2013), but as previously discussed these promoters lack a consensus ZIC3 binding site. Conversely, when tested here with a consensus ZIC3 DNA binding

sequence, the ZIC3-K408X variant was unable to elicit transactivation (Figure 5.4B-C). This finding identified either ZF5, the ZIC3 C-terminus, or both as harbouring sequences critical for transactivation.

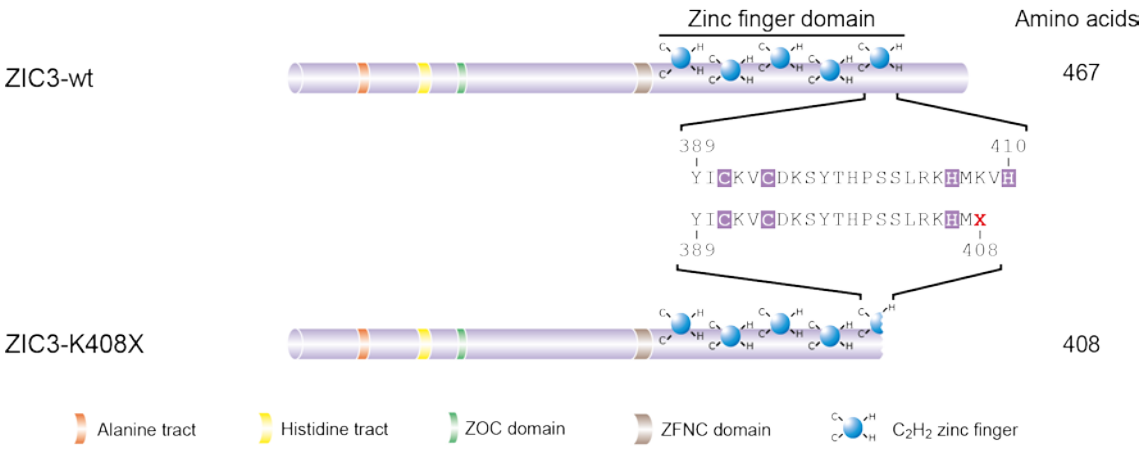


Figure 5.8: Schematic illustration of the K408X mutation. Amino acid sequence of ZIC3 ZF5 (position 389-410) are shown. Conserved cysteine (C) and histidine (H) residues are in white with a purple background. X in red denotes the stop codon (or PTC).

To distinguish these possibilities, a missense mutation was made in ZF5 with the aim of abolishing ZF5 function while leaving the C-terminus intact. The second cysteine (C) of ZF5 was selected for substitution due to its requirement for the ββα structure and Zn^{2+} ion coordination, and serine (S) selected as the substituted residue since it is not a Zn^{2+} ligand (Pace *et al.*, 2014) and the two amino acids differ only in one atom (Figure 5.9). A similar mutation in the second cysteine of ZIC2 ZF4 is known to prevent DNA-binding (Brown *et al.*, 2005) and perturb ZIC2 function in vivo (Elms *et al.*, 2003a). Strikingly, the resulting variant protein (ZIC3-C393S) retains some transactivation ability at both the B:*luc2*:Nanog or B:*luc2*:Z3M2:β-globin reporters (Figure 5.5) indicating that the C-terminus is required for transactivation. It would be interesting to determine whether specific C-terminus deletion of other ZIC proteins also ablates transactivation ability at the ZIC consensus binding sites using the newly generated luciferase assays.

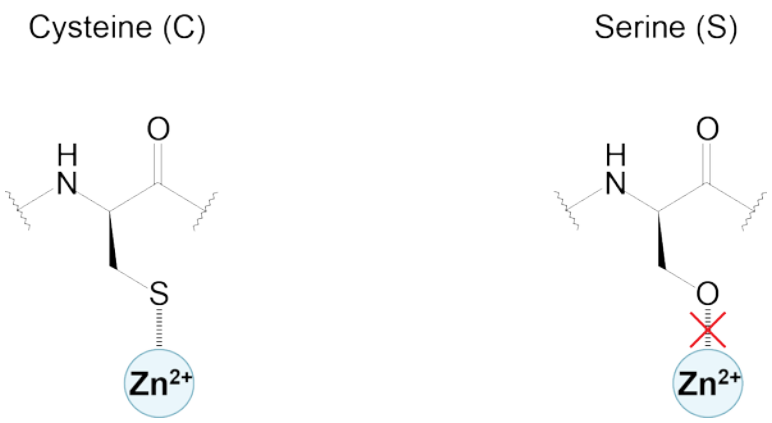


Figure 5.9: Chemical effect of the cysteine to serine conversion on ZFs. Zinc ions (Zn^{2+}) are chelated by cysteine residues using their sulphur atom. Cysteines differ from serine residues by a single atom: sulphur in cysteine is replaced by oxygen in serine. Due to lack of the Zn^{2+} ion-interacting sulphur atom, serine cannot chelate Zn^{2+}.

5.3.2 Each ZF has a role to play in transactivation

Since ZIC proteins contain multiple ZFs, it is possible that specific functions are carried out only by a particular subset of ZFs (Iuchi, 2005), in particular it is thought the ZIC ZF1 and/or ZF2 may have a distinct function from ZF3-5. Using the rationale described above, the second cysteine of each ZF was substituted with a serine residue. This mutation in ZF1-4 rendered the proteins incapable of transactivation (Figure 5.5B-C) indicating that each ZF is individually required for transactivation. To further investigate the function of ZF1, three further missense mutations were assayed. As observed before (Chhin *et al.*, 2007), the in vitro expression of these mutants was reduced in comparison to wild-type ZIC3 (Figure 5.6B-C). This could be attributed to increased degradation of these proteins due to instability in their protein structure as a result of the mutation. Transactivation assays showed that mutants were either hypomorphic (ZIC3-C253S) or nulls (ZIC3-W255G and ZIC3-H286R). Analysis of subcellular localization of these mutants showed that significant amounts of protein can enter the nucleus. The most conceivable explanation for diminished transactivation is perturbed DNA-binding. Each mutation will alter the structural configuration needed for contact with target DNA sites. Comparisons of secondary structure display that the ZFDs of ZIC3-C253S and ZIC3-H286R have

more random coil than wild-type ZIC3 (Hatayama *et al.*, 2008). Coiling of wild-type ZIC3 normally reduces when bound to a Zn^{2+} ion, however in case of ZIC3-C253S and ZIC3-H286R ZFDs random coiling is even higher than wild-type ZIC3 when not bound to a Zn^{2+} ion, indicating that the mutations have impaired coordination of Zn^{2+} ion via the mutant ZFs. CD spectra of ZIC3-W255G showed no difference to wild-type ZIC3, thus the mutation did not have a significant impact on protein secondary structure (Hatayama *et al.*, 2008). Thus far no experiments have been performed to assess the binding affinity of these mutants to target DNA sequences, however based on the crystal structure of GLI (Pavletich *et al.*, 1993) it is possible that ZIC3-ZF1 does not directly bind DNA. The proposed lack of DNA-binding in these mutants, therefore, is most likely due to the inability of their ZF1 to modulate DNA-binding by adjacent ZFs.

5.3.3 The N-terminus contributes to transcription repression

To assess importance of ZIC3 N-terminus for transactivation, the ZOC domain (ZIC3-ZOCdel) or the ZFNC (ZIC3-ZFNCdel) domain or the whole N-terminus (ZIC3-Ndel) were deleted and assayed using B:*luc2*:Nanog and B:*luc2*:Z3M2:β-globin reporters (Figure 5.7B-C). Strikingly, the transactivation ability of the variant proteins depended upon the response element. Specifically, deletion of the entire N-terminus prevented transactivation of the synthetic Z3M2 enhancer but drove hyperstimulation of the *Nanog* promoter, whereas deletion of either the ZOC or ZFNC domains hyper-stimulated the synthetic Z3M2 enhancer but had little effect at the *Nanog* promoter. Two differences between the elements in these reporter constructs that may be responsible for the opposing consequences are: (i) different ZIC3 target sequences (Figure 4.3B and Figure 5.10) and (ii) presence of binding sites for other transcription factors (such as OCT4, SOX2, ESRRB and KLF4) on the *Nanog* promoter (Chen *et al.*, 2008), whereas the Z3M2 synthetic element contains only ZIC3 binding sites.

The ZIC3 target sequence found within the *Nanog* promoter corresponds to that in Motif 1 (Figure 4.7B) for which ZIC3-wt protein has a mild-stimulatory effect when tested in a reporter assay (Figure 4.9B). One possibility is that the N-terminus differentially stabilizes ZIC3 interaction

with different ZIC3 target sequences. This hypothesis could be tested by examining the transactivation ability of the N-terminal deletion constructs at the Z3M1 reporter. An alternative possibility is that the ZIC3 N-terminus interacts with proteins recruited to the *Nanog* promoter (but not the synthetic enhancer). Such an interaction could alter the conformation of the ZIC3-Ndel protein enabling it to participate in transactivation and reveal the strong overall repressive effect of the ZIC3 N-terminus at the *Nanog* promoter. Similarly, context specific activity of both the ZOC and ZFNC domains occur such that they each contribute to transactivation at the *Nanog* promoter. Although it has been shown that ZIC3 activation of the *Nanog* promoter is substantially unaltered when the OCT4 and SOX2 binding sites are mutated (Lim *et al.*, 2010) perhaps ZIC3 interacts with ESRRB, KLF4 or other unidentified proteins to cooperatively activate transcription at the *Nanog* promoter.

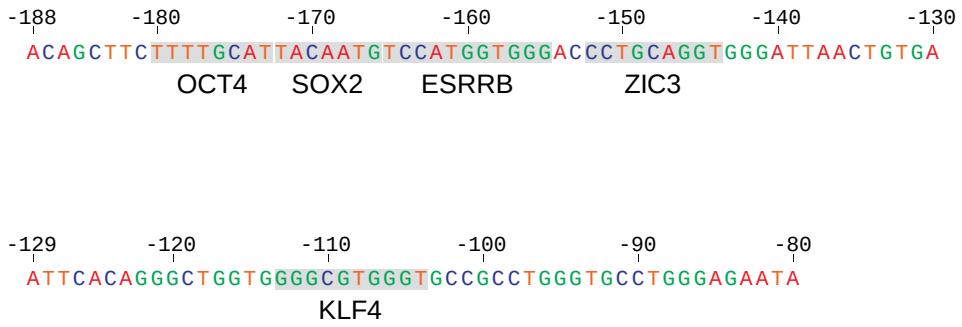


Figure 5.10: Multiple transcription factor binding locus on mouse *Nanog* gene promoter. A 108 bp fragment from mouse genomic DNA (Chr 6: 122707405 - 122707512) was cloned into the reporter plasmid (B:*luc2*). Numbers above the sequence represent distance (in bp) from the major transcription start site as identified by Wu da *et al.* (2005). Bases are coloured: adenine (A) in red; cytosine (C) in blue; guanine (G) in green and thymine (T) in orange. Grey boxes highlight binding sites for transcription factors OCT4, SOX2, ESRRB, ZIC3 and KLF4.

CHAPTER 6

CONCLUDING SUMMARY

ZIC proteins are a family of transcriptional regulators critical for embryonic development. The work presented here aimed to decipher ZIC3 molecular function, details of which remain mostly unknown. This stems from a variety of factors that largely relate to a lack of suitable resources and reagents for ZIC functional studies. At the outset of this project:

i. Most studied mouse alleles were deletion alleles, which reveal the biological role of the protein, but yield no knowledge regarding critical functional domains
ii. Bona-fide ZIC DNA-binding sites were not known prohibiting the production of credible cell-based reporter assays and identification of ZIC3 target genes
iii. In vivo validated protein partners were unknown
iv. Neither antibodies in vivo validated for specificity nor full length recombinant ZIC proteins were available

During the course of this project, novel reagents (the katun mouse allele) and emerging data regarding the ZIC3 DNA-binding site (to make the B:*luc2*:Nanog and B:*luc2*:Z3M2:β-globin reporters) were used to identify ZIC3 functional domains. Characterisation of the katun allele revealed removal of the ZFD and C-terminus renders the protein non-functional. Analysis of several C-terminally truncated mutants verified the importance of ZFD for ZIC protein function and suggested the transactivation domain resides in the C-terminus. The N-terminus appears to have changing roles leading to transcriptional activation or repression, perhaps dependent upon the presence of particular protein partners (co-regulators). The current study significantly advances our understanding of ZIC3 functional domains and suggests several areas for further study.

1. A C-terminal transactivation domain

Typically transactivation domains contain a high percentage of acidic [aspartic acid (D) or glutamic acid (E)], proline (P), glutamine (Q), or serine (S) and threonine (T) amino acid residues (Johnson *et al.*, 1993). Analysis of ZIC3 amino acid sequence (Figure 6.1A) shows that within the N-terminus and the ZFD, residues D, E, P, Q, S and T are sparsely populated. In contrast, the C-terminus (after the ZFD) is dense with D, E, P, Q, S and T residues, with 33 of the 57 residues (~58%) consisting of one of these amino acids. Moreover, the C-terminus of the vertebrate Subgroup A ZICs contains a previously unannotated evolutionary conserved domain (Figure 6.1B). This region (SPAASSGYESSTPP: position 417-431 of human ZIC3) should be tested (via deletion) using the newly developed cell-based reporter assays to assess its functional relevance.

2. N-terminal regions contributing to transactivation

The N-terminus of ZIC3 contains two amino acid repeat regions (Figure 6.1A): (i) an alanine (A)-rich region (position 46-55), which is known to have a transcriptional repression function (Licht *et al.*, 1990), and (ii) a histidine (H)-rich region (position 87-97), which have been shown to target proteins for localization in a nuclear subcompartment, called nuclear speckles (Salichs *et al.*, 2009). Additionally a previously unidentified evolutionary conserved domain (10 amino acids), present in a variety of metazoan species (Figure 6.2), is located in the N-terminus after the ZOC domain. Most likely these regions are not involved in the DNA-binding but may play a role in interacting with protein partners or transactivation. Affinity chromatography or co-immunoprecipitation assays can be done to investigate if deletion of these regions, alters protein-protein interactions. Additionally the cell-based reporter assays (shown here) should be used to assess the requirement of these regions for transactivation.

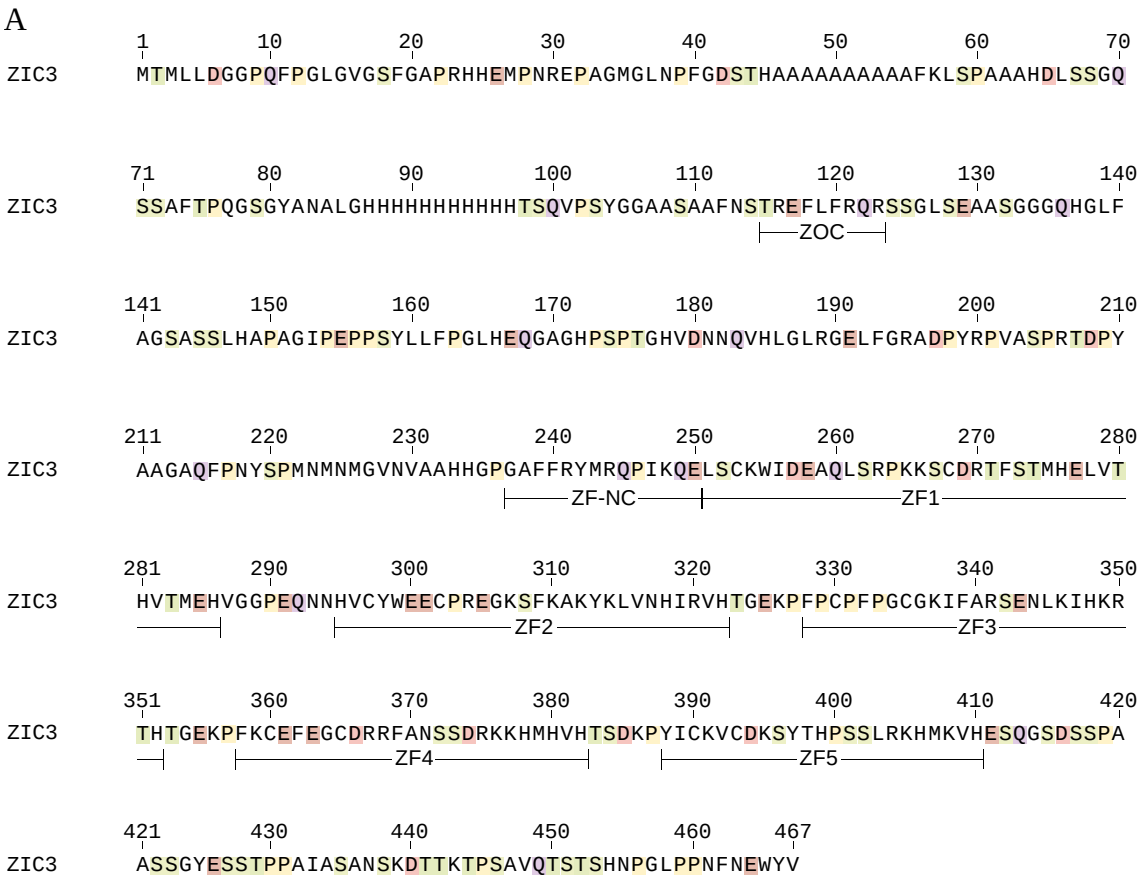


Figure 6.1: C-terminus of ZIC3 contains a high percentage of residues commonly found in transactivation domains. Amino acids known to be involved in transactivation domains of transcription factors are highlighted. Acidic amino acids [aspartic acid (D) and glutamic acid (E)] are highlighted in red. Proline (P) is highlighted in yellow. Glutamine (Q) is highlighted in purple. Serine (S) and threonine (T) are highlighted in green. **(A)** Human ZIC3 amino acid sequence. Bars with ZOC, ZFNC or ZF(1-5) indicate the extent of the ZOC and ZFNC domains and the five C_2H_2 ZFs, respectively. **(B)** Evolutionary conserved region in C-terminus of vertebrate Subgroup A ZIC proteins. *Hs* = *Homo sapiens*; *Mm* = *Mus musculus*; *Xl* = *Xenopus laevis*; *Dr* = *Danio rerio*.

```
 Ap-ZIC      HVIFPGLHDQ
 Sp-ZIC      HVIFPSFHDQ
 Bf-ZIC      HVMFPGFHEQ
 Cj-ZIC      HMLFPGIHDA
Sso-ZIC      HMLFPGIHDA
 Lb-ZIC      HVLFPGLHEP
Hs-ZIC1      HLLFPGLHEQ
Mm-ZIC1      HLLFSGLHEQ
Xl-ZIC1      HLIFPGLHEQ
Dr-ZIC1      HLLFPGLHEQ
Hs-ZIC2      HLLFPGLPEQ
Mm-ZIC2      HLLFSGLPEQ
Xl-ZIC2      HLLFPGIHDQ
Dr-ZIC2      HILFPGIHEQ
Hs-ZIC3      YLLFPGLHEQ
Mm-ZIC3      YLLFPGLHEQ
Xl-ZIC3      HLIFPGLHEQ
Dr-ZIC3      HLLFPGLHDQ
             : :* .: :
```

Figure 6.2: Sequence alignment of N-terminal evolutionary conserved region. Stretch of 10 amino acids (downstream of the ZOC domain), found in a wide variety of metazoan species: *Ap* = *Asterina pectinifera*; *Sp* = *Strongylocentrotus purpuratus*; *Bf* = *Branchiostoma floridae*; *C*j = *Corbicula* sp.; *Sso* = *Spisula solidissima*; *Lb* = *Loligo bleekeri*; *Hs* = *Homo sapiens*; *Mm* = *Mus musculus*; *Xl* = *Xenopus laevis*; *Dr* = *Danio rerio*. Asterick (*) = identical residues. Colon (:) = functionally conserved residues. Period (.) = weak conservation of residue.

3. Further characterization of ZFD variants

The ZF mutants used in this study show either failed or reduced transactivation. This is most likely due to the inability of the mutated ZFs to bind DNA, however this possibility needs to be tested. Proteins with multiple ZFs employ some fingers for DNA-binding while others play a stabilisation role or could be involved in protein-protein interactions (Iuchi, 2005). The particular ZFs of ZIC proteins involved in establishing contacts with DNA, remains to be discovered. Since C_2H_2-type ZFs have a specific ββα structure, Nuclear Magnetic Resonance or X-ray crystallography need to be performed to assess how mutations within particular residues effect the overall tertiary structure. In addition EMSAs and/or ChIP-qPCR can show how mutations in particular locations within ZFs effect DNA-binding. Furthermore the first of four linkers between ZF1 and ZF2 in ZIC proteins is significantly different to the other three and does not correspond to the consensus linker sequence between ZFs in general. Since linker regions within ZF proteins are known to influence DNA-binding affinity (Laity *et al.*, 2000), and suggestions that ZF1 and ZF2 of ZIC proteins are not canonical ZFs, this region warrants an assessment of its requirement for ZIC protein function.

4. Further experiments to link structure with function

i) WNT inhibition: ZIC proteins inhibit β-catenin mediated transcription by binding to the TCF protein via their ZFD (Pourebrahim *et al.*, 2011). Preliminary data shows that each ZF is required for this function, however ZF1-3 are more critical than ZF4 and ZF5. Bimolecular fluorescence complementation assays can be performed to assess how mutations at critical residues within each ZF impact the interaction of ZIC proteins with TCF. Additionally (as suggested above) determining the effect of each mutation on the tertiary structure will aid in understanding the molecular basis of reduced inhibition.

ii) Transcriptional regulatory domains: a common method for identifying transcriptional regulatory domains is the use of the GAL4-UAS fusion reporter system (Lillie *et al.*, 1989, Weston *et al.*, 1989, Pei *et al.*, 1991, Pan *et al.*, 2003, Mizugishi *et al.*, 2004). This system involves creation of fusion proteins, where parts of the protein of interest are fused to the DNA-binding domain of GAL4 (yeast transcriptional activator protein) (Ma, 2014). The fusion construct is co-transfected with a reporter plasmid containing a UAS (upstream activating sequence) followed by reporter CDS (luciferase, GFP, *lacZ*). The GAL4 DNA-binding domain is able to bind to the UAS and if the fused portion of protein of interest contains a transcriptional regulatory domain, changes in reporter expression can be quantitatively assayed. For ZIC3, initial investigations can be done by dividing the *ZIC3* CDS into three subdomains: N-terminus, ZFD, and C-terminus. Once transcriptional activity has been observed from one of these, smaller deletion mutants of 50-75 residues can be created for fine mapping of the transcriptional regulatory domain.

5. Additional mouse mutants

The Australian Phenomics Facility has identified missense mutants in N-Ethyl-N-Nitrosourea (ENU) mutagenized mouse genomes, with three alleles archived at The Australian National University (Table 6.1). Two of these (R→L and Y→C) cause amino acid substitutions with altered biochemical properties and reside in protein regions with putative function. Recovery and

analysis of these alleles may provide additional insight into ZIC3 functional domains. Additionally, the CRISPR-Cas9 genome editing system could be used to delete regions of interest and assessment of their in vivo relevance.

Amino acid substitution	Coordinate (GRCm38)	Amino acid properties	Protein region
R⟶L aa 120	58031539	**R** (arginine): positively charged **L** (leucine): hydrophobic	ZOC domain
V⟶A aa 183	58031728	**V** (valine): hydrophobic **A** (alanine): hydrophobic	N-terminus
Y⟶C aa 424	58034419	**Y** (tyrosine): hydrophobic (aromatic) **C** (cysteine): hydrophobic (non-aromatic)	C-terminus

Table 6.1: Murine *Zic3* missense alleles archived in the Missense Mutation Library. aa: amino acid.

In addition to the role of *ZIC3* in the formation of the L-R axis, it has been shown to maintain pluripotency in embryonic stem cells. Analysis of ZIC3 DNA-binding sites in the mouse (Lim *et al.*, 2010) and zebrafish (Winata *et al.*, 2013) genomes has shown that a fraction of ZIC3 DNA binding sites are associated with promoters, while a significant portion are at distant genomic locations. Additionally binding sites for different transcription factors are present nearby ZIC3 binding sites, indicating that ZIC3 may act in multi-transcription factor complexes. These data suggest that ZIC3 has changing roles and modes of function during embryogenesis. It is hypothesised that in pluripotent stem cells, ZIC3 acts a general transcription factor and directly interacts with the basal transcriptional machinery. In differentiated cells, however, ZIC3 binds to distal elements, perhaps to maintain/complement the differentiated cell state. How this change in preference of binding site occurs remains unknown, however one suggestion is that replacement of maternal transcript with zygotic ones promotes the distal binding (Giraldez *et al.*, 2006). In part, the shift towards binding to enhancers would occur due to changes in chromatin structure and/or presence of particular protein partners. Data shown here indicates

that regions within the N-terminus of ZIC3 influence transcription depending on the spatiotemporal context. These possibilities, nonetheless, need testing at the epigenetic, transcriptomic and proteomic levels. Moreover the embryonic stage when the mutant *Zic3* phenotype first becomes apparent needs to be identified to determine tissues where ZIC3 function is required for prevention of ZIC3-associated congenital defects. A multi-pronged approach that includes discovery of downstream targets of ZIC3 function and analysis of animal models with mutations in specific domains will further aid in understanding the molecular role of ZIC3 during embryogenesis.

Supplementary Information

Appendix 1: Choice of cell line

An important consideration in the design of reporter gene assays is the type of cell line used (Houck *et al.*, 2006). Typically the cell line should reflect the pathological or physiological state wherein the transcription factor normally functions (Yun *et al.*, 2014). Thus using primary cells is the most desirable option, as it will allow translation of in vitro findings into how the transcription factor behaves in vivo. Nonetheless practical considerations limit the use of primary cells, since they are difficult to transfect, slow in proliferation, highly susceptible to contamination and allow limited subdivisions (or passages) (Maurisse *et al.*, 2010). An alternative to primary cells is using established or immortalized cell lines, especially when endogenous factors within the cell are not being assayed. Generally immortalised cell lines are easier to maintain, faster in growth, produce a higher cell yield per flask and allow easy experimental manipulations (such as transfections or chemical treatments) (Freshney, 2010). During the development of the transactivation assay reported here, several established cell lines (C3H/10T1/2, COS-7, HEK293T, NIH3T3, pCHO, and U87) were tested. The primary requirement from these cell lines was high transfection efficiency. With the exception of HEK293T cells, most cells lines displayed relatively low transfection efficiency (Figure S1). In addition, growth and maintenance of HEK293T cells was generally faster and easier than the other cell lines.

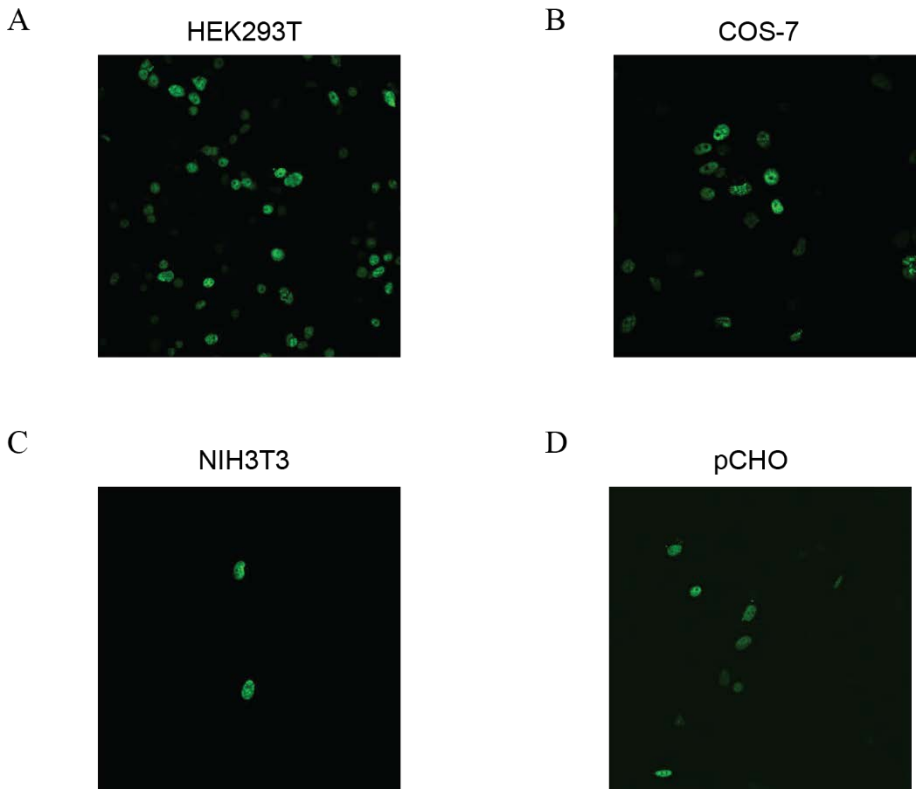


Figure S1: Transfection efficiency across cell lines. HEK293T **(A)**, COS-7 **(B)**, NIH3T3 **(C)** and pCHO **(D)** cells were transfected with EGFP-ZIC3-wt. 24 hours post-transfection cells were viewed under a fluorescence microscope. All images were taken at 100x magnification.

When tested for transactivation of the *APOE* and Z3M synthetic enhancers via wild-type ZIC3 and ZIC3-C365S, none of the cell lines displayed the expected transactivation trends (Figures S2 and S3). This was attributed to endogenous factors within these cells reacting unpredictably to the presence of the synthetic enhancers and/or interfering with ZIC protein function. In contrast, HEK293T cells displayed high levels of luciferase expression via wild-type ZIC3, while transactivation via the DNA-binding mutant (ZIC3-C365S) was negligible. For these reasons and the fact that HEK293T cells have extensively been used for such studies (Braselmann *et al.*, 1993, Lin *et al.*, 2014), these cells were chosen for the transactivation assay reported here. Perhaps once the assay has been established, further experiments can be performed in cell lines that have high endogenous expression of ZIC proteins, such as the cerebellar cell line A40 (Aruga *et al.*, 2000).

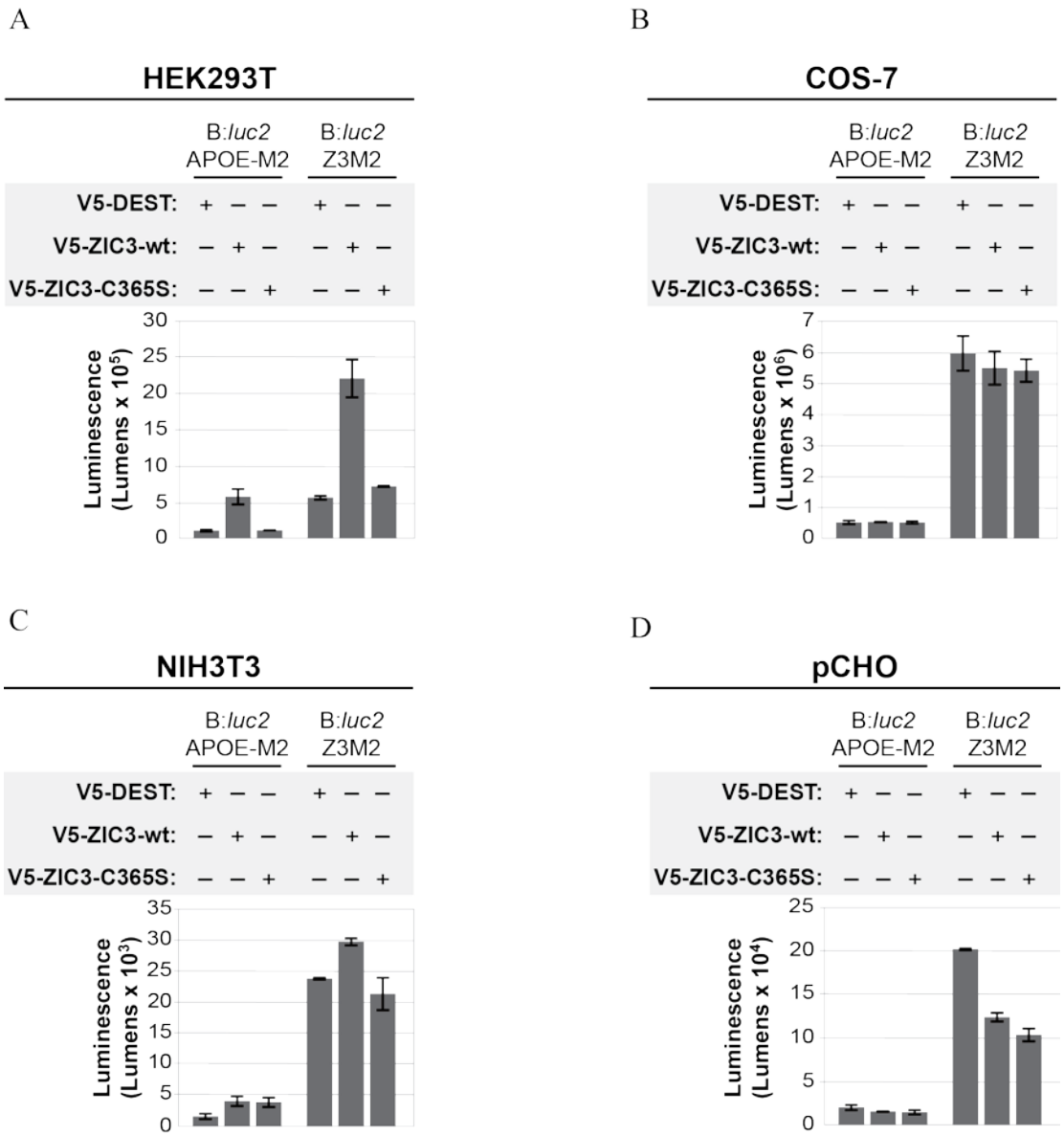


Figure S2: Transactivation via B:*luc2*:APOE-M2 and B:*luc2*:Z3M2 reporters across cell lines. HEK293T **(A)**, COS-7 **(B)**, NIH3T3 **(C)** and pCHO **(D)** cells were co-transfected with the luciferase reporter and expression plasmids shown. 24 hours post-transfection luciferase expression of cells was measured using a luminometer. Graphs display raw values from one experiment. Error bars represent standard deviation between internal repeats. Each experiment was repeated at least 2 times.

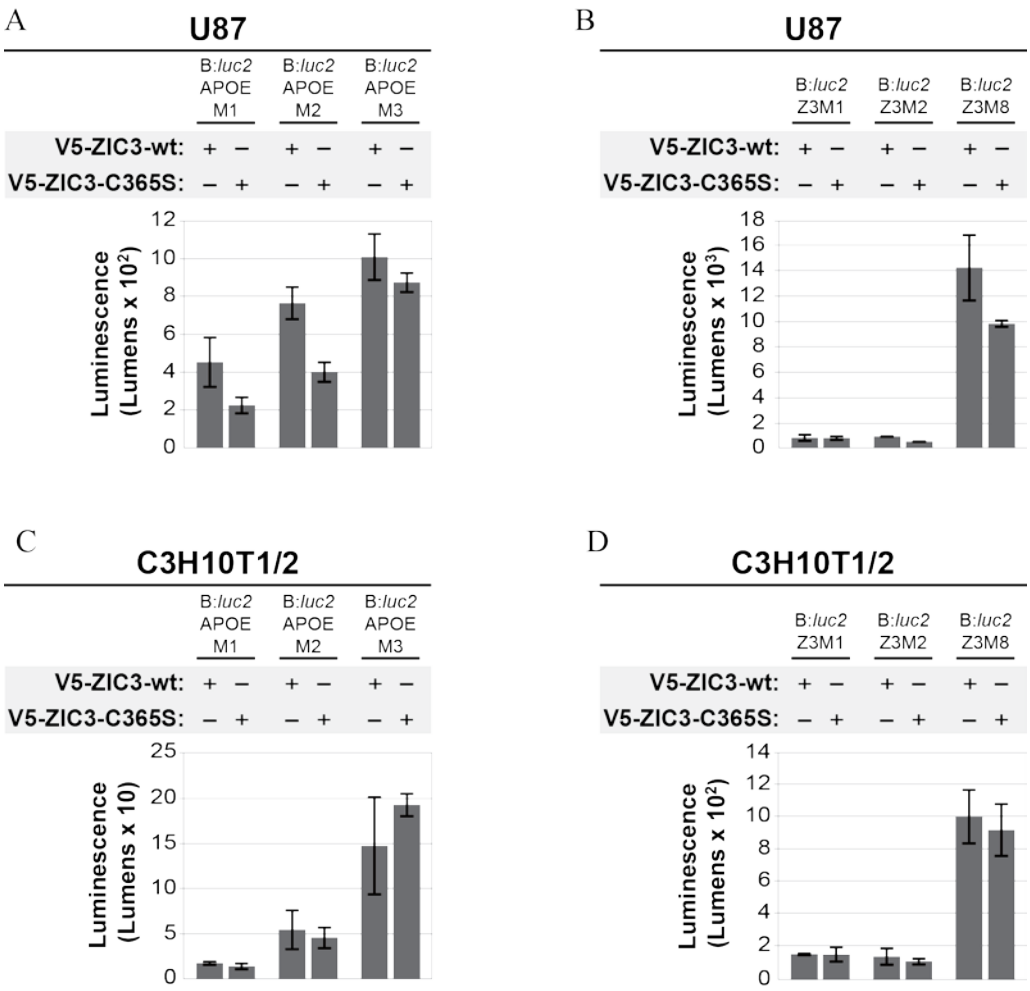


Figure S3: Transactivation via synthetic enhancers in U87 and C3H10T1/2 cells. U87 **(A, B)** and C3H10T1/2 **(C, D)** cells were co-transfected with different combinations of reporter plasmids and expression plasmids as shown. 24 hours post-transfection luciferase expression was measured using a luminometer. Graphs display raw values from one experiment. Error bars represent standard deviation within internal repeats. Each experiment was repeated at least 2 times.

Appendix 2: Genotyping PCR optimisation

Three sets of primers (1 = Ark209-Ark242; 2 = Ark241-Ark242; 3 = Ark241-Ark999) were tested with two different reaction buffers (ThermoPrime ReddyMix 4 and ImmoMix™) on two PCR programs (TD60 and TD65) (Figure S4A-B). From each reaction the expected product was visualized on an agarose gel, however the amplification efficiency was variable. For all primer sets the TD60 program displayed better amplification while there was negligible difference amongst the buffers.

Primer sets 1 and 2 were chosen to test for the HRMA assay. Reactions were performed on the TD60 PCR cycling program using genomic DNA from wild-type and katun animals as template. Melting profiles of samples were analysed using the "Expert Scanning" mode of the LightScanner software. Primer sets 1 and 2 were both able to distinguish between wild-type (grey) and katun DNA (red) (Figure S4C-D). The colours of melting curves were assigned intrinsically by the LightScanner software, based on its ability to distinguish between each genotype. The PCR product of primer set 1 initiated melting at a higher temperature and had a significantly smaller melting temperature range (~90-92°C) in comparison to the same of primer set 2 (~80-84°C). Moreover amplicons of the two samples from primer set 2 displayed a greater difference in melt curves, in comparison to primer set 1.

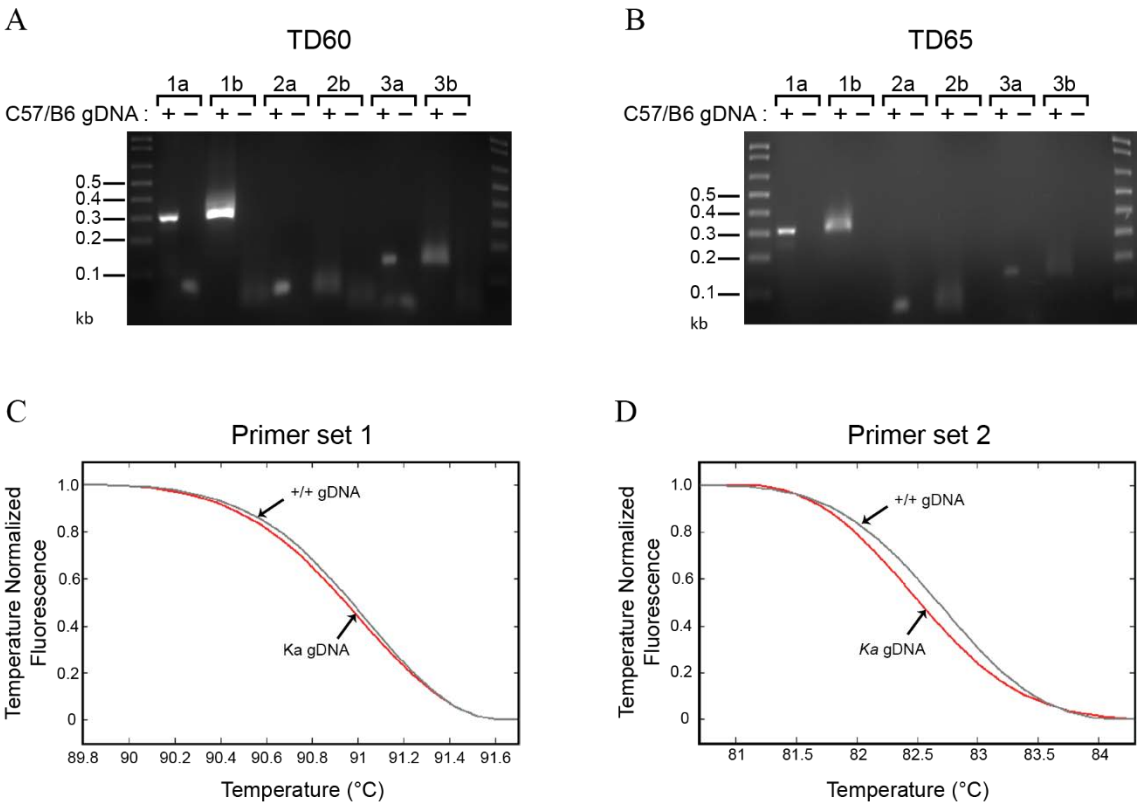


Figure S4: Determining optimal conditions for HRMA assay to genotype katun animals. Diagnostic PCRs performed at two PCR cycling conditions: TD60 **(A)** and TD65 **(B)**, using three different primer combinations (1 = Ark209-Ark242; 2 = Ark241-Ark242; 3 = Ark241-Ark 999) and two reaction buffers (a = ThermoPrime ReddyMix 4; b = ImmoMix™). Each reaction was performed with (+) and without (-) template DNA (C57/B6 genomic DNA). For each reaction the expected product was seen on an agarose gel: 1a and 1b = 304 bp; 2a and 2b = 68 bp; 3a and 3b = 144 bp. Primer sets 1 **(C)** and 2 **(D)** were used in the HRMA assay to distinguish between wild-type (+/+ = grey) and katun (*Ka* = red) genomic DNA. Graphs display shifted melting curves obtained using the "Expert Scanning" mode of the LightScanner software.

REFERENCES

Ahmed, J.N., Ali, R.G., Warr, N., Wilson, H.M., Bellchambers, H.M., Barratt, K.S., Thompson, A.J. & Arkell, R.M. (2013) A murine Zic3 transcript with a premature termination codon evades nonsense-mediated decay during axis formation. *Disease Models & Mechanims,* **6**(3): 755-767.

Ali, R.G., Bellchambers, H.M. & Arkell, R.M. (2012) Zinc fingers of the cerebellum (Zic): Transcription factors and co-factors. *The international journal of biochemistry & cell biology,* **44**(11): 2065-2068.

Almond, B., Hawkins, E., Stecha, P., Garvin, D., Paguio, A., Butler, B., Beck, M., Wood, M. & Wood, K. (2003) A new luminescence: Not your average click beetle. *Promega Notes,* **85**: 11-14.

Altman, D.G. & Bland, J.M. (2005) Standard deviations and standard errors. *BMJ,* **331**(7521): 903.

Amoutzias, G.D., Robertson, D.L., Van de Peer, Y. & Oliver, S.G. (2008) Choose your partners: dimerization in eukaryotic transcription factors. *Trends in Biochemical Sciences,* **33**(5): 220-229.

Aruga, J. (2004) The role of Zic genes in neural development. *Molecular and Cellular Neuroscience,* **26**(2): 205-221.

Aruga, J., Kamiya, A., Takahashi, H., Fujimi, T.J., Shimizu, Y., Ohkawa, K., Yazawa, S., Umesono, Y., Noguchi, H., Shimizu, T., Saitou, N., Mikoshiba, K., Sakaki, Y., Agata, K. & Toyoda, A. (2006) A wide-range phylogenetic analysis of Zic proteins: implications for correlations between protein structure conservation and body plan complexity. *Genomics,* **87**(6): 783-792.

Aruga, J., Minowa, O., Yaginuma, H., Kuno, J., Nagai, T., Noda, T. & Mikoshiba, K. (1998) Mouse Zic1 is involved in cerebellar development. *Journal of Neuroscience,* **18**(1): 284-293.

Aruga, J., Nagai, T., Tokuyama, T., Hayashizaki, Y., Okazaki, Y., Chapman, V.M. & Mikoshiba, K. (1996) The mouse zic gene family. Homologues of the Drosophila pair-rule gene odd-paired. *Journal of Biological Chemistry,* **271**(2): 1043-1047.

Aruga, J., Shimoda, K. & Mikoshiba, K. (2000) A 5' segment of the mouse Zic1 gene contains a region specific enhancer for dorsal hindbrain and spinal cord. *Brain Res Mol Brain Res,* **78**(1-2): 15-25.

Aruga, J., Yokota, N., Hashimoto, M., Furuichi, T., Fukuda, M. & Mikoshiba, K. (1994) A novel zinc finger protein, zic, is involved in neurogenesis, especially in the cell lineage of cerebellar granule cells. *Journal of Neurochemistry,* **63**(5): 1880-1890.

Arya, M., Shergill, I.S., Williamson, M., Gommersall, L., Arya, N. & Patel, H.R. (2005) Basic principles of real-time quantitative PCR. *Expert Review of Molecular Diagnostics,* **5**(2): 209-219.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.F., Coburn, D., Newburger, D.E., Morris, Q., Hughes, T.R. & Bulyk, M.L. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science,* **324**(5935): 1720-1723.

Barratt, K.S., Glanville-Jones, H.C. & Arkell, R.M. (2014) The Zic2 gene directs the formation and function of node cilia to control cardiac situs. *Genesis,* **52**(6): 626-635.

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell,* **116**(2): 281-297.

Bedard, J.E., Haaning, A.M. & Ware, S.M. (2011) Identification of a novel ZIC3 isoform and mutation screening in patients with heterotaxy and congenital heart disease. *PLoS One,* **6**(8): e23755.

Bedard, J.E., Purnell, J.D. & Ware, S.M. (2007) Nuclear import and export signals are essential for proper cellular trafficking and function of ZIC3. *Human Molecular Genetics,* **16**(2): 187-198.

Bedell, M.A., Jenkins, N.A. & Copeland, N.G. (1997) Mouse models of human disease. Part I: techniques and resources for genetic analysis in mice. *Genes Dev,* **11**(1): 1-10.

Behrens, J., von Kries, J.P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R. & Birchmeier, W. (1996) Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature,* **382**(6592): 638-642.

Benedyk, M.J., Mullen, J.R. & DiNardo, S. (1994) Odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in Drosophila embryos. *Genes and Development,* **8**(1): 105-117.

Berg, J.M. (1988) Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc Natl Acad Sci U S A,* **85**(1): 99-102.

Berg, J.M., Tymoczko, J.L. & Stryer, L. (2002). RNA Synthesis and Splicing. *Biochemistry.* New York: W.H. Freeman.

Berwick, D.C., Diss, J.K., Budhram-Mahadeo, V.S. & Latchman, D.S. (2010) A simple technique for the prediction of interacting proteins reveals a direct Brn-3a-androgen receptor interaction. *J Biol Chem,* **285**(20): 15286-15295.

Blainey, P., Krzywinski, M. & Altman, N. (2014) Points of significance: replication. *Nature Methods,* **11**(9): 879-880.

Bogani, D., Warr, N., Elms, P., Davies, J., Tymowska-Lalanne, Z., Goldsworthy, M., Cox, R.D., Keays, D.A., Flint, J., Wilson, V., Nolan, P. & Arkell, R. (2004) New semidominant mutations that affect mouse development. *Genesis,* **40**(2): 109-117.

Boyle, P. & Despres, C. (2010) Dual-function transcription factors and their entourage: unique and unifying themes governing two pathogenesis-related genes. *Plant Signal Behav,* **5**(6): 629-634.

Braselmann, S., Graninger, P. & Busslinger, M. (1993) A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. *Proc Natl Acad Sci U S A,* **90**(5): 1657-1661.

Brivanlou, A.H. & Darnell, J.E., Jr. (2002) Signal transduction and the control of gene expression. *Science,* **295**(5556): 813-818.

Brogna, S. & Wen, J. (2009) Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology,* **16**(2): 107-113.

Bronstein, I., Fortin, J., Stanley, P.E., Stewart, G.S. & Kricka, L.J. (1994) Chemiluminescent and bioluminescent reporter gene assays. *Anal Biochem,* **219**(2): 169-181.

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. *Human Molecular Genetics,* **14**(3): 411-420.

Bullock, P.A. (1997) The initiation of simian virus 40 DNA replication in vitro. *Crit Rev Biochem Mol Biol,* **32**(6): 503-568.

Burt, S.M., Carter, T.J. & Kricka, L.J. (1979) Thermal characteristics of microtitre plates used in immunological assays. *J Immunol Methods,* **31**(3-4): 231-236.

Butler, J.E. & Kadonaga, J.T. (2001) Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev,* **15**(19): 2515-2519.

Carmi, R., Boughman, J.A. & Rosenbaum, K.R. (1992) Human situs determination is probably controlled by several different genes. *Am J Med Genet,* **44**(2): 246-249.

Carrel, T., Purandare, S.M., Harrison, W., Elder, F., Fox, T., Casey, B. & Herman, G.E. (2000) The X-linked mouse mutation Bent tail is associated with a deletion of the Zic3 locus. *Hum Mol Genet,* **9**(13): 1937-1942.

Cartwright, P. & Helin, K. (2000) Nucleocytoplasmic shuttling of transcription factors. *Cellular & Molecular Life Sciences,* **57**(8-9): 1193-1206.

Chabot, A., Shrit, R.A., Blekhman, R. & Gilad, Y. (2007) Using reporter gene assays to identify cis regulatory differences between humans and chimpanzees. *Genetics,* **176**(4): 2069-2076.

Chang, Y.F., Imam, J.S. & Wilkinson, M.F. (2007) The nonsense-mediated decay RNA surveillance pathway. *Annual Review of Biochemistry,* **76**: 51-74.

Chen, C.M., Krohn, J., Bhattacharya, S. & Davies, B. (2011) A comparison of exogenous promoter activity at the ROSA26 locus using a PhiiC31 integrase mediated cassette exchange approach in mouse ES cells. *PLoS One,* **6**(8): e23376.

Chen, X. & Sullivan, P.F. (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J,* **3**(2): 77-96.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., Loh, Y.H., Yeo, H.C., Yeo, Z.X., Narang, V., Govindarajan, K.R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W.K., Clarke, N.D., Wei, C.L. & Ng, H.H. (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell,* **133**(6): 1106-1117.

Chesnoy, S. & Huang, L. (2000) Structure and function of lipid-DNA complexes for gene delivery. *Annu Rev Biophys Biomol Struct,* **29**: 27-47.

Chhin, B., Hatayama, M., Bozon, D., Ogawa, M., Schon, P., Tohmonda, T., Sassolas, F., Aruga, J., Valard, A.G., Chen, S.C. & Bouvagnet, P. (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. *Human Mutation,* **28**(6): 563-570.

Choo, Y. & Klug, A. (1993) A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic Acids Res,* **21**(15): 3341-3346.

Choo, Y. & Klug, A. (1994) Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A,* **91**(23): 11168-11172.

Clemens, K.R., Zhang, P., Liao, X., McBryant, S.J., Wright, P.E. & Gottesfeld, J.M. (1994) Relative contributions of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. *J Mol Biol,* **244**(1): 23-35.

Clevers, H. (2006) Wnt/beta-catenin signaling in development and disease. *Cell,* **127**(3): 469-480.

Cokol, M., Nair, R. & Rost, B. (2000) Finding nuclear localization signals. *EMBO Reports,* **1**(5): 411-415.

Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L. & Myers, R.M. (2006) Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res,* **16**(1): 1-10.

Cowan, J., Tariq, M. & Ware, S.M. (2014) Genetic and functional analyses of ZIC3 variants in congenital heart disease. *Hum Mutat,* **35**(1): 66-75.

Czarnecka, E., Verner, F.L. & Gurley, W.B. (2012) A strategy for building an amplified transcriptional switch to detect bacterial contamination of plants. *Plant Mol Biol,* **78**(1-2): 59-75.

D'Alessandro, L.C., Casey, B. & Siu, V.M. (2011) Situs Inversus Totalis and a Novel ZIC3 Mutation in a Family with X-linked Heterotaxy. *Congenital Heart Disease*.

D'Alessandro, L.C., Latney, B.C., Paluru, P.C. & Goldmuntz, E. (2013) 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**(4): 792-802.

Dalby, B., Cates, S., Harris, A., Ohki, E.C., Tilkins, M.L., Price, P.J. & Ciccarone, V.C. (2004) Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications. *Methods,* **33**(2): 95-103.

Daniels, D.L. & Weis, W.I. (2005) Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nature Structural & Molecular Biology,* **12**(4): 364-371.

Day, D.A. & Tuite, M.F. (1998) Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *Journal of Endocrinology,* **157**(3): 361-371.

de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. & Subramani, S. (1987) Firefly luciferase gene: structure and expression in mammalian cells. *Mol Cell Biol,* **7**(2): 725-737.

de Wet, J.R., Wood, K.V., Helinski, D.R. & DeLuca, M. (1985) Cloning of firefly luciferase cDNA and the expression of active luciferase in Escherichia coli. *Proc Natl Acad Sci U S A,* **82**(23): 7870-7873.

Desterro, J.M., Rodriguez, M.S. & Hay, R.T. (2000) Regulation of transcription factors by protein degradation. *Cellular and Molecular Life Sciences,* **57**(8-9): 1207-1219.

Doppler, S.A., Werner, A., Barz, M., Lahm, H., Deutsch, M.A., Dressen, M., Schiemann, M., Voss, B., Gregoire, S., Kuppusamy, R., Wu, S.M., Lange, R. & Krane, M. (2014) Myeloid zinc finger 1 (Mzf1) differentially modulates murine cardiogenesis by interacting with an Nkx2.5 cardiac enhancer. *PLoS One,* **9**(12): e113775.

Elms, P., Scurry, A., Davies, J., Willoughby, C., Hacker, T., Bogani, D. & Arkell, R. (2004) Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. *Gene Expression Patterns,* **4**(5): 505-511.

Elms, P., Siggers, P., Haines, H., Alexander, E., Greenfield, A. & Arkell, R. (2003a) Characterization of the Kumba allele of Zic2. *Genetical Research,* **81**(3): 234-234.

Elms, P., Siggers, P., Napper, D., Greenfield, A. & Arkell, R. (2003b) Zic2 is required for neural crest formation and hindbrain patterning during mouse development. *Developmental Biology,* **264**(2): 391-406.

Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. (1996) Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure,* **4**(10): 1171-1180.

Everett, L.M. & Crabb, D.W. (1999) Sensitivity of virally-driven luciferase reporter plasmids to members of the steroid/thyroid/retinoid family of nuclear receptors. *J Steroid Biochem Mol Biol,* **70**(4-6): 197-201.

Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T. & Rhodes, D. (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature,* **366**(6454): 483-487.

Fan, F. & Wood, K.V. (2007) Bioluminescent assays for high-throughput screening. *Assay and drug development technologies,* **5**(1): 127-136.

Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., Wenz, M., Northrop, J.P., Ringold, G.M. & Danielsen, M. (1987) Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc Natl Acad Sci U S A,* **84**(21): 7413-7417.

Festing, M.F. (2001) Guidelines for the design and statistical analysis of experiments in papers submitted to ATLA. *Altern Lab Anim,* **29**(4): 427-446.

Foecking, M.K. & Hofstetter, H. (1986) Powerful and versatile enhancer-promoter unit for mammalian expression vectors. *Gene,* **45**(1): 101-105.

Fraga, H. (2008) Firefly luminescence: a historical perspective and recent developments. *Photochem Photobiol Sci,* **7**(2): 146-158.

Freedman, L.P., Luisi, B.F., Korszun, Z.R., Basavappa, R., Sigler, P.B. & Yamamoto, K.R. (1988) The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature,* **334**(6182): 543-546.

Freshney, R.I. (2010). *Culture of animal cells: a manual of basic technique and specialized applications,* Hoboken, New Jersey, Wiley-Blackwell.

Freund, J.N., Domon-Dell, C., Kedinger, M. & Duluc, I. (1998) The Cdx-1 and Cdx-2 homeobox genes in the intestine. *Biochem Cell Biol,* **76**(6): 957-969.

Fujimi, T.J., Hatayama, M. & Aruga, J. (2012) Xenopus Zic3 controls notochord and organizer development through suppression of the Wnt/beta-catenin signaling pathway. *Developmental Biology,* **361**(2): 220-231.

Furushima, K., Murata, T., Kiyonari, H. & Aizawa, S. (2005) Characterization of Opr deficiency in mouse brain: subtle defects in dorsomedial telencephalon and medioventral forebrain. *Developmental Dynamics,* **232**(4): 1056-1061.

Furushima, K., Murata, T., Matsuo, I. & Aizawa, S. (2000) A new murine zinc finger gene, Opr. *Mechanisms of development,* **98**(1-2): 161-164.

Gamsjaeger, R., Liew, C.K., Loughlin, F.E., Crossley, M. & Mackay, J.P. (2007) Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci,* **32**(2): 63-70.

Gebbia, M., Ferrero, G.B., Pilia, G., Bassi, M.T., Aylsworth, A., Penman-Splitt, M., Bird, L.M., Bamforth, J.S., Burn, J., Schlessinger, D., Nelson, D.L. & Casey, B. (1997) X-linked situs abnormalities result from mutations in ZIC3. *Nature Genetics,* **17**(3): 305-308.

Gebelein, B. & Urrutia, R. (2001) Sequence-specific transcriptional repression by KS1, a multiple-zinc-finger-Kruppel-associated box protein. *Mol Cell Biol,* **21**(3): 928-939.

Ghazawi, I., Cutler, S.J., Low, P., Mellick, A.S. & Ralph, S.J. (2005) Inhibitory effects associated with use of modified Photinus pyralis and Renilla reniformis luciferase vectors in dual reporter assays and implications for analysis of ISGs. *J Interferon Cytokine Res,* **25**(2): 92-102.

Giannini, A., Mazor, M., Orme, M., Vivanco, M., Waxman, J. & Kypta, R. (2004) Nuclear export of alpha-catenin: overlap between nuclear export signal sequences and the beta-catenin binding site. *Experimental Cell Research,* **295**(1): 150-160.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J. & Schier, A.F. (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science,* **312**(5770): 75-79.

Glait, C., Ravid, D., Lee, S.W., Liscovitch, M. & Werner, H. (2006) Caveolin-1 controls BRCA1 gene expression and cellular localization in human breast cancer cells. *FEBS Lett,* **580**(22): 5268-5274.

Grimm, S.L. & Nordeen, S.K. (1999) Luciferase reporter gene vectors that lack potential AP-1 sites. *Biotechniques,* **27**(2): 220-222.

Grinberg, I. & Millen, K.J. (2005) The ZIC gene family in development and disease. *Clinical Genetics,* **67**(4): 290-296.

Grinberg, I., Northrup, H., Ardinger, H., Prasad, C., Dobyns, W.B. & Millen, K.J. (2004) Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nature Genetics,* **36**(10): 1053-1055.

Grinblat, Y., Gamse, J., Patel, M. & Sive, H. (1998) Determination of the zebrafish forebrain: induction and patterning. *Development,* **125**(22): 4403-4416.

Gudikote, J.P., Imam, J.S., Garcia, R.F. & Wilkinson, M.F. (2005) RNA splicing promotes translation and RNA surveillance. *Nat Struct Mol Biol,* **12**(9): 801-809.

Han, K.K. & Martinage, A. (1992) Post-translational chemical modification(s) of proteins. *International Journal of Biochemistry,* **24**(1): 19-28.

Harrison, S.M., Dunwoodie, S.L., Arkell, R.M., Lehrach, H. & Beddington, R.S. (1995) Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo. *Development,* **121**(8): 2479-2489.

Hatayama, M. & Aruga, J. (2010) Characterization of the tandem CWCH2 sequence motif: a hallmark of inter-zinc finger interactions. *BMC Evolutionary Biology,* **10**: 53.

Hatayama, M., Ishiguro, A., Iwayama, Y., Takashima, N., Sakoori, K., Toyota, T., Nozaki, Y., Odaka, Y.S., Yamada, K., Yoshikawa, T. & Aruga, J. (2011) Zic2 hypomorphic mutant mice as a schizophrenia model and ZIC2 mutations identified in schizophrenia patients. *Sci Rep,* **1**: 16.

Hatayama, M., Tomizawa, T., Sakai-Kato, K., Bouvagnet, P., Kose, S., Imamoto, N., Yokoyama, S., Utsunomiya-Tate, N., Mikoshiba, K., Kigawa, T. & Aruga, J. (2008) Functional and structural basis of the nuclear localization signal in the ZIC3 zinc finger domain. *Human Molecular Genetics,* **17**(22): 3459-3473.

Hawley-Nelson, P., Ciccarone, V. & Moore, M.L. (2008) Transfection of cultured eukaryotic cells using cationic lipid reagents. *Curr Protoc Mol Biol,* **Chapter 9**: Unit 9 4.

Heard, E., Clerc, P. & Avner, P. (1997) X-chromosome inactivation in mammals. *Annu Rev Genet,* **31**: 571-610.

Herman, G.E. & El-Hodiri, H.M. (2002) The role of ZIC3 in vertebrate development. *Cytogenetic and Genome Research,* **99**(1-4): 229-235.

Hershko, A. & Ciechanover, A. (1998) The ubiquitin system. *Annual Review of Biochemistry,* **67**: 425-479.

Herskowitz, I. (1987) Functional inactivation of genes by dominant negative mutations. *Nature,* **329**(6136): 219-222.

Hicks, G.R. & Raikhel, N.V. (1995) Protein import into the nucleus: an integrated view. *Annual Review of Cell & Developmental Biology,* **11**: 155-188.

Himeda, C.L., Barro, M.V. & Emerson, C.P., Jr. (2013) Pax3 synergizes with Gli2 and Zic1 in transactivating the Myf5 epaxial somite enhancer. *Dev Biol,* **383**(1): 7-14.

Houck, K.A., Bocchinfuso, W.P., Dowless, M.S. & Borchert, K.M. (2006). Reporter Gene Assays for Drug Discovery. *In:* MINOR, L. K. (ed.) *Handbook of Assay Development in Drug Discovery.* CRC Press.

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**(20): 3791-3811.

Ibrahim, N.M., Marinovic, A.C., Price, S.R., Young, L.G. & Frohlich, O. (2000) Pitfall of an internal control plasmid: response of Renilla luciferase (pRL-TK) plasmid to dihydrotestosterone and dexamethasone. *Biotechniques,* **29**(4): 782-784.

Iuchi, S. (2001) Three classes of C2H2 zinc finger proteins. *Cellular and Molecular Life Sciences,* **58**(4): 625-635.

Iuchi, S. (2005). Zinc Finger Proteins. *In:* KULDELL, N. (ed.) *From Atomic Content to Cellular Function.* New York: Kluwer Academic/Plenum Publishers.

Jacobs, G.H. (1992) Determination of the base recognition positions of zinc fingers from sequence analysis. *EMBO J,* **11**(12): 4507-4517.

Jacquet, M.A., Ehrlich, R. & Reiss, C. (1989) In vivo gene expression directed by synthetic promoter constructions restricted to the -10 and -35 consensus hexamers of E. coli. *Nucleic Acids Res,* **17**(8): 2933-2945.

Jans, D.A., Xiao, C.Y. & Lam, M.H. (2000) Nuclear targeting signal recognition: a key control point in nuclear transport? *Bioessays,* **22**(6): 532-544.

Jiang, T., Xing, B. & Rao, J. (2008) Recent developments of biological reporter technology for detecting gene expression. *Biotechnology and Genetic Engineering Reviews,* **25**: 41-75.

Jiang, Z., Zhu, L., Hu, L., Slesnick, T.C., Pautler, R.G., Justice, M.J. & Belmont, J.W. (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**(5): 879-889.

Jimenez-Sanchez, G., Childs, B. & Valle, D. (2001) Human disease genes. *Nature,* **409**(6822): 853-855.

Johnson, H.M., Torres, B.A., Green, M.M., Szente, B.E., Siler, K.I., Larkin, J., 3rd & Subramaniam, P.S. (1998) Cytokine-receptor complexes as chaperones for nuclear translocation of signal transducers. *Biochemical & Biophysical Research Communication,* **244**(3): 607-614.

Johnson, P.F., Sterneck, E. & Williams, S.C. (1993) Activation domains of transcriptional regulatory proteins. *The Journal of Nutritional Biochemistry,* **4**(7): 386-398.

Kadonaga, J.T. (1998) Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell,* **92**(3): 307-313.

Kearns-Jonker, M. (2006). *Congenital Heart Disease: Molecular Diagnostics*, Humana Press

Kiefer, S.M., Ohlemiller, K.K., Yang, J., McDill, B.W., Kohlhase, J. & Rauchman, M. (2003) Expression of a truncated Sall1 transcriptional repressor is responsible for Townes-Brocks syndrome birth defects. *Hum Mol Genet,* **12**(17): 2221-2227.

Kim, Y.S., Nakanishi, G., Lewandoski, M. & Jetten, A.M. (2003) GLIS3, a novel member of the GLIS subfamily of Kruppel-like zinc finger proteins with repressor and activation functions. *Nucleic Acids Res,* **31**(19): 5513-5525.

Kitaguchi, T., Nagai, T., Nakata, K., Aruga, J. & Mikoshiba, K. (2000) Zic3 is involved in the left-right specification of the Xenopus embryo. *Development,* **127**(22): 4787-4795.

Klemm, J.D., Schreiber, S.L. & Crabtree, G.R. (1998) Dimerization as a regulatory mechanism in signal transduction. *Annual Review of Immunology,* **16**: 569-592.

Kloosterman, W.P. & Plasterk, R.H. (2006) The diverse functions of microRNAs in animal development and disease. *Developmental Cell,* **11**(4): 441-450.

Klug, A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem,* **79**: 213-231.

Klug, A. & Schwabe, J.W. (1995) Protein motifs 5. Zinc fingers. *FASEB J,* **9**(8): 597-604.

Koebernick, K. & Pieler, T. (2002) Gli-type zinc finger proteins as bipotential transducers of Hedgehog signaling. *Differentiation,* **70**(2-3): 69-76.

Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B. & Clevers, H. (1997) Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science,* **275**(5307): 1784-1787.

Koyabu, Y., Nakata, K., Mizugishi, K., Aruga, J. & Mikoshiba, K. (2001) Physical and functional interactions between Zic and Gli proteins. *J Biol Chem,* **276**(10): 6889-6892.

Kricka, L.J., Carter, T.J., Burt, S.M., Kennedy, J.H., Holder, R.L., Halliday, M.I., Telford, M.E. & Wisdom, G.B. (1980) Variability in the adsorption properties of microtitre plates used as solid supports in enzyme immunoassay. *Clin Chem,* **26**(6): 741-744.

Kuehl, K.S. & Loffredo, C. (2002) Risk factors for heart disease associated with abnormal sidedness. *Teratology,* **66**(5): 242-248.

Kurtzman, A.L. & Schechter, N. (2001) Ubc9 interacts with a nuclear localization signal and mediates nuclear localization of the paired-like homeobox protein Vsx-1 independent of SUMO-1 modification. *Proceedings of the National Academt of Sciences USA,* **98**(10): 5602-5607.

Lacks, S. & Greenberg, B. (1975) A deoxyribonuclease of Diplococcus pneumoniae specific for methylated DNA. *J Biol Chem,* **250**(11): 4060-4066.

Laity, J.H., Dyson, H.J. & Wright, P.E. (2000) DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. *J Mol Biol,* **295**(4): 719-727.

Landolin, J.M., Johnson, D.S., Trinklein, N.D., Aldred, S.F., Medina, C., Shulha, H., Weng, Z. & Myers, R.M. (2010) Sequence features that drive human promoter function and tissue specificity. *Genome Res,* **20**(7): 890-898.

Landy, A. (1989) Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem,* **58**: 913-949.

Latchman, D.S. (1990) Eukaryotic transcription factors. *The Biochemical Journal,* **270**(2): 281-289.

Latchman, D.S. (1997) Transcription factors: an overview. *International Journal of Biochemistry & Cell Biology,* **29**(12): 1305-1312.

Latchman, D.S. (2008). *Eukaryotic Transcription Factors,* Amsterdam, Elsevier.

Lay, M.J. & Wittwer, C.T. (1997) Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR. *Clin Chem,* **43**(12): 2262-2267.

Layden, M.J., Meyer, N.P., Pang, K., Seaver, E.C. & Martindale, M.Q. (2010) Expression and phylogenetic analysis of the zic gene family in the evolution and development of metazoans. *EvoDevo,* **1**(1): 12.

Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. & Wright, P.E. (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science,* **245**(4918): 635-637.

Li, X., Eastman, E.M., Schwartz, R.J. & Draghia-Akli, R. (1999) Synthetic muscle promoters: activities exceeding naturally occurring regulatory sequences. *Nat Biotechnol,* **17**(3): 241-245.

Li, Y.C., Hayes, S. & Young, A.P. (1994) Transactivation of the 'promoterless' luciferase-encoding vectors pXP1 and pXP2 by C/EBP alpha. *Gene,* **138**(1-2): 257-258.

Licht, J.D., Grossel, M.J., Figge, J. & Hansen, U.M. (1990) Drosophila Kruppel protein is a transcriptional repressor. *Nature,* **346**(6279): 76-79.

Lillie, J.W. & Green, M.R. (1989) Transcription activation by the adenovirus E1a protein. *Nature,* **338**(6210): 39-44.

Lim, L.S., Hong, F.H., Kunarso, G. & Stanton, L.W. (2010) The pluripotency regulator Zic3 is a direct activator of the Nanog promoter in ESCs. *Stem Cells,* **28**(11): 1961-1969.

Lim, L.S., Loh, Y.H., Zhang, W., Li, Y., Chen, X., Wang, Y., Bakre, M., Ng, H.H. & Stanton, L.W. (2007) Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Molecular Biology of the Cell,* **18**(4): 1348-1358.

Lin, X., Antalffy, B., Kang, D., Orr, H.T. & Zoghbi, H.Y. (2000) Polyglutamine expansion down-regulates specific neuronal genes before pathologic changes in SCA1. *Nat Neurosci,* **3**(2): 157-163.

Lin, Y.C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van de Peer, Y., Tavernier, J. & Callewaert, N. (2014) Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun,* **5**: 4767.

Livak, K.J. (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genetic Analysis,* **14**(5-6): 143-149.

Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H. & Matsudaira, P. (2007). *Molecular Cell Biology,* New York, W. H. Freeman and Company.

Lundin, A. (2014) Optimization of the firefly luciferase reaction for analytical purposes. *Advances in Biochemical Enginerring/Biotechnology,* **145**: 31-62.

Ma, J. (2014). Promoter Fusions to Study Gene Expression. *eLS.* Chichester: John Wiley & Sons, Ltd.

Maddox, C.B., Rasmussen, L. & White, E.L. (2008) Adapting Cell-Based Assays to the High Throughput Screening Platform: Problems Encountered and Lessons Learned. *JALA Charlottesv Va,* **13**(3): 168-173.

Mader, S. & White, J.H. (1993) A steroid-inducible promoter for the controlled overexpression of cloned genes in eukaryotic cells. *Proc Natl Acad Sci U S A,* **90**(12): 5603-5607.

Malakoff, D. (2000) The rise of the mouse, biomedicine's model mammal. *Science,* **288**(5464): 248-253.

Malo, N., Hanley, J.A., Cerquozzi, S., Pelletier, J. & Nadon, R. (2006) Statistical practice in high-throughput screening data analysis. *Nature Biotechnology,* **24**(2): 167-175.

Mantovani, R., Malgaretti, N., Nicolis, S., Giglioni, B., Comi, P., Cappellini, N., Bertero, M.T., Caligaris-Cappio, F. & Ottolenghi, S. (1988) An erythroid specific nuclear factor binding to the proximal CACCC box of the beta-globin gene promoter. *Nucleic Acids Res,* **16**(10): 4299-4313.

Maquat, L.E. & Li, X. (2001) Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. *RNA,* **7**(3): 445-456.

Matsugi, T., Morishita, K. & Ihle, J.N. (1990) Identification, nuclear localization, and DNA-binding activity of the zinc finger protein encoded by the Evi-1 myeloid transforming gene. *Mol Cell Biol,* **10**(3): 1259-1264.

Maurisse, R., De Semir, D., Emamekhoo, H., Bedayat, B., Abdolmohammadi, A., Parsi, H. & Gruenert, D.C. (2010) Comparative transfection of DNA into primary and transformed mammalian cells from different lineages. *BMC Biotechnol,* **10**(9).

McGuigan, F.E. & Ralston, S.H. (2002) Single nucleotide polymorphism detection: allelic discrimination using TaqMan. *Psychiatr Genet,* **12**(3): 133-136.

McKillup, S. (2011). *Statistics explained: an introductory guide for life scientists*, Cambridge University Press.

Megarbane, A., Salem, N., Stephan, E., Ashoush, R., Lenoir, D., Delague, V., Kassab, R., Loiselet, J. & Bouvagnet, P. (2000) X-linked transposition of the great arteries and incomplete penetrance among males with a nonsense mutation in ZIC3. *Eur J Hum Genet,* **8**(9): 704-708.

Melchior, F., Schergaut, M. & Pichler, A. (2003) SUMO: ligases, isopeptidases and nuclear pores. *Trends in Biochemical Sciences,* **28**(11): 612-618.

Michalek, J.L., Besold, A.N. & Michel, S.L. (2011) Cysteine and histidine shuffling: mixing and matching cysteine and histidine residues in zinc finger proteins to afford different folds and function. *Dalton Trans,* **40**(47): 12619-12632.

Miller, J., McLachlan, A.D. & Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J,* **4**(6): 1609-1614.

Miraglia, L.J., King, F.J. & Damoiseaux, R. (2011) Seeing the light: luminescent reporter gene assays. *Comb Chem High Throughput Screen,* **14**(8): 648-657.

Mitchell, P.J. & Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science,* **245**(4916): 371-378.

Miyamoto, H., Rahman, M., Takatera, H., Kang, H.Y., Yeh, S., Chang, H.C., Nishimura, K., Fujimoto, N. & Chang, C. (2002) A dominant-negative mutant of androgen receptor coregulator ARA54 inhibits androgen receptor-mediated prostate cancer growth. *J Biol Chem,* **277**(7): 4609-4617.

Mizugishi, K., Aruga, J., Nakata, K. & Mikoshiba, K. (2001) Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI proteins. *Journal of Biological Chemistry,* **276**(3): 2180-2188.

Mizugishi, K., Hatayama, M., Tohmonda, T., Ogawa, M., Inoue, T., Mikoshiba, K. & Aruga, J. (2004) Myogenic repressor I-mfa interferes with the function of Zic family proteins. *Biochemical and Biophysical Research Communications,* **320**(1): 233-240.

Muratani, M. & Tansey, W.P. (2003) How the ubiquitin-proteasome system controls transcription. *Nature Reviews Molecular Cell Biology,* **4**(3): 192-201.

Nagai, T., Aruga, J., Takada, S., Gunther, T., Sporle, R., Schughart, K. & Mikoshiba, K. (1997) The expression of the mouse Zic1, Zic2, and Zic3 gene suggests an essential role for Zic genes in body pattern formation. *Developmetal Biology,* **182**(2): 299-313.

Nagy, E. & Maquat, L.E. (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends in Biochemical Sciences,* **23**(6): 198-199.

Nakamura, T., Saito, D., Kawasumi, A., Shinohara, K., Asai, Y., Takaoka, K., Dong, F., Takamatsu, A., Belo, J.A., Mochizuki, A. & Hamada, H. (2012) Fluid flow and interlinked feedback loops establish left-right asymmetric decay of Cerl2 mRNA. *Nat Commun,* **3**: 1322.

Nejepinska, J., Malik, R., Moravec, M. & Svoboda, P. (2012) Deep sequencing reveals complex spurious transcription from transiently transfected plasmids. *PLoS One,* **7**(8): e43283.

Nigg, E.A. (1997) Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature,* **386**(6627): 779-787.

Nilsen, T.W. & Graveley, B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature,* **463**(7280): 457-463.

Nolan, P.M., Peters, J., Strivens, M., Rogers, D., Hagan, J., Spurr, N., Gray, I.C., Vizor, L., Brooker, D., Whitehill, E., Washbourne, R., Hough, T., Greenaway, S., Hewitt, M., Liu, X., McCormack, S., Pickford, K., Selley, R., Wells, C., Tymowska-Lalanne, Z., Roby, P., Glenister, P., Thornton, C., Thaung, C., Stevenson, J.A., Arkell, R., Mburu, P., Hardisty, R., Kiernan, A., Erven, A., Steel, K.P., Voegeling, S., Guenet, J.L., Nickols, C., Sadri, R., Nasse, M., Isaacs, A., Davies, K., Browne, M., Fisher, E.M., Martin, J., Rastan, S., Brown, S.D. & Hunter, J. (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet,* **25**(4): 440-443.

Nordeen, S.K. (1988) Luciferase reporter gene vectors for analysis of promoters and enhancers. *Biotechniques,* **6**(5): 454-458.

Norris, D.P. (2012) Cilia, calcium and the basis of left-right asymmetry. *BMC Biol,* **10**: 102.

Nusslein-Volhard, C. & Wieschaus, E. (1980) Mutations affecting segment number and polarity in Drosophila. *Nature,* **287**(5785): 795-801.

Ohno, S. (1970). *Evolution by gene duplication*, Springer-Verlag.

Oliver, D.G., Sanders, A.H., Hogg, R.D. & Hellman, J.W. (1981) Thermal gradients in microtitration plates. Effects on enzyme-linked immunoassay. *J Immunol Methods,* **42**(2): 195-201.

Pace, N.J. & Weerapana, E. (2014) Zinc-binding cysteines: diverse functions and structural motifs. *Biomolecules,* **4**(2): 419-434.

Paguio, A., Almond, B., Fan, F., Stecha, P., Garvin, D., Wood, M. & Wood, K. (2005) pGL4 Vectors: A new generation of luciferase reporter vectors. *Promega Notes,* **89**: 7-10.

Paguio, A., Stecha, P., Wood, K.V. & Fan, F. (2010) Improved dual-luciferase reporter assays for nuclear receptors. *Curr Chem Genomics,* **4**: 43-49.

Pan, G.J. & Pei, D.Q. (2003) Identification of two distinct transactivation domains in the pluripotency sustaining factor nanog. *Cell Res,* **13**(6): 499-502.

Parekh, R.B. & Rohlff, C. (1997) Post-translational modification of proteins and the discovery of new medicine. *Current Opinion in Biotechnology,* **8**(6): 718-723.

Patel, K., Kumar, A. & Durani, S. (2007) Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures. *Biochim Biophys Acta,* **1774**(10): 1247-1253.

Pavletich, N.P. & Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science,* **252**(5007): 809-817.

Pavletich, N.P. & Pabo, C.O. (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science,* **261**(5129): 1701-1707.

Pazzagli, M., Devine, J.H., Peterson, D.O. & Baldwin, T.O. (1992) Use of bacterial and firefly luciferases as reporter genes in DEAE-dextran-mediated transfection of mammalian cells. *Anal Biochem,* **204**(2): 315-323.

Pei, D.Q. & Shih, C.H. (1991) An "attenuator domain" is sandwiched by two distinct transactivation domains in the transcription factor C/EBP. *Mol Cell Biol,* **11**(3): 1480-1487.

Peifer, M., Berg, S. & Reynolds, A.B. (1994) A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell,* **76**(5): 789-791.

Pérez-Pinera, P., Menéndez-González, M. & Vega, J.A. (2006) Deletion of DNA sequences of using a polymerase chain reaction based approach. *Electronic Journal of Biotechnology,* **9**(5): 604-609.

Pillai, R.S., Bhattacharyya, S.N. & Filipowicz, W. (2007) Repression of protein synthesis by miRNAs: how many mechanisms? *Trends in Cell Biology,* **17**(3): 118-126.

Plant, N.J., Ogg, M., Crowder, M. & Gibson, G.G. (2000) Control and statistical analysis of in vitro reporter gene assays. *Anal Biochem,* **278**(2): 170-174.

Polozov, R.V., Sivozhelezov, V.S., Chirgadze, Y.N. & Ivanov, V.V. (2015) Recognition rules for binding of Zn-Cys2His2 transcription factors to operator DNA. *J Biomol Struct Dyn,* **33**(2): 253-266.

Posokhova, E., Shukla, A., Seaman, S., Volate, S., Hilton, M.B., Wu, B., Morris, H., Swing, D.A., Zhou, M., Zudaire, E., Rubin, J.S. & St Croix, B. (2015) GPR124 functions as a WNT7-specific coactivator of canonical beta-catenin signaling. *Cell Rep,* **10**(2): 123-130.

Pourebrahim, R., Houtmeyers, R., Ghogomu, S., Janssens, S., Thelie, A., Tran, H.T., Langenberg, T., Vleminckx, K., Bellefroid, E., Cassiman, J.J. & Tejpar, S. (2011) Transcription factor Zic2 inhibits Wnt/beta-catenin protein signaling. *Journal of Biological Chemistry,* **286**(43): 37732-37740.

Purandare, S.M., Ware, S.M., Kwan, K.M., Gebbia, M., Bassi, M.T., Deng, J.M., Vogel, H., Behringer, R.R., Belmont, J.W. & Casey, B. (2002) A complex syndrome of left-right axis, central nervous system and axial skeleton defects in Zic3 mutant mice. *Development,* **129**(9): 2293-2302.

Quinn, M.E., Haaning, A. & Ware, S.M. (2012) Preaxial polydactyly caused by Gli3 haploinsufficiency is rescued by Zic3 loss of function in mice. *Hum Mol Genet,* **21**(8): 1888-1896.

Redden, H. & Alper, H.S. (2015) The development and characterization of synthetic minimal yeast promoters. *Nat Commun,* **6**: 7810.

Reed, G.H., Kent, J.O. & Wittwer, C.T. (2007) High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics,* **8**(6): 597-608.

Rio, D.C., Clark, S.G. & Tjian, R. (1985) A mammalian host-vector system that regulates expression and amplification of transfected genes by temperature induction. *Science,* **227**(4682): 23-28.

Roberts, M.L. (2011). The use of functional genomics in synthetic promoter design. *In:* LOPES, H. (ed.) *Computational Biology and Applied Bioinformatics.* INTECH Open Access Publisher.

Rodriguez, M.S., Dargemont, C. & Hay, R.T. (2001) SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J Biol Chem,* **276**(16): 12654-12659.

Romoren, K., Thu, B.J., Bols, N.C. & Evensen, O. (2004) Transfection efficiency and cytotoxicity of cationic liposomes in salmonid cell lines of hepatocyte and macrophage origin. *Biochimica et Biophysica Acta,* **1663**(1-2): 127-134.

Rosenthal, N. (1987) Identification of regulatory elements of cloned genes with functional assays. *Methods in Enzymology,* **152**: 704-720.

Rudd, M.D., Gonzalez-Robayna, I., Hernandez-Gonzalez, I., Weigel, N.L., Bingman, W.E., 3rd & Richards, J.S. (2007) Constitutively active FOXO1a and a DNA-binding domain mutant exhibit distinct co-regulatory functions to enhance progesterone receptor A activity. *J Mol Endocrinol,* **38**(6): 673-690.

Ryan, R.F. & Darby, M.K. (1998) The role of zinc finger linkers in p43 and TFIIIA binding to 5S rRNA and DNA. *Nucleic Acids Res,* **26**(3): 703-709.

Salero, E., Perez-Sen, R., Aruga, J., Gimenez, C. & Zafra, F. (2001) Transcription factors Zic1 and Zic2 bind and transactivate the apolipoprotein E gene promoter. *The Journal of Biological Chemistry,* **276**(3): 1881-1888.

Salichs, E., Ledda, A., Mularoni, L., Alba, M.M. & de la Luna, S. (2009) Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genet,* **5**(3): e1000397.

Schagat, T., Paguio, A. & Kopish, K. (2007) Normalizing genetic reporter assays: approaches and considerations for increasing consistency and statistical significance. *Cell Notes,* **17**: 9-12.

Schlabach, M.R., Hu, J.K., Li, M. & Elledge, S.J. (2010) Synthetic design of strong promoters. *Proc Natl Acad Sci U S A,* **107**(6): 2538-2543.

Sharma, A., Antoku, S., Fujiwara, K. & Mayer, B.J. (2003) Functional interaction trap: a strategy for validating the functional consequences of tyrosine phosphorylation of specific substrates in vivo. *Mol Cell Proteomics,* **2**(11): 1217-1224.

Sherf, B.A., Navarro, S.L., Hannah, R.R. & Wood, K.V. (1996) Dual-Luciferase® reporter assay: An advanced co-reporter technology integrating firefly and Renilla luciferase assays. *Promega Notes,* **57**: 2-9.

Siebring-van Olst, E., Vermeulen, C., de Menezes, R.X., Howell, M., Smit, E.F. & van Beusechem, V.W. (2013) Affordable luciferase reporter assay for cell-based high-throughput screening. *J Biomol Screen,* **18**(4): 453-461.

Smale, S.T. & Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annu Rev Biochem,* **72**: 449-479.

Soni, S., Pchelintsev, N., Adams, P.D. & Bieker, J.J. (2014) Transcription factor EKLF (KLF1) recruitment of the histone chaperone HIRA is essential for beta-globin gene expression. *Proc Natl Acad Sci U S A,* **111**(37): 13337-13342.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A. & Soreq, H. (2005) Function of alternative splicing. *Gene,* **344**: 1-20.

Stanley, P.E. (1992) A survey of more than 90 commercially available luminometers and imaging devices for low-light measurements of chemiluminescence and bioluminescence, including instruments for manual, automatic and specialized operation, for HPLC, LC, GLC and microtitre plates. Part 1: Descriptions. *Journal of Bioluminescence and Chemiluminescence,* **7**(2): 77-108.

Stegmaier, P., Kel, A.E. & Wingender, E. (2004) Systematic DNA-binding domain classification of transcription factors. *Genome Informatics,* **15**(2): 276-286.

Stewart, M. (2007) Molecular mechanism of the nuclear protein import cycle. *Nature Reviews Molecular Cell Biology,* **8**(3): 195-208.

Stubbs, L., Sun, Y. & Caetano-Anolles, D. (2011) Function and Evolution of C2H2 Zinc Finger Arrays. *Subcell Biochem,* **52**: 75-94.

Sutherland, M.J., Wang, S., Quinn, M.E., Haaning, A. & Ware, S.M. (2013) Zic3 is required in the migrating primitive streak for node morphogenesis and left-right patterning. *Hum Mol Genet,* **22**(10): 1913-1923.

Sutherland, M.J. & Ware, S.M. (2009) Disorders of left-right asymmetry: heterotaxy and situs inversus. *Am J Med Genet C Semin Med Genet,* **151C**(4): 307-317.

Suzuki, M., Gerstein, M. & Yagi, N. (1994) Stereochemical basis of DNA recognition by Zn fingers. *Nucleic Acids Res,* **22**(16): 3397-3405.

Thomsen, N., Ali, R.G., Ahmed, J.N. & Arkell, R.M. (2012) High resolution melt analysis (HRMA); a viable alternative to agarose gel electrophoresis for mouse genotyping. *PloS one,* **7**(9): e45252.

Thorne, N., Inglese, J. & Auld, D.S. (2010) Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology. *Chem Biol,* **17**(6): 646-657.

Toyama, R., Gomez, D.M., Mana, M.D. & Dawid, I.B. (2004) Sequence relationships and expression patterns of zebrafish zic2 and zic5 genes. *Gene Expression Patterns,* **4**(3): 345-350.

Trinklein, N.D., Aldred, S.J., Saldanha, A.J. & Myers, R.M. (2003) Identification and functional analysis of human transcriptional promoters. *Genome Res,* **13**(2): 308-312.

Tsuji, T., Sun, Y., Kishimoto, K., Olson, K.A., Liu, S., Hirukawa, S. & Hu, G.F. (2005) Angiogenin is translocated to the nucleus of HeLa cells and is involved in ribosomal RNA transcription and cell proliferation. *Cancer Res,* **65**(4): 1352-1360.

Vaeth, M., Muller, G., Stauss, D., Dietz, L., Klein-Hessling, S., Serfling, E., Lipp, M., Berberich, I. & Berberich-Siebelt, F. (2014) Follicular regulatory T cells control humoral autoimmunity via NFAT2-regulated CXCR5 expression. *J Exp Med,* **211**(3): 545-561.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nature Review Genetics,* **10**(4): 252-263.

Vassen, L., Fiolka, K., Mahlmann, S. & Moroy, T. (2005) Direct transcriptional repression of the genes encoding the zinc-finger proteins Gfi1b and Gfi1 by Gfi1b. *Nucleic Acids Res,* **33**(3): 987-998.

Vaux, D.L., Fidler, F. & Cumming, G. (2012) Replicates and repeats - what is the difference and is it significant? A brief discussion of statistics and experimental design. *EMBO Reports,* **13**(4): 291-296.

Wabed, A.S. & Tang, X. (2010). Analysis of Variance (ANOVA) *In:* SALKIND, N. J. (ed.) *Encyclopedia of Research Design.* Thousand Oaks, CA, USA: SAGE Publications, Inc.

Ware, S.M. & Belmont, J.W. (2006a). ZIC3, CFC1, ACVR2B, and LEFTY2 and the Visceral Heterotaxies. *In:* EPSEIN, C., ERICKSON, R. & WYNSHAW-BORIS, A. (eds.) *Inborn errors of development: The molecular basis of clinical disorders of morphogenesis.* New York: Oxford University Press.

Ware, S.M., Harutyunyan, K.G. & Belmont, J.W. (2006b) Heart defects in X-linked heterotaxy: evidence for a genetic interaction of Zic3 with the nodal signaling pathway. *Dev Dyn,* **235**(6): 1631-1637.

Ware, S.M., Harutyunyan, K.G. & Belmont, J.W. (2006c) Zic3 is critical for early embryonic patterning during gastrulation. *Developmental Dynamics,* **235**(3): 776-785.

Ware, S.M., Peng, J., Zhu, L., Fernbach, S., Colicos, S., Casey, B., Towbin, J. & Belmont, J.W. (2004) Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. *American Journal of Human Genetics,* **74**(1): 93-105.

Warr, N., Powles-Glover, N., Chappell, A., Robson, J., Norris, D. & Arkell, R.M. (2008) Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. *Human Molecular Genetics,* **17**(19): 2986-2996.

Watson, J. (2011) Using bioluminescent assays as quantitative cell-based and biochemical biosensors. *Comb Chem High Throughput Screen,* **14**(8): 645-647.

Wei, X., Henke, V.G., Strubing, C., Brown, E.B. & Clapham, D.E. (2003) Real-time imaging of nuclear permeation by EGFP in single intact cells. *Biophysical journal,* **84**(2 Pt 1): 1317-1327.

Weirauch, M.T. & Hughes, T.R. (2011) A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell Biochem,* **52**: 25-73.

Wells, L., Vosseller, K. & Hart, G.W. (2001) Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science,* **291**(5512): 2376-2378.

Wen, W., Meinkoth, J.L., Tsien, R.Y. & Taylor, S.S. (1995) Identification of a signal for rapid export of proteins from the nucleus. *Cell,* **82**(3): 463-473.

Wessels, M.W., Kuchinka, B., Heydanus, R., Smit, B.J., Dooijes, D., de Krijger, R.R., Lequin, M.H., de Jong, E.M., Husen, M., Willems, P.J. & Casey, B. (2010) Polyalanine expansion in the ZIC3 gene leading to X-linked heterotaxy with VACTERL association: a new polyalanine disorder? *Journal of Medical Genetics,* **47**(5): 351-355.

Weston, K. & Bishop, J.M. (1989) Transcriptional activation by the v-myb oncogene and its cellular progenitor, c-myb. *Cell,* **58**(1): 85-93.

Winata, C.L., Kondrychyn, I., Kumar, V., Srinivasan, K.G., Orlov, Y., Ravishankar, A., Prabhakar, S., Stanton, L.W., Korzh, V. & Mathavan, S. (2013) Genome wide analysis reveals Zic3 interaction with distal regulatory elements of stage specific developmental genes in zebrafish. *PLoS Genet,* **9**(10): e1003852.

Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G. & Pryor, R.J. (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem,* **49**(6 Pt 1): 853-860.

Wolberger, C. (1999) Multiprotein-DNA complexes in transcriptional regulation. *Annu Rev Biophys Biomol Struct,* **28**: 29-56.

Wolfe, S.A., Nekludova, L. & Pabo, C.O. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct,* **29**: 183-212.

Wu da, Y. & Yao, Z. (2005) Isolation and characterization of the murine Nanog gene promoter. *Cell Res,* **15**(5): 317-324.

Xu, L. & Massague, J. (2004) Nucleocytoplasmic shuttling of signal transducers. *Nature Review Molecular Cell Biology,* **5**(3): 209-219.

Yang, Y., Hwang, C.K., Junn, E., Lee, G. & Mouradian, M.M. (2000) ZIC2 and Sp3 repress Sp1-induced activation of the human D1A dopamine receptor gene. *J Biol Chem,* **275**(49): 38863-38869.

Yokota, N., Aruga, J., Takai, S., Yamada, K., Hamazaki, M., Iwase, T., Sugimura, H. & Mikoshiba, K. (1996) Predominant expression of human zic in cerebellar granule cell lineage and medulloblastoma. *Cancer Research,* **56**(2): 377-383.

Yun, C. & Dasgupta, R. (2014) Luciferase reporter assay in Drosophila and mammalian tissue culture cells. *Curr Protoc Chem Biol,* **6**(1): 7-23.

Zhang, F., White, R.L. & Neufeld, K.L. (2001) Cell density and phosphorylation control the subcellular localization of adenomatous polyposis coli protein. *Mol Cell Biol,* **21**(23): 8143-8156.

Zhang, X., Chen, H.Z. & Rovin, B.H. (2003) Unexpected sensitivity of synthetic Renilla luciferase control vectors to treatment with a cyclopentenone prostaglandin. *Biotechniques,* **35**(6): 1144-1146, 1148.

Zheng, R. & Blobel, G.A. (2010) GATA Transcription Factors and Cancer. *Genes Cancer,* **1**(12): 1178-1188.

Zhu, L. & Huq, E. (2011) Mapping functional domains of transcription factors. *Methods Mol Biol,* **754**: 167-184.

Zhu, L., Peng, J.L., Harutyunyan, K.G., Garcia, M.D., Justice, M.J. & Belmont, J.W. (2007) Craniofacial, skeletal, and cardiac defects associated with altered embryonic murine Zic3 expression following targeted insertion of a PGK-NEO cassette. *Frontiers in Bioscience,* **12**: 1680-1690.