The Function of *TCF4* in Pitt-Hopkins Syndrome and Schizophrenia

Marc P. Forrest

Cardiff University

A thesis submitted for the degree of Doctor of Philosophy

2013

Supervisors: Prof. Derek Blake and Prof Lawrence Wilkinson

This thesis is dedicated to my Dad,
Ronald E. L. FORREST

In loving memory

Abstract

Genome-wide association studies have identified *TCF4* (transcription factor 4) as a susceptibility gene for schizophrenia. In addition, rare *TCF4* mutations cause Pitt Hopkins syndrome (PTHS), a severe form of intellectual disability associated with characteristic facial features, developmental delay and autonomic dysfunction. *TCF4* belongs to the basic helix-loop-helix family (bHLH) of transcription factors that play a central role in development, however the precise function of this gene in the brain is unknown. In this study, I use molecular and cellular techniques to improve our understanding of TCF4 function in the brain.

Modeling of PTHS-associated missense mutations in transfected cells showed that *TCF4* mutations affecting the DNA-binding domain cause mislocalisation of the mutant protein. DNA-binding domain mutations also impaired dimerisation and attenuated transcriptional activity at the *NRXN1* and *CNTNAP2* promoters. *TCF4* mutations affecting other domains of the protein had context-specific deficits in dimerisation and transactivation under the same conditions.

Microarray analysis of SH-SY5Y cells where all TCF4 isoforms had been knocked down identified gene expression changes affecting cellular processes including epithelial-to-mesenchymal transition, apoptosis and neurodevelopment. However, isoform-specific knockdown experiments showed that TCF4-A and TCF4-B isoforms affect distinct biological processes such as the cell cycle, chromatin modification (TCF4-B), cell adhesion and cytoskeletal remodeling (TCF4-A).

Finally, mass spectrometry was used to identify TCF4-associated proteins in SH-SY5Y cells. These experiments showed that TCF4-B and TCF4-A co-purified with proteins involved in chromatin organization, cell cycle control and RNA splicing. In addition to the bHLH factors HAND2 and TWIST2, TCF4 co-purified with components of the nuclear co-repressor complex. TCF4 also has multiple phosphorylation sites in both activation domains suggesting that TCF4 function may be regulated by kinase signaling.

In conclusion, these data provide a mechanistic insight into the function of TCF4 that may advance our understanding of disease processes in PTHS and schizophrenia.

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Derek Blake for his remarkable support and guidance during the course of my PhD and until the very last moments of writing this thesis. I am grateful for his scientific insight that has allowed me to develop as a researcher. I would like to thank my supervisor Prof Lawrence Wilkinson for his interest in my work and his inspirational views on science.

I owe a huge debt of gratitude to Adrian for the time and effort he has put into supporting my work, for his friendship, and for the stimulating discussions that have helped shape my views in biological research. I am also very grateful to Caz for her help on training me in the lab, for her friendship and her love of coffee. I would like to thank my fellow PhD student Ria for her friendship and her contributions to my work.

Many thanks to all the great people on the (ex) 2nd floor for making science enjoyable at all times and for providing laughs and coffee breaks on demand.

Thank you to my friends Russell, Charlie, Ben and the rest of the Cardiff team for providing endless entertainment and for keeping me sane during the writing of this thesis. A big thank you goes out to all my friends in Geneva for offering the best possible times when I need a break.

An enormous thank you to my Mum for her continuous love and support, her positive energy, skype chats, and extraordinary food. I must also say a huge thank you to my sister Sophie who always has the right advice and is always there when I need her. Finally, a big thank you to my Dad for all his love and support and for motivating me to succeed in science.

Table of Contents

Chapter 1

Introduction

Human genetic studies have been instrumental in uncovering part of the genetic etiology of complex psychiatric disorders (Sullivan et al., 2012). However, understanding the function of genes that contribute to disease processes in these disorders presents a major challenge in biology and medicine. *Transcription factor 4* (*TCF4*) is a gene that has been implicated in the genetics of intellectual disability and schizophrenia. Rare mutations affecting the coding regions of *TCF4* cause Pitt-Hopkins syndrome (PTHS), a severe form of intellectual disability associated with a range of other developmental abnormalities (Amiel et al., 2007). Furthermore, common variation in *TCF4* is associated with a small but robust increase in risk for schizophrenia (Stefansson et al., 2009). The genetic association of common and rare variants in *TCF4* to neurodevelopmental disorders underscores its importance for brain development and function. These discoveries therefore offer a unique opportunity to understand the function of this gene in the context of the nervous system. Advancing our knowledge of the fundamental biology of *TCF4* will be an essential part of deciphering disease mechanisms related to schizophrenia and intellectual disability.

TCF4 is a member of the basic helix-loop-helix (bHLH) family of transcription factors that have an important role in a number of developmental processes (Massari and Murre, 2000). Many members of the bHLH family are involved in regulating neurodevelopment although the function of TCF4 in the nervous system remains largely unknown. The following chapter is divided into two sections: The first summarises the current knowledge on the molecular and cellular functions of TCF4 whilst the second provides an overview of PTHS and schizophrenia and their genetic association to *TCF4*.

1.1. *TCF4* Nomenclature

Although *TCF4* (GeneID: 6925) is the gene's official Human Genome Organisation (HUGO) symbol, in the literature it is often referred to as E2-2, immunoglobulin transcription factor 2 (ITF2) or SL3-3 enhancer factor (SEF2). Importantly, transcription factor 4 should not be confused with transcription factor 7-like 2 (*TCF7L2)*, a gene on chromosome 10q25.3 that is also referred to *T-cell factor 4,* and therefore shares the *TCF4* acronym. The *T-cell factor 4* (*TCF7L2, TCF4*) gene encodes a high mobility group (HMG) box-containing transcription factor of the Wnt signaling pathway, and is not related to the work in this thesis.

1.2 *TCF4*: gene discovery and alternative splicing

In humans, *TCF4* is located on chromosome 18q21.1 that spans approximately 450kb and has at least 20 coding exons. The *TCF4* cDNA was initially discovered as its cognate protein was able to bind the μE5 heavy chain and κE2 light chain immunoglobulin enhancers (Henthorn et al., 1990). In this study, the *TCF4* cDNA was found to encode a 623 amino acid isoform that was designated ITF2 for 'immunoglobulin transcription factor 2'. The μE5 and κE2 enhancer sequences both shared an E-box (Ephrussi-box) DNA element (CACCTG) that was recognised as an important binding site (Murre et al., 1989a; Henthorn et al., 1990). Independently, the canonical TCF4 sequence, encoding a 667 amino acid protein (SEF2-1B), was isolated from human thymocytes along with truncated TCF4 isoforms TCF4-A (SEF2-1A) (Corneliussen et al., 1991). These isoforms were identified by their ability to bind an enhancer of the murine leukemia virus SL3-3 and were termed SEF2 for 'SL3-3 enhancer factors 2'. Orthologues (mouse and rat) of TCF4 were found to regulate other E-box-containing regulatory regions such as the rat tyrosine hydroxlase (TH) enhancer and the human somatostatin receptor-2 (SSTR2) promoter (Yoon and Chikaraishi, 1994; Pscherer et al., 1996). Together, these studies established TCF4 as a conserved bHLH transcription factor that binds E-box sequences in the promoters and enhancers of certain genes.

Since these initial discoveries, an extensive list of alternatively spliced TCF4 transcripts has been identified (Figure 1.1). Many of these transcripts have different 5' exons and may be transcribed from alternative intragenic promoters. The most

detailed study of *TCF4* alternative splicing in humans identified 18 transcripts with distinct N-terminal sequences that result from the use of alternative start sites within the gene (Sepp et al., 2011). *TCF4* transcripts can contain an alternatively spliced exon of 12 bp that code for the amino acids "RSRS". Transcripts that contain the insertion are denoted with a "+" whilst transcripts that do not contain the insertion are denoted with a "-" (Corneliussen et al., 1991; Sepp et al., 2011). For the purpose of this thesis,"TCF4" will refer to the canonical TCF4-B^{+} isoform that contains 671 amino acids (including RSRS) unless otherwise specified.

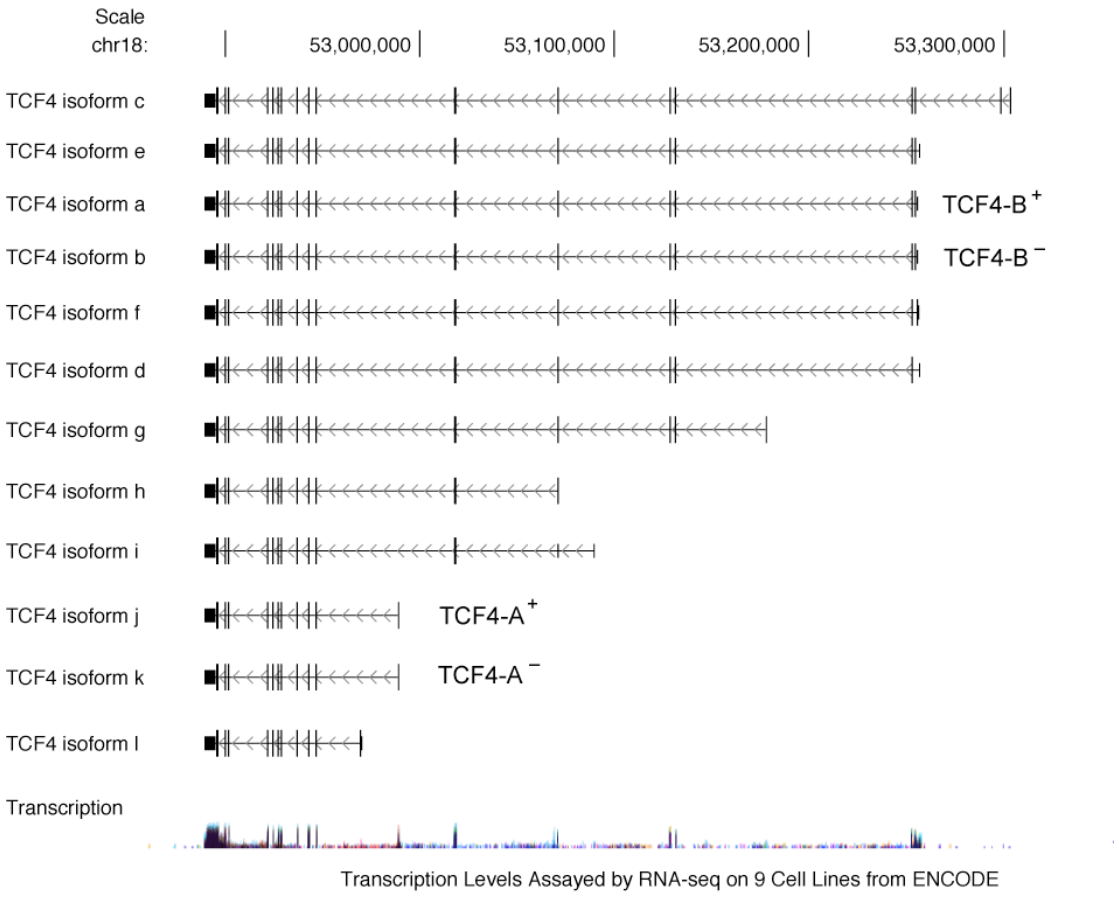


Figure 1.1 Transcriptional architecture of *TCF4*.
TCF4 transcripts from the NCBI Reference Sequences Database (RefSeq) are shown. Some of the mRNA transcripts described by Sepp *et al.* 2011 are not in the RefSeq database (see Table 5.1) (Sepp et al., 2011). The level of transcription in 9 different cell lines measured by RNA-seq is indicated in the track below the *TCF4* mRNA schematic. The image was modified from data in the UCSC genome browser.

1.3 E-protein Isoforms

The E-proteins are a class of bHLH transcription factors that are orthologous to the *Drosophila* protein daughterless (da) (Murre et al., 1989a; Bain and Murre, 1998). In humans, these proteins include TCF4, Transcription factor 3 (TCF3, E2A) and Transcription factor 12 (TCF12, HEB). E-protein genes encode at least two distinct alternatively spliced protein isoforms (Figure 1.2). The *TCF3* (*E2A*) gene is capable of producing two highly similar protein isoforms called E12 and E47 that differ uniquely in the bHLH region of their sequence, giving them different DNA binding properties (Sun and Baltimore, 1991). In contrast, the TCF12 (*HEB)* gene produces two protein isoforms that differ in their N-terminal domain (Wang et al., 2006b). The alternative HEB isoform (HEBAlt) is 170 amino acids shorter than the canonical HEB isoform (sometimes called HEBCan). *TCF4* is able to produce a wide variety of transcripts with different N-termini, however only two protein isoforms have been consistently described, TCF4-A (SEF2-1A, E2-2A, ITF2A) and TCF4-B (TCF4, SEF2-1B, E2-2B, ITF2B) (Sepp et al., 2011). Similarly to the HEB isoforms, TCF4-A is a truncated form of TCF4-B, missing the first 183 amino acids. Multiple sequence alignment of the four most studied E-protein isoforms (TCF4-B, HEBcan, E47, E12) reveals their high degree of sequence similarity, especially in regions of the bHLH domain and the adjacent domain C that is unique to E-proteins (amino acids 566-644).

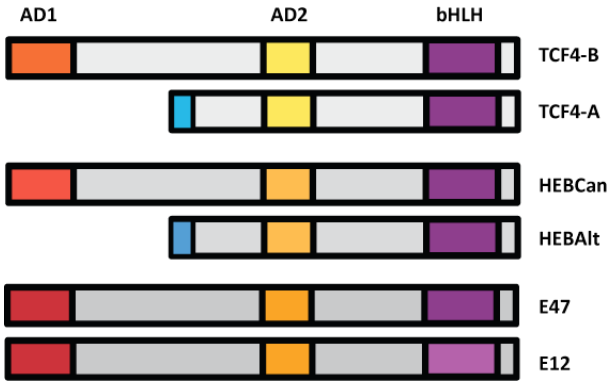


Figure 1.2 Structure of E-protein isoforms. The illustration depicts the structure of the two alternatively spliced protein isoforms for TCF4, TCF3 (E2A) and TCF12 (HEB). The TCF3 isoforms (E47, E12) contain two activation domains (AD1 and AD2) and only differ slightly in their basic helix-loop-helix domain (bHLH). *TCF4* and *TCF12* (HEB) encode full length isoforms (TCF4-B and HEBCan) and one shorter protein isoform that lack the N-terminal AD1 (TCF4-A and HEBAlt).

1.4 The bHLH family

The bHLH family is organised into seven functional classes according to tissue expression, dimerisation capability and DNA binding specificity (Appendix I) (Massari and Murre, 2000). bHLH family members are characterised by their core bHLH structural domain and their ability to bind E-boxes (consensus sequence "CANNTG"). TCF4 and the other E-proteins form the Class I bHLH proteins. E-proteins are widely expressed in different tissues and are able to form stable homodimers that interact with E-boxes on DNA (Ellenberger et al., 1994). Class I proteins also form heterodimers with class II (e.g. ASCL1, MYOD), class V (ID proteins) and class VI (HES1) members of the bHLH family that regulate its DNA binding properties (Figure 1.3). Class II transcription factors have tissue-restricted expression patterns and are potent inducers of cell-type specification such as MyoD for myogenesis or ASCL1 for neurogenesis (Tapscott et al., 1988; Turner and Weintraub, 1994; Farah et al., 2000). E-proteins allow other bHLH to bind to DNA and activate transcription, making E-proteins key regulators biological activity (Murre et al., 1989b; Longo et al., 2008).

Conversely, class V proteins and class VI transcription factors are characterised by their inhibitory or repressive activities. The ID proteins (ID1-4, Class V) are orthologous to the *Drosophila* protein extramacrochaetae (emc) and lack the basic region of the bHLH sequence rendering them unable to bind DNA (Benezra et al., 1990). ID proteins heterodimerise with E-proteins and sequester them into inactive complexes that limit their availability for class II transcription factors (Sun et al., 1991). The expression of ID proteins affects the balance between cell growth and differentiation (Lyden et al., 1999; Zebedee and Hara, 2001). The hairy and enhancer of split (HES) proteins (Class VI) are unique as they contain a proline residue in their basic region. This unique structure enables them to bind N-boxes (consensus sequence "CACNAG") rather than E-boxes (Akazawa et al., 1992; Ohsako et al., 1994). HES proteins also heterodimerise with class I factors and inhibit their DNA binding properties in a dominant negative manner (Akazawa et al., 1992; Sasai et al., 1992). HES proteins play a central role in maintaining progenitor cells in an undifferentiated state (Kageyama et al., 2007). The balance of bHLH factors during development is therefore essential in regulating E-protein activity as well as the

lineage specification of different cells types (Massari and Murre, 2000; Jogi et al., 2002).

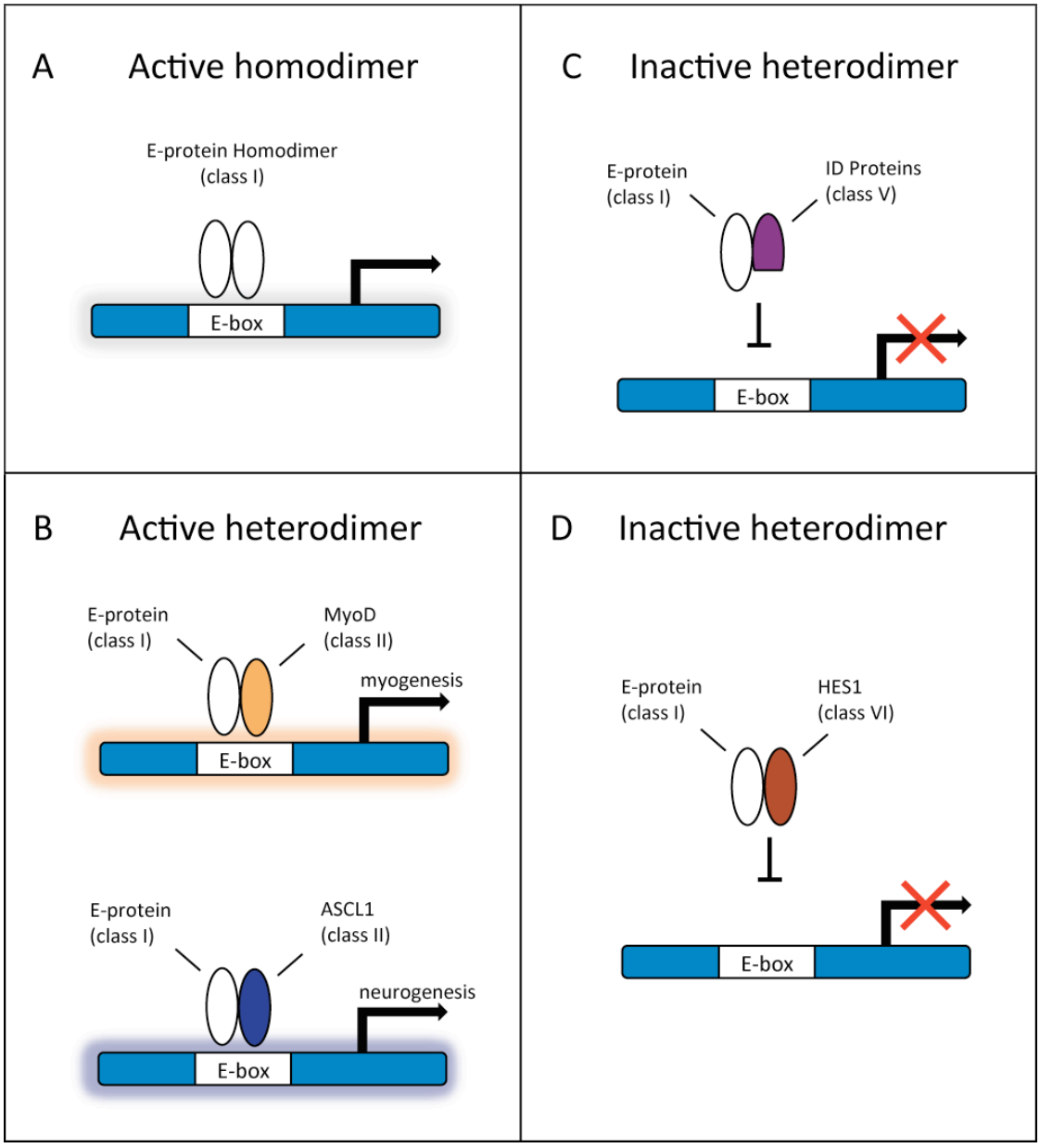


Figure 1.3 Model of transcriptional regulation by bHLH proteins. The schematic depicts the general mechanisms by which bHLH proteins regulate E-box containing promoters or enhancers. (**A**) E-proteins (class I), such as TCF4 can self-associate as homodimers to bind DNA and activate or repress transcription. (**B**) E-proteins associate with class II transcription factors such as MYOD or ASCL1 to form DNA binding heterodimers. E-protein/Class II heterodimers activate lineage specific gene expression programs such as myogenesis or neurogenesis. (**C**) E-proteins can heterodimerise with ID proteins (class V) which prevents them from binding DNA. ID proteins lack the crucial basic domain required for DNA binding. (**D**) E-proteins can heterodimerise with class VI proteins such as HES1. HES1 can interact with E-proteins and inhibit their DNA binding activity.

1.5 Structure and function of the TCF4 protein

The canonical human TCF4 (TCF4-B$^+$, NP_001077431) protein sequence is 671 amino acids and is encoded by at least 18 exons and includes the RSRS insertion sequence (Figure 1.4). Critical to its function, the TCF4 protein contains an assortment of structural and function domains that regulate multiple aspects of its biochemistry including transcriptional activity, DNA binding, protein dimerisation and cellular localisation. The N-terminal region of the protein contains regions that modulate transcriptional activity and a nuclear localisation signal (NLS) that controls subcellular location of TCF4. The C-terminal region of the protein contains the bHLH domain and domain C that us essential for the DNA binding and dimerisation properties of the protein. TCF4 isoforms all share the same C-terminus but vary in composition at their N-termini, multiplying the properties of their transcriptional and biological activity.

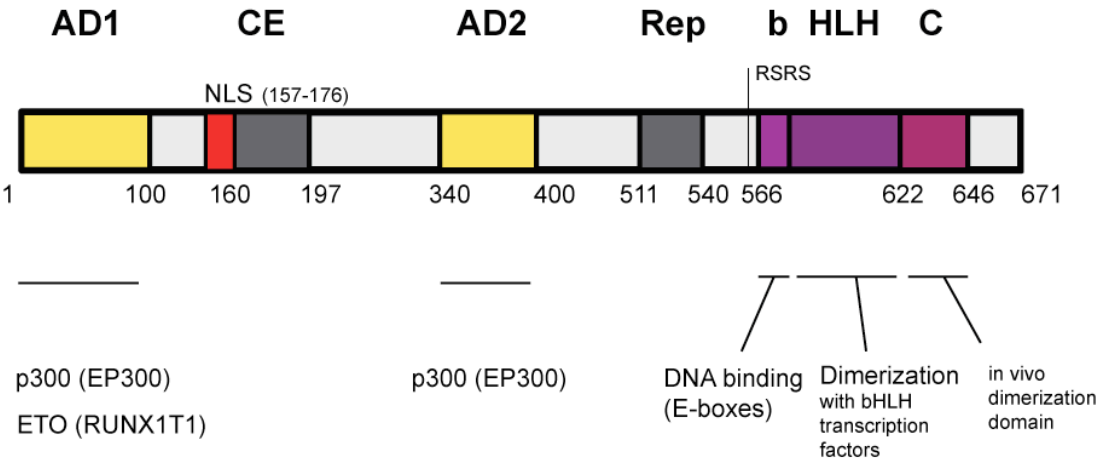


Figure 1.4. TCF4 protein structure. The illustration shows the conceptual organisation of the TCF4 protein domains and their summarised functions. AD1 and AD2 are proposed to interact with the p300 co-activator whilst AD1 also participates in ETO co-repressor recruitment. The basic region of the bHLH domain mediates DNA binding to E-boxes whilst the helix-hoop-helix forms an interaction surface for dimerisation. Domain C is required for *in vivo* homodimerisation. Numbers represent the approximate amino acid positions of each domain in full length TCF4 (TCF4-B$^+$; 671 amino acids including RSRS insertion). Abbreviations and amino acid positions: Activation domain 1 (AD1; 1-100), CE repressor domain (CE; 160-179), nuclear localisation signal (NLS; 157-176), activation domain 2 (AD2; 340-400), Rep repressor domain (Rep; 511-540), basic domain (b; 566-583), helix-loop-helix domain (583-622), domain C (C; 622-646). Illustration not to scale.

1.5.1 Activation domains and transcriptional co-factors

The E-proteins all contain two activation domains (AD) that have been defined by deletion mapping and mutagenesis (Figure 1.4). These domains are highly conserved across the E-proteins and are required for activity in yeast and mammalian cells (Quong et al., 1993; Massari et al., 1996). AD1 is located within the first 100 amino acids of TCF4, a region sharing 45% similarity with the other E-proteins (Massari et al., 1996) (Figure 1.4). Within this region is a highly conserved alpha-helical structure (amino acids 11-38) containing the LDFS motif that is essential for transactivation properties (Massari et al., 1996). Interestingly, the transactivation properties of the AD1 domain vary across mammalian cell lines, with the highest activity found in neuroblastoma cells (Aronheim et al., 1993). The TCF4-A isoform is truncated at its N-terminus and is devoid of AD1 (Figure 1.2).

AD2 is predicted to be a loop-helix (LH, amino acids 340-400) structure or leucine zipper (LZ, amino acids 379-400) that is highly conserved and present in all described TCF4 isoforms (Quong et al., 1993; Sepp et al., 2011) (Figure 1.4). In general, AD2 has relatively less activation potential than AD1 in many cell lines, although it is highly active in cells derived from the pancreas (Aronheim et al., 1993). Further *in vivo* experiments in *Zebrafish* have identified cell-type specificity for AD1 and AD2. AD1 activates expression in all major cell types of the embryo (skin, sclerotome, myotome, notochord, and nervous system) whilst AD2 is principally active in the myotome (muscle precusors) (Argenton et al., 1996).

In yeast, AD1-mediated transactivation is dependent on the SAGA (Spt-Ada-Gcn5 acetyltransferase) complex, a multifunctional co-activator that regulates transcription (Massari et al., 1999; Wu et al., 2004). AD1 can directly interact with the SAGA complex in a manner that is dependent on the integrity of its alpha helical structure and LDFS motif (Massari et al., 1999). In mammalian cells, AD1 and AD2 cooperate to recruit p300 (EP300) or its close paralogue CBP (collectively referred to as p300/CBP) that are transcriptional co-activators with histone acetyltransferase activity (Bradney et al., 2003; Bayly et al., 2004). The KIX domain of p300/CBP and the two alpha-helical surfaces of AD1 and AD2 mediate the interaction between these proteins (Bayly et al., 2004; Denis et al., 2012). A PCET motif (LXXLL) is critical

for the interaction with AD1 and overlaps with the LDFS sequence important for SAGA recruitment (Denis et al., 2012). Interestingly the PCET motif in AD1 is also able to interact with the transcriptional co-repressor ETO (RUNX1T1) (Zhang et al., 2004). ETO represses TCF4 by competing with p300 for the AD1 binding site, abrograting its association the with histone deacetylase complexes, and thereby repressing transcription (Hug and Lazar, 2004; Zhang et al., 2004). Although these interactions have not yet been studied in the context of the nervous system, E-protein activation domains remain essential interfaces for the recruitment of transcriptional co-regulator complexes that have import roles in physiological processes (Bayly et al., 2004; Bhalla et al., 2008).

1.5.2 Repressor domains and the nuclear localisation signal

In addition to AD1 and 2, two repressor domains have also been identified by deletion mapping of the TCF4 protein sequence (Figure 1.4). The CE repressor is located between the AD1 and AD2 activation domains (amino acids 160-197, Figure 1.4). Deletion of the CE repressor from TCF4 leads to an augmentation of transcriptional activity in mammalian cells, beyond that of the full length TCF4 protein (Herbst and Kolligs, 2008). The CE repressor is also sufficient to abolish the transcriptional activity mediated by AD1 when the two domains are directly fused together without intervening sequences (Herbst and Kolligs, 2008). The CE repressor is therefore a modulator of transcriptional activity that can interact with AD1 to govern transcriptional output.

Imbedded in the CE repressor domain are the residues encoding an NLS (Figure 1.4). This motif is characterised by two clusters of basic amino acids that are important to direct TCF4 to the nucleus (RRR, amino acids 157-159; KVKK, amino acids 171-174; Figure 1.4) (Sepp et al., 2011). TCF4 isoforms that contain these residues have an exclusively nuclear localisation whereas isoforms that lack this motif, such as TCF4-A, are present in both the nucleus and the cytoplasm of cells. Isoforms lacking an NLS are able to translocate to the nucleus by dimerising with other bHLH family members that do contain an NLS in a so called ‘piggy-back’ mechanism (Sepp et al., 2011).

The second repressor region, termed the Rep domain, is located between AD2 and the bHLH domain (amino acids 511-540, Figure 1.4). The Rep domain can repress both AD1 and AD2 independently in a selection of mammalian cell lines (Markus et al., 2002b). The Rep domain plays a role in maintaining E-protein homodimers in an inactive state on certain enhancers (Markus et al., 2002b). The intramolecular repression mediated by the Rep domain may be necessary for AD2 to function in particular heterodimeric complexes (Markus et al., 2002b). Intramolecular repressor domains are therefore able to regulate the E-protein activation domains in a context dependent manner (Markus et al., 2002b; Wong et al., 2008).

1.5.3 Structure and function of C-terminal domains

TCF4 and other members of the bHLH family possess as their defining feature a structural motif consisting of two alpha-helices connected by an intervening loop termed the bHLH domain (Phillips, 1994). The N-terminal part of this domain contains a highly conserved set of positively charged amino acids that make up the basic region and participates in DNA interactions (Voronova and Baltimore, 1990; Ellenberger et al., 1994). The remainder of the bHLH motif is a surface that favours homo- and hetero- dimerisation through the assembly of a stable four-helix bundle with a defined hydrophobic core (Ellenberger et al., 1994; Longo et al., 2008). The structure of E-protein dimers exploits the central two-fold symmetry of the E-box ("CACCTG") with each monomer contacting an E-box half site ("CAC" or "CTG") (Ellenberger et al., 1994). Directly C-terminal to the bHLH domain is a highly conserved set of amino acids termed domain C. This domain is exclusively present in E-proteins and is critical for dimerisation in physiological contexts (Goldfarb et al., 1998).

1.6 Ca^{2+}-dependent regulation of TCF4

The DNA binding properties of E-proteins are regulated by Ca^{2+} binding proteins such as Ca^{2+}-dependent calmodulin (CaM), S100α and S100β (Corneliussen et al., 1994; Onions et al., 1997; Larsson et al., 2001, 2005). These proteins are all Ca^{2+}-sensing molecules that have important roles in intracellular signaling (Hermann et al., 1998). CaM is ubiquitously expressed whereas S100α and S100β are predominantly expressed in muscle and glial cells respectively but are also found in neurons

(Corneliussen et al., 1994; Donato et al., 2013). In the presence of Ca^{2+}, CaM can selectively inhibit the DNA binding of E-protein homodimers whereas E-protein heterodimers are much less sensitive (Corneliussen et al., 1994). Ca^{2+}-dependent proteins interact directly with the basic region of the bHLH domain, involved in DNA binding (Onions et al., 1997). Consistent with these interactions, altering levels of CaM or Ca^{2+} in cells has a direct effect on E-protein-mediated transcription (Hauser et al., 2008a; Hauser et al., 2008b).

1.7 TCF4 expression

TCF4 is widely expressed during development and in adult tissues (Soosaar et al., 1994; Shain et al., 1997; Brockschmidt et al., 2007; de Pontual et al., 2009). In mice, *Tcf4* (*ME2*) is expressed in the developing nervous system including the cortex, cerebellum, pons, medulla and spinal chord (Soosaar et al., 1994; Pscherer et al., 1996; Dorflinger et al., 1999). TCF4 also has high expression in non-neuronal tissues such as skeletal muscle, subepidermal connective tissue and in the developing limbs (Soosaar et al., 1994; Dorflinger et al., 1999). Later in development (e18) and during adulthood, *Tcf4* mRNA is detected at high levels in the hippocampus, olfactory bulb, ventricular zones, rostal migratory stream (RMS), cerebellum and at a lower levels in the neocortex and amygdala (Soosaar et al., 1994; Dorflinger et al., 1999). The widespread expression during developmental and more specific expression pattern during adulthood has led to the proposal that *Tcf4* could have different regulatory roles in these two developmental stages (Soosaar et al., 1994). *Tcf4* expression is also distinct from the genes encoding the other E-proteins (Chiaramello et al., 1995). E-protein gene expression is generally high during neurodevelopment however, *Tcf4* expression persists in proliferative regions of the adult brain where *Tcf3* and *Tcf12* expression declines (Neuman et al., 1993; Uittenbogaard and Chiaramello, 1999; Ravanpay and Olson, 2008; Brzozka et al., 2010). This places TCF4 in a unique position to regulate bHLH transcription factor networks in the adult brain.

Examination of human *TCF4* mRNA revealed a similar expression pattern to that of the mouse, with high *TCF4* expression identified in human foetal and adult brain (Sepp et al., 2011). During human development, *TCF4* is expressed throughout the central nervous system (CNS), sclerotome, lateral plate mesoderm and pharyngeal

arch mesenchyme (de Pontual et al., 2009). *TCF4* is also expressed in NCAM (neural cell adhesion molecule)-expressing sympathetic, parasympathetic and enteric ganglia and the ventricular zones of the CNS (de Pontual et al., 2009). In the adult brain, TCF4 expression is highest in the cerebellum but is also prominent in the cerebral cortex (Sepp et al., 2011). *TCF4* mRNA is also present in non-neuronal tissues such as the spleen, uterus, lung, thymus and placenta (Sepp et al., 2011). *In situ* hybridization analysis of adult human brain demonstrated that *TCF4* is expressed in the dentate gyrus and in the CA1-CA3 regions of the hippocampus, neurons of subiculum and parahippocampal gyrus of the cortex, and cerebellar granule neurons (Sepp et al., 2011).

1.8 Non-neuronal functions of TCF4

Consistent with its widespread expression throughout development, TCF4 has been implicated in a number of cellular and disease processes. In addition to neuronal functions, TCF4 has been studied for its role in the development of immune system and epithelial-to-mesenchymal transition (EMT) (Kee, 2009; Cano and Portillo, 2010). TCF4 is also a highly significant risk factor for Fuchs' endothelial corneal dystrophy (FECD) (Baratz et al., 2010).

E-proteins are critical regulators of lineage specification in the immune system (Kee, 2009). Analysis of *Tcf4* knockout mice has revealed that *Tcf4* expression can effect B- and T- cell development. Although Tcf4 is not absolutely required for B cell development, *Tcf4* knockout mice display a reduced number of pro-B cells demonstrating a role for Tcf4 in early B-cell development (Zhuang et al., 1996). Expression of Tcf4 may be required in marginal versus follicular zone fate decisions during pro-B cell expansion (Wikstrom et al., 2006). Specific deletion of *Tcf4* can also lead to a partial block in T-lymphocyte development when targeted postnatally (Bergqvist et al., 2000). In contrast, *Tcf4* has crucial role in plasmocytoid dendritic cell development (pDC), a unique immune cell type specialised in type I interferon secretion in response to viral nucleic acids (Cisse et al., 2008). *Tcf4* is highly expressed in pDCs and it is essential for the development and maintenance of the pDC phenotype (Ghosh et al., 2010). *Signal transducer and activator of transcription 5* (*Stat5*) controls the expression of *Id2* and *Tcf4* during pDC development(Li et al.,

2012a). Heterozygous *Tcf4* mutant mice have decreased expression of many pDC-associated genes (Cisse et al., 2008). Interestingly, PTHS patients show similar defects in pDC development and function to mice that only have one functional *Tcf4* allele (Cisse et al., 2008).

In epithelial cells, TCF4 is emerging as an important regulator of EMT (Cano and Portillo, 2010). EMT is a complex cellular process that is required for embryonic development, tissue repair and cancer metastasis. Epithelial cells undergoing EMT lose their apico-basal polarity and intercellular junctions to become migratory acquiring a mesenchymal cell phenotype (Lim and Thiery, 2012). This process enables cells to penetrate the extracellular matrix and coordinate tissue remodeling. A hallmark of EMT is the loss of *E-cadherin* (epithelial-cadherin, *CDH1*) expression, although multiple epithelial and mesenchymal markers are differentially expressed during the EMT process (Kalluri and Weinberg, 2009). Over-expression of TCF4 in epithelial MDCK cells leads to a potent induction of EMT (Sobrado et al., 2009). Cells over-expressing TCF4 acquire a motile and highly invasive phenotype that is associated with change in a number of EMT markers (Sobrado et al., 2009). E-cadherin expression is down-regulated with other epithelial markers whilst mesenchymal markers such as N-cadherin (neural cadherin, *CDH2*), vimentin and fibronectin are upregulated in these cells (Sobrado et al., 2009). The phenotype is similar to epithelial cells over-expressing the transcription factors SNAI1, SNAI2, and TCF3 (E47) that are also important regulators of EMT (Cano et al., 2000; Perez-Moreno et al., 2001; Bolos et al., 2003). However, TCF4 appears to be acting downstream of these transcription factors and is not required to maintain the mesenchymal cell fate (Moreno-Bueno et al., 2006; Sobrado et al., 2009). In agreement with EMT functioning during cancer metastasis, TCF4 expression is also dysregulated in a number of human cancers (Kolligs et al., 2002; Zhai et al., 2002; Vicent et al., 2008; Herbst et al., 2009). However, by contrast to TCF4 mediated effects in epithelial cells, TCF4 expression is associated with inhibition of migration in vascular endothelial cells underscoring its context dependent activities (Tanaka et al., 2010).

Common variation in *TCF4* is associated with increased risk of FECD, a condition affecting 5% of the U.S. population over 40 years of age (Baratz et al., 2010). FECD

is caused by a progressive degeneration of the corneal endothelium that is accompanied by deposition of abnormal extracellular matrix in the cornea (Schmedt et al., 2012). These aberrations lead to corneal edema and reduced vision in patients. A genome-wide association study identified *TCF4* as highly significant (rs613872, $P = 2.3 \times 10^{-26}$) risk factor in typical FECD (Baratz et al., 2010). This association increased the odds of having FECD by a factor of 30 and allowed cases to be predicted from controls with 76% accuracy, suggesting TCF4 is a major contributor to FECD. The risk allele rs613872 is also associated with FECD disease severity and central corneal thickness, enforcing its role in the endothelial organisation of the cornea (Igo et al., 2012). However, rs613872 was not associated to a change in corneal endothelial cell density or morphology in a study of 445 young adults (Mackey et al., 2012). In addition to the rs613872 allele, a separate study identified a trinucleotide expansion (TGC) in intron 3 of *TCF4* that is present in a high proportion of FECD cases (Wieben et al., 2012). A repeat length of 50 or more trinucleotides was identified in 79% patients (n = 66) and 3% of controls (n = 63) although the mechanism by which this expansion causes susceptibility to FECD is unknown.

1.9 TCF4 and neurodevelopment

The association of *TCF4* variants with neurological and neuropsychiatric disease has prompted attention to its role in brain development and function (Blake et al., 2010).Our current understanding of TCF4 in this context is limited, although emerging evidence suggests that it contributes to neurodevelopmental events and has an impact on brain function.

1.9.1 The proneural genes and neurogenesis

In the vertebrate embryo, neurodevelopment begins with the creation of the neuroepithelium that contains neural stem cells (or neuroblasts) with self-renewing properties. Neuroepithelial cells can divide symmetrically to produce two neuroblasts with stem cell characteristics or divide asymmetrically to generate intermediate progenitors (Huttner and Brand, 1997; Farkas and Huttner, 2008). During asymmetric division, intermediate progenitors derived from neuroepithelial cells lose their epithelial characteristics and migrate away from the basal lamina. Intermediate progenitors have a limited proliferation potential and eventually differentiate into mature neurons and glia. The process of neurogenesis is dependent on the expression of the proneural bHLH transcription factors of the *Drosophila atonal* (*ato*) and *achaete-scute* (*ac/sc*) families (Bertrand et al., 2002). Proneural genes are first expressed in neuroepithelial cells that are committed to a neural fate but are also implicated in the delamination and migration of intermediate progenitors as well as the final steps of neuronal differentiation (Bertrand et al., 2002). The correct spatial and temporal distribution of proneural genes in the developing brain is crucial for the specification of different neuronal subtypes (Bertrand et al., 2002).

In vertebates, the proneural proteins are members of the class II bHLH transcription factors and accordingly, their activity can be modulated by the expression of cell intrinsic transcription factor networks, E-proteins, ID proteins and HES proteins, to coordinate cortical development (Ross et al., 2003). In agreement with this model, TCF4 is known to heterodimerise with a variety of proneural proteins including neurogenic differentiation 1 (NEUROD1), neurogenic differentiation 2 (NEUROD2), achaete-scute complex homolog 1 (ASCL1), achaete-scute complex homolog 3 (ASCL3) and atonal homolog 1 (ATOH1) (Persson et al., 2000; Yoshida et al., 2001;

Flora et al., 2007; Ravanpay and Olson, 2008; Brzozka et al., 2010). TCF4 is also able to heterodimerise with other bHLH family members with important roles in neurodevelopment such as the ID proteins (ID1-4), HES1 and heart and neural crest derivatives expressed transcript 1 (HAND1, dHAND) (Jogi et al., 2002; Liu et al., 2004; Murakami et al., 2004).

In mice, *Ascl1* is expressed in the ventral telencephalon, a region of the brain that gives rise to GABAergic and cholinergic neurons (Guillemot and Joyner, 1993; Wilson and Rubenstein, 2000). In *Ascl1* knockout mice, the loss of proneural activity leads to the loss of neural progenitor cells and a decrease in GABAergic interneurons in the cortex (Casarosa et al., 1999). *Ngn2* expression is complimentary to that of *Ascl1* and is distributed over the dorsal telencephalon. Absence of Ngn2 and Ascl1 exacerbates the depletion of neuronal progenitors originating in the ventral and dorsal regions of the telencephalon demonstrating their combined importance in early neural fate commitment (Fode et al., 2000).

Neurod1 is a downstream mediator of Ngn2 and is expressed in immature and differentiating neurons (Lee et al., 2000). Consistent with this expression pattern, Neurod1 and Neurod2 are required for the terminal differentiation of granule cells in the dentate gyrus and cerebellum (Miyata et al., 1999; Schwab et al., 2000). *Neurod1* expression persists in adulthood where it has a role in postnatal neurogenesis and survival of hippocampal and olfactory neurons (Gao et al., 2009; Boutin et al., 2010). Studies on adult mouse brain have established that Tcf4 and Neurod1 physically interact in the hippocampus and cerebellum (Brzozka et al., 2010).

Thus, through interactions with proneural proteins, TCF4 may participate in the early stages of neuronal cell commitment or later in neuronal differentiation. TCF4 transcripts are also found in regions undergoing adult neurogenesis such as the hippocampus and olfactory bulb. This expression pattern coupled to its role in regulating proneural function, suggests that TCF4 may play a part in prenatal neurodevelopment in addition to postnatal neuron generation.

1.9.2 Tcf4 Mouse models

Mouse knockout studies have been instrumental in determining the function of bHLH transcription factors during development (Casarosa et al., 1999; Lyden et al., 1999; Miyata et al., 1999; Fode et al., 2000). Studies on the proneural genes have established the bHLH family as critical regulators of neurogenesis, however the role of Tcf4 and other E-proteins in this process is poorly understood (Ik Tsen Heng and Tan, 2003). This is presumably because E-proteins are to a certain extent, functionally redundant and individual E-protein knockout mice have no gross neurological phenotype (Flora et al., 2007; Ravanpay and Olson, 2008). In spite of this, it has become clear that E-proteins have non-redundant unique properties in particular contexts that will require further detailed characterisation (Zhuang et al., 1998).

Homozygous *Tcf4* knockout (*Tcf4*$^{-/-}$) mice die postnatally from undetermined causes, suggesting that *Tcf4* has a crucial role in normal development (Zhuang et al., 1996). While the brains of *Tcf4*$^{-/-}$ mice have no overt morphological abnormalities, specific defects in Atoh1-dependent hindbrain structures have been identified (Flora et al., 2007). As mentioned above, Atoh1 is a proneural protein that forms transcriptionally active complexes with Tcf4. Both Atoh1 and Tcf4 are expressed in the rhombic lip, a region of the developing hindbrain that matures to form the granule cells of the cerebellum and five distinct brainstem nuclei (Flora et al., 2007). To determine the function of Tcf4 in Atoh1-dependent neural progenitors, heterozygous Atoh1 knockout (*Atoh1*$^{+/-}$) mice were crossed with viable heterozygous Tcf4 mutant (*Tcf4*$^{+/-}$) mice. On the *Atoh1*$^{+/-}$ background, *Tcf4*$^{-/-}$ mice display disrupted development of the pontine nucleus (PN) – one of the five brainstem nuclei specified by precursor cells in the rhombic lip (Flora et al., 2007). Here, *Tcf4* deletion causes a reduction in the number of neurons and an accumulation of ectopic neurons that fail to migrate to their correct location in the pontine nucleus. Heterozygous *Tcf4* knockout mice display less severe migratory deficits than the double mutant, demonstrating a dose dependent response to Tcf4 levels. The defects in hindbrain development are remarkably specific, as the development of other Atoh1-dependent structures remains apparently intact. The Ascl1-dependent development of the *locus coeruleus* is also normal in *Tcf4*$^{-/-}$ mice, suggesting that Tcf4 can specify very distinct progenitor cells

by dimerising with partner bHLH proteins (Flora et al., 2007). Importantly, these deficits are highly specific to Tcf4, as pontine development in *Tcf3* and *Tcf12* knockout backgrounds were normal (Flora et al., 2007). The requirement of Tcf4 heterodimers to specify particular neuronal precursors reveals the unique functions of Tcf4 in the brain and suggests that delineation of context specific Tcf4 interactions will be essential to understand its function in neurogenesis.

A mouse model over-expressing *Tcf4* has also been generated to explore the function of *Tcf4* in the adult brain (Brzozka et al., 2010). The transgenic Tcf4 mouse (*Tcf4tg*) has an approximately 1.5 fold increase in *Tcf4* expression in the forebrain structures of the hippocampus, cortex and amygdala. In behavioural tests, *Tcf4tg* mice showed no defects in exploratory, motor, and motivational behavior however they had significantly reduced freezing in the contextual and cued fear memory tests (Brzozka et al., 2010). Deficits in higher order cortical processing measured by trace fear conditioning were also observed in *Tcf4tg* mice (Brzozka and Rossner, 2013). Furthermore, the mice displayed impaired sensorimotor gating measured by pre-pulse inhibition (PPI); a common behavioural endophenotype found in animal models of schizophrenia and patients (Braff and Geyer, 1990; Brzozka et al., 2010). Examination of gene expression changes in the hippocampus identified increased expression of the circadian clock gene *Per2* and *Id2* that is known to regulate Tcf4 function (Brzozka et al., 2010). This mouse model demonstrates that the precise control of Tcf4 dosage is critical for certain aspects of cognitive processing.

1.10 Pitt-Hopkins Syndrome

1.10.1 Genetics of PTHS

Pitt-Hopkins syndrome (PTHS) was first described in 2 unrelated children with severe intellectual disability (ID), similar facial features and episodes of hyperventilation followed by apnea and cyanosis (Pitt and Hopkins, 1978). Very few cases of PTHS were reported until two groups independently discovered that haploinsufficiency of *TCF4* was the cause of the disorder (Amiel et al., 2007; Zweier et al., 2007). Determining the genetic cause of the PTHS has greatly increased the recognition of the disorder. 112 diagnosed patients are currently reported in the literature that have a wide spectrum of *de novo* heterozygous mutations including partial and whole gene deletions and a variety of point mutations (Table 1.1) (Amiel et al., 2007; Brockschmidt et al., 2007; Zweier et al., 2007; Giurgea et al., 2008; de Pontual et al., 2009; Takano et al., 2010; Lehalle et al., 2011; Marangi et al., 2011; Whalen et al., 2012b). The patients range from 20 months to 32 years of age, with a mean age of 11 years (Whalen et al., 2012b). Inherited cases of PTHS are rare but have been described in individuals where one of the parents was a somatic mosaic for a *TCF4* mutation (de Pontual et al., 2009; Kousoulidou et al., 2013; Steinbusch et al., 2013). Mosaic status can impact on disease severity for certain mutation types (Rossi et al., 2012; Kousoulidou et al., 2013; Steinbusch et al., 2013). PTHS is caused by loss-of-function of a singe *TCF4* allele, leading to haploinsufficiency of the gene product. In agreement with this hypothesis, frameshift, missense and nonsense mutations found in patients induce similar defects in TCF4-mediated transactivation on luciferase reporter constructs (Zweier et al., 2007; de Pontual et al., 2009). However, missense mutations affecting the bHLH domain appear to be more damaging then those outside the bHLH domain in *in vitro* assays (Sepp et al., 2012b).

Table 1.1 Range of *TCF4* mutations that cause PTHS. The number of patients with each mutation type and its corresponding percentage is presented in the table. This table was modified from data published by Whalen et al. 2012.

Mutation type	Number of patients	% of patients
Deletions		
Whole gene deletion	23	**21**
Partial gene deletion	10	**9**
Point mutations		
Frameshift	33	**29**
Nonsense	18	**16**
Splice site	6	**5**
Missense	22	**20**
Total	112	**100**

1.10.2 Clinical symptoms

A constellation of symptoms has been described in PTHS patients but few genotype-phenotype correlations have been documented (Table 1.2) (Whalen et al., 2012b). The most prominent features are severe intellectual disability associated with distinctive facial features, developmental delay, stereotypic movements, absent speech and breathing abnormalities. Epilepsy is reported in nearly 40% of cases (Table 1.2). Initially, a single report documented a higher incidence of epilepsy in patients with missense mutations (7/8, 88%) compared to other mutations types (6/36, 17%) (Rosenfeld et al., 2009). However, this relationship remains to be confirmed since more patents are being identified that have epilepsy without missense mutations (Marangi et al., 2011). A study of 10 cases found that many of the behavioural abnormalities exhibited by PTHS patients could be associated to autism spectrum disorders, such as severe impairments in communication and language, difficulties in social engagement, fascinations with specific objects, and intense motor stereotypies (Van Balkom et al., 2012). The facial gestalt in PTHS is recognisable clinically by a wide mouth with prominent Cupid's bow upper lip, a broad and beaked nasal bridge and flaring nostrils, deep-set eyes and mildly cup-shaped, fleshy ears (Zweier et al., 2007). Non-neurological symptoms are also very common; PTHS patients often present with constipation and/or ocular anomalies such strabismus, myopia, and astigmatism. Particularities of the hands and feet are noticeable and have been suggested as a useful diagnostic criterion (Lehalle et al.,

2011). Patients are mostly within the normal ranges of height and weight (± 2 SD) although of these patients, 81% fall at the lower end of normal growth parameters (0 to –2 SD) (Whalen et al., 2012b). Finally, a number of brain abnormalities have been detected in PTHS patients using magnetic resonance imaging (MRI) (Table 1.3). The most common irregularity is agenesis or hypoplasia of the corpus callosum (40%) however decreased hippocampal volume or enlarged ventricles are recorded in over 30% of patients. The structural abnormalities observed by MRI strongly support a role for TCF4 in regulating normal brain morphology and development.

Table 1.2 Neurological and behavioural abnormalities in PTHS patients. A number of characteristic symptoms have been described in PTHS; the most common (>30%) symptoms are presented in the table. This table was modified from data published by Whalen et al. 2012.

Clinical Feature	Observed	Total	%
Psychomotor development			
Severe ID or DD	31	31	**100**
Hypotonia	19	26	**81**
Delayed walking	24	24	**100**
Walking achieved	21	30	**47**
Ataxic gait	18	23	**60**
Absent language or <10 words	32	32	**100**
Stereotypic movements			
Positive history	29	31	**77**
Arm flapping	20	25	**62**
Hand nibbling/biting	13	24	**45**
Movement of fingers	11	23	**48**
Movement of wrists	10	21	**48**
Hand washing	9	20	**57**
Head stereotypies	9	22	**42**
Median line	11	21	**62**
Behavior			
Smiling appearance	25	28	**91**
Harm to self	15	25	**52**
Anxiety	17	21	**57**
Agitation	23	26	**68**
Agitation outbursts	15	21	**71**
Unmotivated laughing	17	27	**55**

Breathing anomalies			
Hyperventilation	19	31	**56**
Apnoea	13	28	**54**
Cyanosis	6	28	**32**
Seizures			
History of seizures	6	30	**39**

Table 1.3 MRI brain abnormalities exhibited by PTHS patients. Patients with PTHS have been scanned by magnetic resonance imaging (MRI) to detect specific abnormalities in brain structure. The number of patients affected by each abnormality is presented. This table was modified from data published by Whalen et al. 2012.

Cerebral MRI	Observed	Total	%
Normal	30	85	**35**
Corpus callosum: hypoplasia or agenesis	34	85	**40**
Temporal lobe hyperintensity	13	35	**37**
Small hippocampi	15	44	**34**
Ventricular dilatation	20	64	**31**
Minor posterior fossa abnormalities (other than dentate nuclei)	7	25	**28**
Cortical atrophy	4	21	**19**
Dentate nuclei hyperintensity	4	24	**17**
Abnormal myelinisation / reduced white matter volume	4	32	**13**
Frontal lobe hypoplasia	4	42	**10**
Large caudate nuclei	4	46	**9**

Despite PTHS symptomatology having strongly associated features, confirmation of the diagnosis is impeded by the phenotypic similarities with Rett (RTT; MIM# 312750), Angelman (AS; MIM# 105830) and Mowat-Wilson syndromes (MWS; MIM# 235730) that share the traits of severe intellectual disability, absent speech, delayed walking, epilepsy, microcephaly and a happy disposition (Marangi et al., 2011; Armani et al., 2012; Whalen et al., 2012b). In a group of 86 patients suspected to have AS but with normal UBE3A (ubiquitin protein ligase E3A) sequencing and methylation analysis, two percent were found to have *TCF4* mutations (Takano et al., 2010). PTHS has also been recommended as differential diagnosis for α-thalassemia X-linked intellectual disability syndrome (ATRX; MIM# 301040) (Takano et al., 2011). Finally, TCF4 mutations have been identified in a few patients with milder,

non-syndromic neurodevelopmental disorders, indicating that *TCF4* mutations may be associated with a broader clinical phenotype than is currently recognised (Kalscheuer et al., 2008; Rauch et al., 2012; Talkowski et al., 2012; Hamdan et al., 2013).

1.11 Schizophrenia

1.11.1 Schizophrenia and neurodevelopment

Schizophrenia is a severe, life-long, psychiatric disorder that usually manifests in early adulthood, and is characterised by auditory hallucinations, delusions, cognitive deficits, and affective retraction. The disorder is estimated to affect up to 1% of the world population and is associated with substantial morbidity, mortality and societal costs (Knapp et al., 2004; Saha et al., 2007). Psychological, pharmacological and genetic studies have advanced our general understanding of the etiology of schizophrenia (Keshavan et al., 2008; Tandon et al., 2008). The prevailing hypothesis is that schizophrenia is a neurodevelopmental disorder induced by genetic and environmental risk factors (Lewis and Levitt, 2002). Twin studies estimate that genetics could account for up to 80% of the disease susceptibility (Sullivan et al., 2003). Environmental risks come from a potentially diverse range of sources that include biological, psychological, social, and physical stresses compounding genetic susceptibility during development (Tsuang, 2000). The neurodevelopmental hypothesis is supported by lack of neurodegenerative processes during the course of the disease, and by the fact that affected individuals show cognitive and social impairment before the first episode of the disease (Lewis and Levitt, 2002). Early neurodevelopmental defects caused by genetic and/or environmental insults may be responsible for an altered developmental trajectory causing brain dysfunction (Insel, 2010). In addition, later developmental maturation processes may be defective, causing brain dysconnectivity through aberrant synaptic pruning and axonal myelination (Insel, 2010). Together, the neurodevelopmental deficits are thought to alter brain circuitry, connectivity and synaptic plasticity resulting in symptoms of schizophrenia (Stephan et al., 2009).

1.11.2 Schizophrenia genetics

Schizophrenia is complex and heterogeneous genetic disorder (Rodriguez-Murillo et al., 2012). The high heritability of schizophrenia is well established although the search for robust genetic associations has, until recently, yielded few compelling results. Two major models for the genetic basis of the disorder have been proposed, which differ fundamentally in their conception of the genetic architecture and heritability of the disorder (Mitchell and Porteous, 2011). The most common

conceptualisation of the disorder is the common disease–common variant (CDCV) model of schizophrenia (Lohmueller et al., 2003). This model proposes that schizophrenia is caused by the inheritance of multiple genetic variants of small effect that are common in the population. The additive effect of each variant causes a genetic burden of susceptibility to the disease. In contrast, the multiple rare variants model proposes that each case of schizophrenia is caused by a single rare variant of large effect (McClellan et al., 2007). In this model, rare variants can affect different genes in different families or individuals making the disorder genetically heterogeneous. Neither model excludes the contribution from environment factors.

Human genetic studies of schizophrenia have so far provided evidence for both rare and common variants conferring risk to disease suggesting that each can contribute to the genetic etiology (Sullivan et al., 2012). Large case-control genome-wide association (GWA) studies of schizophrenia have begun to uncover common disease variants of small effect sizes (odds ratio (OR) < 1.3) (Stefansson et al., 2009; Ripke et al., 2011; Steinberg et al., 2011; Ripke et al., 2013). Focusing on results that achieve genome-wide significance in large case-control studies ($P < 5x10^{-8}$), 14 schizophrenia susceptibility loci have been discovered to date including variants associated to the *MHC*, *MIR137*, *ZNF804A* and *TCF4* genes (Sullivan et al., 2012). Consistent with a CDCV model, up to 23% of variation in susceptibility to schizophrenia is estimated to be conferred by common single nucleotide polymorphisms (SNPs) (Lee et al., 2012). Schizophrenia is also predicted to have high degree of shared genetic etiology with bipolar disorder (68%) and major depressive disorder (48%) and a lower concordance with autism (16%), as measured by common SNPs (Cross-Disorder Group of the Psychiatric Genomics et al., 2013).

In contrast, several studies have identified an important role for rare variants in the etiology of schizophrenia; in the form of single or multi- gene deletions and duplications (copy number variants, CNVs) (International Schizophrenia, 2008; Walsh et al., 2008; Levinson et al., 2011; Mitchell and Porteous, 2011). Currently, eight rare copy number variants of strong effect (odds ratio between 4–20) and with consistent replication have been described for schizophrenia, including the most highly associated regions of 22q11.21 (OR 20.3) and 16p11.2 (OR 9.5) (Sullivan et al., 2012). However, rare mutations that confer high risk of schizophrenia can also

cause other neurodevelopmental disorders such as bipolar disorder, major depression, autism, attention deficit hyperactivity disorder (ADHD), epilepsy and mental retardation (Sebat et al., 2009). The shared genetic risk factors between multiple neuropsychiatric disorders suggest that they have a common pathophysiology. Functional gene set analyses of common and rare variants have determined that susceptibility genes converge on diverse biological networks such as chromatin remodeling, cell adhesion, axon guidance, neuronal cell mobility and synaptic signaling, supporting a neurodevelopmental basis for the disorder (Walsh et al., 2008; O'Dushlaine et al., 2011; Gilman et al., 2012; Lips et al., 2012). The most recent advance in schizophrenia genetics has focused on trying to find single nucleotide variants (SNVs) in schizophrenia patients using exome-sequencing technologies (Need et al., 2012). This high-throughput methodology is expected to deliver a new source of genetic variability with moderate effect sizes, however large studies and careful interpretation of the genetic data will be required to identify causal variants with confidence (Piton et al., 2013).

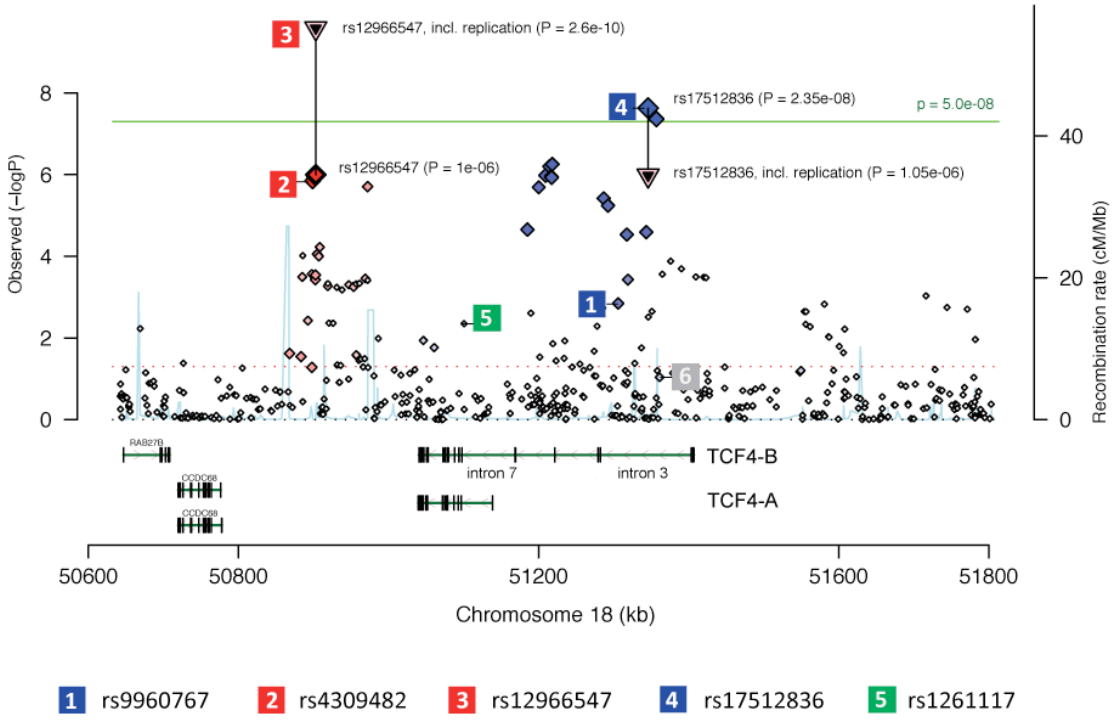
The widespread use of genomic microarray-based SNP genotyping and next-generation sequencing technologies has also paved the way for the discovery *de novo* CNVs and SNVs in schizophrenia (Kirov et al., 2012; Xu et al., 2012). *De novo* mutations have the potential to be highly damaging, as they represent a sporadic source of rare genetic variation that has not been confronted with purifying, natural selection (Veltman and Brunner, 2012). It has already been proposed that patients with schizophrenia carry a higher burden of *de novo* mutations than unaffected relatives indicating that these rare mutations participate in the genetic etiology of the disorder (Girard et al., 2011; Xu et al., 2011; Xu et al., 2012).

1.11.3 TCF4 and the genetic association to schizophrenia

As mentioned above, TCF4 is one of the few genes robustly associated with schizophrenia through common genetic variation (Sullivan et al., 2012). Three independent genetic loci in and near *TCF4* have been associated with schizophrenia (Figure 1.5). The initial discovery identified SNP rs9960767, located in intron 3 of *TCF4,* as genome wide significant ($P = 4.1\times10^{-9}$; OR 1.23) in a large meta-analysis of 12,945 schizophrenia cases 34,591 controls (Stefansson et al., 2009) (Figure 1.5).

A second variant, downstream of the *TCF4* gene (rs4309482) reached genome-wide significance ($P = 7.8\times10^{-9}$; OR 1.09) in a follow-up study with 18,206 cases and 42,536 controls (Figure 1.5) (Steinberg et al., 2011). Two more variants were identified in a GWAS mega-analysis from the Psychiatric Genetics Consortium (PGC). rs12966547 was genome-wide significant in the combined data set of 17,839 schizophrenia cases and 33,859 controls ($P = 2.6\times10^{-10}$; OR 1.09) whereas rs17512836 (Figure 1.5) was significant in the initial stage with 9,394 cases and 12,462 controls ($P = 2.35 \times 10^{-08}$; OR 1.40) (Ripke et al., 2011). A fifth SNP (rs1261117), in intron 8 of *TCF4*, was the most significant variant in a family based replication study of GWAS results ($P = 2.53\times10^{-10}$, OR 1.60) (5 green, Figure 1.5) (Aberg et al., 2013). Although five *TCF4*-associated SNPs have been identified as genome-wide significant in different studies, SNPs cluster into three independent genetic loci ($r^2<0.1$) due to linkage disequilibrium (Figure 1.5). The SNP cluster downstream of *TCF4* (Figure 1.5) is also located near *CCDC68* (coiled-coil domain containing 68), a gene with unknown function. Interestingly, SNP rs17512836 (Figure 1.5) nearly reached genome wide significance (1.05×10^{-6}) in a large cross-disorder GWAS investigating shared risk variants between autism spectrum disorder (ASD), ADHD, bipolar disorder, major depressive disorder, and schizophrenia. Modelling analysis revealed that the best-fit model for the rs17512836 SNP encompassed both schizophrenia and ASD, suggesting that *TCF4* could also contribute to ASD. The functional effects of SNPs surrounding *TCF4* are currently unknown however many SNPs associated with schizophrenia risk appear to affect gene expression (Bacanu et al., 2013).

Figure 1.5 Distribution of genome-wide significant SNPs in the *TCF4* gene. The location of genome-wide significant SNPs from GWAS of schizophrenia is illustrated (1-5). The graphic displays data from the Psychiatric Genetics Consortium study (Ripke et al., 2011), however genome-wide significant SNPs from other studies have been annotated (Stefansson et al., 2009; Steinberg et al., 2011). SNPs labeled with different coloured numbers represent genetically independent loci ($r^2 < 0.1$). The r^2 values and the GWAS graphic were obtained from the Broad Institute website (Section 2.10.1).



1.11.4 TCF4 risk alleles and cognitive endophenotypes

The impact of schizophrenia-associated *TCF4* risk variants on cognition and information processing has been analysed by neuropsychological testing. In a sample of 401 schizophrenia patients, the *TCF4* variant rs9960767 was demonstrated to influence verbal memory in the Rey Auditory Verbal Learning Test (Lennertz et al., 2011). Impaired verbal memory is among the most prominent cognitive deficits in schizophrenia patients (Toulopoulouand and Murray, 2004). Contrary to expectations, carriers of the schizophrenia risk allele (C-allele) showed better recognition in verbal memory, suggesting a role of this variant in the development of memory-related brain structures. *TCF4* genotype did not impact on various other cognitive tests measuring the domains of attention and executive function.

The effect of the *TCF4* variant rs9960767 on sensorimotor gating was also assessed in 107 healthy volunteers and a schizophrenia spectrum group with 113 patients (Quednow et al., 2011). Sensory gating is an essential psychological process that allows the filtering of sensory information during cognitive tasks (Cromwell et al., 2008). This mechanism is disrupted in several psychiatric disorders and is considered a schizophrenia endophenotype (Braff and Geyer, 1990). PPI is a translational task that can be used to measure sensorimotor gating in humans and animals (Braff and Geyer, 1990). In PPI, subjects presented with a weak pre-stimulus (pre-pulse) are able to inhibit their reaction to a strong stimulus (Li et al., 2009). Carriers of the schizophrenia risk variant in both the healthy control and patient groups produced a statistically significant reduction in PPI, indicative of an impaired inhibitory response (Quednow et al., 2011). Remarkably, PPI is also reduced in transgenic mice that over-express TCF4 (Brzozka et al., 2010) (Section 1.9.2). A follow up study measured auditory sensory gating assessed by P50 suppression of the auditory evoked potential (Quednow et al., 2012). P50 suppression has been related to attentional performance, working memory, and behavioral inhibition and is another measure of gating function (Lijffijt et al., 2009). In this large multi-center study, 1,821 healthy volunteers were genotyped for 21 different *TCF4* polymorphisms. The schizophrenia risk alleles for four polymorphisms (rs9960767, rs10401120rs, rs17597926, and 17512836) were associated with a highly significant reduction in P50 suppression. This decrease in P50 suppression was more pronounced in smokers than in non-smokers, suggesting an interaction between *TCF4* risk alleles and smoking behaviour on cognitive functioning (Quednow et al., 2012). These studies indicate that *TCF4* may influence key mechanisms regulating information processing that could contribute to the cognitive deficits and other endophenotypes observed in schizophrenia.

1.12 Thesis aims and structure

The importance of TCF4 in brain developmental and function is apparent from its genetic association to debilitating neurodevelopmental disorders. However, the function of TCF4 in the nervous system is largely unknown. Accordingly, our understanding of TCF4-mediated pathophysiology in PTHS and schizophrenia is scarce. Preliminary evidence from mouse models and psychological testing suggests that TCF4 has a role in neurodevelopment and cognitive function, but the molecular basis underlying these processes has not yet been addressed. Consequently, the objective of this study was to gain further insight into the molecular and cellular function of TCF4 in order to better understand TCF4-associated disease mechanisms in PTHS and schizophrenia. Experiments in this thesis were designed to address the following aims:

1) To determine the effect of PTHS-associated missense mutations on the function of TCF4 (Chapter 3).

2) To understand the effects of TCF4 on the cellular transcriptome in SH-SY5Y cells (Chapter 4).

3) To characterise endogenous TCF4 isoforms and identify novel TCF4-interacting proteins (Chapter 5),

4) To understand the effects of different TCF4 isoforms on gene expression in SH-SY5Y cells (Chapter 6).

The novel findings presented in this thesis are summarised and discussed in Chapter 7.

Chapter 2

Materials and Methods

2.1 General

All chemicals used for experiments were of analytical grade and were purchased from Fisher or Sigma-Aldrich unless otherwise stated. All molecular biology reagents were purchased from New England Biolabs (NEB) and Promega unless otherwise stated. All kits were used as per manufacturer's instructions.

2.2 Cell strains and media

2.2.1 Bacterial cell culture

Plasmid propagation was performed using *Escherichia coli* (*E.coli*) XL1-Blue cells (Stratagene), genotype: *recA1 endA1 gyrA496 thi-1 hsdR17 supE44 relA1 lac [F'proAB lacI^{q}ZΔM15 Tn10 (Tet^r)]*. For expression of fusion proteins with the pET-32 (Novagen) and pGEX-4T (Promega) vector constructs, the *Escherichia coli* BL21-Gold (DE3) strain was used (Stratagene), genotype: *E.coli B F$^-$ ompT hsdS($r_B^-m_B^-$) dcm$^+$ Tet^r gal λ (DE3) endA The.* Bacterial cells were grown in Luria-Bertani (LB) media (liquid culture) and on LB-agar media (plate culture) (Tryptone (pancreatic digest of casein) 10 g/l, yeast extract 5 g/l, NaCl 5 g/l and agar 15 g/l). Media was autoclaved for 30 min at 115^{o}C and cooled to below 50^{o}C prior to addition of the appropriate antibiotic. The final concentrations of antibiotics used were: ampicillin 100 μg/ml (LB-amp), kanamycin 30 μg/ml (LB-kan), tetracycline 100 μg/ml (LB-tet). In liquid culture, *E. coli* strains were grown for 16-20 h at 37°C in LB with shaking at 200 rpm for all recombinant DNA methods. In solid culture, E. coli strains were grown on LB plates for 16-20 h and stored at 4°C for up to one month. Before picking colonies, the LB plates were removed from storage one hour before use, inverted and allowed to dry. For long term storage of *E. coli* strains, fresh cultures in

mid-log phase (optical density measured at 600 nm (OD600) = 0.6-0.8) were frozen in 15% (v/v) glycerol and stored at -80°C.

2.2.2 Mammalian cell culture

COS-7 (African green monkey kidney), HEK293T (human embryonic kidney 293T) and SH-SY5Y (neuroblastoma) cell lines were purchased from the European Collection of Cell Cultures (ECACC) and the American Type Culture Collection (ATCC). The 9E10 (anti-myc) hybridoma and 12G10: a-tubulin (isotype IgG1) cell lines were obtained from the Developmental Studies Hybridoma Bank (DSHB) at the University of Iowa. All cell lines were cultured in Dulbecco's modified Eagle medium (DMEM, Invitrogen) supplemented with 10% (v/v) foetal calf serum (FCS, PAA laboratories) and 1% (v/v) penicillin/streptomycin except for siRNA knockdown experiments were antibiotic-free medium was used. All cells were grown as adherent cultures in T175 cm^2 tissue culture flasks containing 30ml of cell culture medium. Cultures were maintained at 37^oC with 5% CO_2 in a humidified incubator. Cells were passaged at 80% confluence to ensure the cultures remained healthy. Cells were washed twice with 10 ml $Ca^{2+}$and Mg^{2+} free Hanks balanced salt solution (HBSS, Sigma) and detached with 3 ml 1× trypsin-EDTA (Sigma) at 37^oC for 2-5 min. Trypsin was inactivated with addition of 7ml growth media followed by collection of cells by centrifugation at 2000 rpm for 3 min at room temperature. The cell pellet was resuspended in 10 ml fresh media and reseeded at an appropriate density. For medium term storage of cell lines the cell pellet was resuspended in freezing media (90% (v/v) FCS, 10% (v/v) DMSO) and aliquoted in cryovials at $3x10^6$ cells/vial, followed by storage at -80^oC.

2.3 Molecular biology

2.3.1 Plasmid Vectors

Name	Description	Antibiotic Selection	Supplier
pCMV-Myc	**Myc fusion protein**. Mammalian expression vector that expresses proteins containing an N-terminal c-myc epitope tag.	Ampicillin	Clontech
pCMV-HA	**HA fusion protein**. Mammalian expression vector that expresses proteins containing an N-terminal hemagglutinin epitope tag.	Ampicillin	Clontech
pEYFP-C1	**GFP fusion protein**. Mammalian expression vector that expresses proteins as an N-terminal	Kanamycin	Clontech

	fusion of enhanced yellow fluorescent protein.		
pGL3-Basic	**Luciferase reporter construct**. contains a modified coding region for the firefly (Photinus pyralis) luciferase enzyme. Regulatory DNA sequences can be cloned upstream to monitor transcriptional activity in eukaryotic cells.	Ampicillin	Promega
pGEX-4T1	**GST fusion protein**. *E.coli* expression vector that utilises the *tac* promoter to control expression of cloned insert. Expression is induced upon addition of IPTG in any *E.coli* host though preferably BL21. The cloned insert is expressed with a glutathione S-transferase (GST) tag.	Ampicillin	Promega
pET-32	**Thioredoxin and His-tagged fusion protein**. E.coli expression vector that utilises the T7 lac promoter to control the expression of cloned insert. Expression is induced upon addition of IPTG in the host strain BL21. The cloned insert is expressed with N-terminal thioredoxin (TRX) and 6 x histidine (His) fusion tags. (	Ampicillin	Novagen

2.3.2 Plasmid construction

cDNA sequences were amplified from the ProQuest™ Human Fetal Brain cDNA Library (Invitrogen) and mutagenised where necessary (Section 2.3.4). All constructs were created according to cloning methods described in Sections 2.4.2, 2.4.3, 2.4.7. The primers used for construct generation are provided in Appendix II. The same full length TCF4 sequence was used for the GFP-TCF4, myc-TCF4 and HA-TCF4 constructs (ENST00000565018). The HA-NEUROD1 and HA-ID2 constructs were prepared by Mrs. M. Doyle (DJ Blake Lab, Cardiff University). The HA-ASCL1, GFP-G358V, GFP-D535G, GFP-R578W and GFP-A614V constructs were prepared by Mrs. R. Chapman (DJ Blake Lab, Cardiff University).

Name	Insert	Vector	Restriction sites
GFP-TCF4	Full length human TCF4-B$^+$ (671 amino acids)	pEYFP-C1	*EcoRI*/*SalI*
GFP-G358V	GFP-TCF4 with amino acid number 358/671 converted from glycine to valine	pEYFP-C1	*EcoRI*/*SalI*
GFP-D535G	GFP-TCF4 with amino acid number 535/671 converted from serine to glycine	pEYFP-C1	*EcoRI*/*SalI*
GFP-R578P	GFP-TCF4 with amino acid number 576/671 converted from arginine to proline	pEYFP-C1	*EcoRI*/*SalI*
GFP-R580W	GFP-TCF4 with amino acid number 580/671 converted from arginine to tryptophan	pEYFP-C1	*EcoRI*/*SalI*
GFP-A614V	GFP-TCF4 with amino acid number 614/671 converted from alanine to valine	pEYFP-C1	*EcoRI*/*SalI*
myc-TCF4	Full length human TCF4-B$^+$ (671 amino acids)	pCMV-Myc	*EcoRI*/*NotI*
myc-G358V	myc-TCF4 with amino acid number	pCMV-Myc	*EcoRI*/*SalI*

	358/671 converted from glycine to valine		
myc-D535G	myc-TCF4 with amino acid number 535/671 converted from serine to glycine	pCMV-Myc	*EcoRI*/*SalI*
myc-R578P	myc-TCF4 with amino acid number 576/671 converted from arginine to proline	pCMV-Myc	*EcoRI*/*SalI*
myc-R580W	myc-TCF4 with amino acid number 580/671 converted from arginine to tryptophan	pCMV-Myc	*EcoRI*/*SalI*
myc-A614V	myc-TCF4 with amino acid number 614/671 converted from alanine to valine	pCMV-Myc	*EcoRI*/*SalI*
HA-TCF4	Full length human TCF4-B^+ (671 amino acids)	pCMV-HA	*EcoRI*/*NotI*
HA-ASCL1	Full length human ASCL1 (ENST00000266744)	pCMV-HA	*EcoRI*/*KpnI*
HA-ATOH1	Full length human ATOH1 (ENST00000306011)	pCMV-HA	*EcoRI*/*SalI*
HA-NEUROD1	Full-length human NEUROD1 (ENST00000295108)	pCMV-HA	*SalI*/*NotI*
HA-ID2	Full-length human ID2 ((ENST00000331129)	pCMV-HA	*SalI*/*NotI*
GST-TCF4	Amino acids 361-554 of the full length TCF4-B^+ sequence (exons 14-18)	pGEX_4T2	*EcoRI*/*SalI*
TRX-TCF4	Amino acids 361-554 of the full length TCF4-B^+ sequence (exons 14-18)	pET_32c	*EcoRI*/*SalI*
pcDNA-E47	The pcDNA-E47 plasmid was a kind gift from Dr.Carme Gallego, Molecular Biology Institute of Barcelona, Barcelona.	pcDNA3.1	*EcoRI*/*EcoRI*
myc-E47	Full length E47 cDNA cloned from the pcDNA-E47 plasmid	pCMV-Myc	*EcoRI*/*SalI*
pLuc-CNTNAP2_F	2434bp of the CNTNAP2 promoter (5' to 3') (chr7:145811615-145814049)	pGL3-Basic	*HindIII*/*HindIII*
pLuc-CNTNAP2_R	2434bp of the CNTNAP2 promoter (3' to 5') (chr7:145811615-145814049)	pGL3-Basic	*HindIII*/ *HindIII*
pLuc-NRXN1β _F	599bp of the NRXN1β promoter (5' to 3') (chr2:50574779-50575378)	pGL3-Basic	*HindIII*/ *HindIII*
pLuc-NRXN1β _R	599bp of the NRXN1β promoter (3' to 5') (chr2:50574779-50575378)	pGL3-Basic	*HindIII*/ *HindIII*

2.3.3 Polymerase chain reaction (PCR)

For polymerase chain reaction (PCR), oligonucleotide primers were designed complementary to 18-24 base pairs of the target sequence with an estimated melting temperature of 60°C. PCR primers are listed in Appendix II. When required, restriction endonuclease sites were added to the 5' end of the primer sequence. To increase the efficiency of restriction enzyme cleavage, three base pairs were included at the 5' end of the restriction enzyme recognition sequence. PCR was routinely performed using Red Taq (Sigma). When high fidelity PCR was required, Easy A (Stratagene) was used. A typical PCR reaction with cycling parameters is provided below. As an optimisation step, DMSO was optionally added to a final concentration of 5% (v/v). PCRs were performed using a GeneAmp PCR system 9700 (Applied

Biosystems) or a PTC-220 DNA Engine Dyad Thermal Cycler (MJ Research Inc.). PCR products were separated by gel electrophoresis using 1-3% agarose gels as described in Section 2.3.6.

A typical PCR reaction with cycling parameters consisted of the following:

Reagent	Volume (µl)	Final concentration	Program
10× PCR buffer	5	1×	1. 95°C for 5 min
Forward primer (10µM)	1	0.2µM	2. 95°C for 30 s
Reverse primer (10µM)	1	0.2µM	3. 56-62°C for 30 s
Deoxynucleotide mix (10mM)	1	200µM	4. 72°C 1 kb / min
Template (10ng/ul)	1-5	10-50ng	5. cycle steps 2-5 20-40 times
DNA polymerase (5 U/µl)	1	5U	6. 72°C for 7 min
Molecular grade water	Up to 50		7. hold at 4°C

2.3.4 Site-directed mutagenesis

Site-directed mutagenesis was used to introduce mutations into cDNA plasmid constructs. The primer design and protocol was based on the Quick-change site-directed mutagenesis system and utilised the Pfu-Ultra high-fidelity DNA polymerase (Stratagene). A typical mutagenesis reaction with cycling parameters is provided below. 10 µl of reaction product was run on a 1% (w/v) agarose gel to confirm amplification of the DNA. 1 µl of *DpnI* restriction enzyme was added to the remaining reaction mix and incubated for 2 h at 37^{o}C to digest the methylated, double stranded non-mutated DNA template. 2-5 µl of the digest product was used to transform 50-100 µl of chemically competent XL1-Blue cells using the heat shock method (Section 2.4.6). Multiple colonies were picked for plasmid purification (Miniprep) and the DNA was subsequently sequenced to confirm successful mutagenesis (Section 2.4.7).

A typical mutagenesis reaction with cycling parameters consisted of the following:

Reagent	Volume (µl)	Final concentration	Program
10× Pfu-Ultra reaction buffer	5	1×	1. 95°C for 2 min
Forward primer (100µM)	1.25	2.5µM	2. 95°C for 30 s
Reverse primer (100µM)	1.25	2.5µM	3. 55°C for 1 min

Deoxynucleotide mix (10mM)	1	200µM	4. 68°C 1 kb / min
DMSO	2.5	5% (v/v)	5. cycle steps 2-5 16-18 times
DNA template	1-5	5-50ng	6. hold at 4°C
Pfu-Ultra DNA polymerase	1	2.5U	-
molecular grade water	Up to 50		-

2.3.5 PCR screening of colonies

Bacterial colonies were resuspended in 50 µl sterile LB medium using radiation sterilised inoculation needles (Fisher Scientific). 5 µl was used as a template for the PCR reaction. A scaled-down standard PCR reaction using Red-Taq DNA polymerase was used with the appropriate primers to check for the presence of insert, with a reaction volume of 25 µl. The PCR reaction was resolved on an agarose gel and positive clones identified by presence of a band of correct size. For positive clones, the remaining 45 µl of colony resuspension was used to inoculate 5ml liquid media plus appropriate antibiotic, and incubated for 12-16 h at 37^{o}C, 200 rpm. Plasmid DNA was prepared from this culture as described in Section 2.4.1.

2.3.6 Agarose gel electrophoresis

Agarose gels were typically made as 1% (w/v) agarose (Invitrogen) in 1× TAE buffer (40mM tris-acetic acid, 10mM EDTA pH 8.0). However, 0.5-2.0% gels were used according to the final size of the expected DNA fragments. The agarose was dissolved in 1× TAE by heating in a microwave oven on full power for 1 min cycles and cooled to 50^{o}C. Ethidium bromide solution was added to a final concentration of 0.5 mg/ml, and the gel solution poured into a gel former with comb inserted. The set gel was transferred to an electrophoresis tank and submerged in 1× TAE buffer. Samples were mixed with 6x gel loading buffer (30% (v/v) glycerol, 20 mM EDTA pH 8.0, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol in water) before loading into wells. A 1Kb Plus DNA Ladder (Invitrogen) was used as a size standard. The gel was run at 100 V for approximately 30 min or until an appropriate resolution was achieved. Ethidium bromide stained nucleic acid was visualised using a UV transilluminator system (Bio-Rad Gel Doc 2000 or UVP Bioimaging Systems AC1 Auto Darkroom).

2.3.7 Gel extraction of DNA

For the extraction of DNA from agarose gels, the QIAquick Gel Extraction kit (QIAGEN) was used according to manufacturers' instructions. Samples were eluted in 30-50 ml of elution buffer or molecular biology grade water.

2.4 Cloning

2.4.1 Preparation of plasmid DNA

DNA was isolated from either small-scale cultures (10 ml) or large-scale cultures (250 ml) of *E.coli*. The QIAGEN Spin Miniprep or Plasmid Maxiprep kits were used for all plasmid preparations according to manufacturers' instructions. DNA was quantified using a Nanodrop spectrophotometer (GE Healthcare).

2.4.2 Restriction digestion of DNA

All restriction enzymes were purchased from NEB. For preparation of vectors for cloning, 1 µg of purified vector was cut with 20 U of the appropriate enzymes in a final reaction volume of 50 µl for 2 h at 37°C. The reaction mix was additionally treated with 10 U of calf intestinal alkaline phosphatase (CIP) for a further 1 h at 37°C to minimise re-circularisation of the vector DNA. Digest reactions were resolved on an agarose gel, and the appropriate DNA fragment purified by gel extraction (Section 2.3.7). For digestion of purified PCR products, the entire purified product was mixed with the appropriate 10× buffer and restriction enzyme (approximately 20 U) for 1-2 h at 37°C. The digest was purified using the Qiagen PCR purification kit. To confirm the presence of an insert in a new vector, 1 µl of DNA was digested with the appropriate restriction enzyme (approximately 10 U) and 10× buffer in a total volume of 20 µl for 1 h at 37°C. The restriction digest products were separated by agarose gel electrophoresis.

2.4.3 Ligation of DNA sequences into expression vectors

Ligation reactions of digested DNA were performed using T4 DNA ligase (Promega). A typical reaction consisted of a 3:1 molar ratio of insert to vector. 10µl ligation reactions contained 1 µl 10× ligase buffer, 1 µl T4 DNA ligase. Reactions

were centrifuged briefly before incubation overnight at 4°C. Control reactions containing insert only or vector only were prepared to confirm complete digestion of the insert and vector. Typically, 5 µl of the ligation reaction was used to transform E. coli (Section 2.4.6).

2.4.4 Large-scale preparation of chemically competent E.coli XL1-Blue cells

An *E. coli* XL1-Blue glycerol stock was streaked onto LB- tetracycline plates and grown overnight at 37°C. A single colony was inoculated into 10 ml LB-tet media and grown overnight (37°C, 200 rpm). 5ml of the overnight culture was used to inoculate 500ml LB-tet, which was grown at 37°C with agitation (200 rpm) until the cells reached mid-log phase (OD600 approximately 0.5). The culture was cooled on ice for 2 h and the bacteria were pelleted by centrifugation at 4000 rpm for 20 min at 4°C. The pellet was gently resuspended in 250 ml fresh, ice-cold, filter sterilised salt buffer (100 mM calcium chloride ($CaCl_2$), 70 mM manganese chloride ($MnCl_2$), 40 mM sodium acetate (NaOAc) pH 5.5) and incubated on ice for 45 min. After incubation, the bacteria were pelleted by centrifugation at 4000rpm for 10 min at 4°C. The pellet was resuspended in a total of 50 ml salt buffer and 11.5 ml ice-cold filter sterilised 80% (v/v) glycerol was added drop-wise with gentle agitation, to give a final concentration of 15% (v/v). Single aliquots of 500 µl were stored at -80°C.

2.4.5 Small-scale preparation of chemically competent E.coli BL21

A single colony was used to inoculate 10 ml LB media that was incubated overnight at 37^{o}C, 200 rpm. 500 µl of the overnight culture was used to inoculate 10 ml LB media and was incubated at 37^{o}C, 200 rpm until OD_{600} reached 0.2-0.3 (approximately 90 min). Cells were harvested by centrifugation at 4000 rpm, 15 min, 4^{o}C, and the pellet resuspended in 5 ml of ice-cold 100 mM $CaCl_2$. The cells were incubated on ice for 10 min, centrifuged again (as above) and the pellet resuspended in 400ml 100mM $CaCl_2$. Cells were stored at 4^{o}C for a maximum of 7 days.

2.4.6 Transformation of bacteria using the heat-shock method

An aliquot of chemically competent cells was thawed on ice before use and the contents mixed thoroughly. 5 µl of ligation mix or 1 µl diluted plasmid (10 ng) were mixed with 50 µl of competent cells in a pre-chilled 15 ml polypropylene tube

(Falcon 2059, BD Bioscience) and incubated on ice for 30 min. Cells were heat-shocked for 45 s at 42°C and immediately cooled on ice for a further 2 min before addition of 1 ml LB media. Cells were incubated at 37°C for 30-60 min to allow recovery before transfer to a 1.5 ml microcentrifuge tube and centrifugation at 13,000 rpm, 1 min, and the pellet resuspended in 200 ml LB media. The cell suspension was plated onto LB agar plates supplemented with appropriate antibiotic. The plates were incubated overnight at 37°C.

2.4.7 DNA sequencing

Sequencing reactions were carried out by the Geneservice Oxford DNA sequencing service at the Department of Biochemistry, University of Oxford. All sequencing was performed using ABI BigDye Terminators V3.1 (Applied Biosystems). The sequencing chromatogram traces were checked using Chromas Lite, version 2.01 (Technelysium Pty Ltd).

2.5 Protein analysis

2.5.1 Sample preparation

For the detection of recombinant proteins in mammalian cells, $7.5x10^5$ SH-SY5Y or $5x10^5$ HEK293T cells were seeded into 6-well plates, transfected after 24h, and harvested after a further 24 h. To detect endogenous protein in SH-SY5Y, $7.5x10^5$ cells were seeded into 6-well plates and grown for 48 h. Cells were washed twice with PBS and lysed directly in 250 µl 2× SDS-sample buffer (0.125 M Tris pH 6.8, 4% sodium dodecyl sulphate (SDS), 20% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 0.001 (w/v) bromophenol blue). Lysed cells were scraped, transferred to a 1.5ml microcentrifuge tube and sonicated using a UP50H Ultrasonic Processor (Hielscher). 10-20 µl was used for analysis by SDS-PAGE (Section 2.5.2). For quantitation of protein levels after siRNA treatment, transfected SH-SY5Y cells were washed twice with PBS and lysed in 100 µl of RIPA (150mM NaCl, 50mM Tris pH 8.0, 1% (v/v) Triton X-100, 0.5% (w/v) sodium deoxycholate, and 1mM EGTA) buffer. Lysed cells were scraped, transferred to a 1.5ml microcentrifuge tube and sonicated with a UP50H Ultrasonic Processor (Hielscher). The lysates were clarified by centrifugation at 13’000 rpm for 10 min and the protein content of the supernatant was quantified using a BCA assay (Pierce). The required amount of protein was diluted in 2× sample

buffer and analysed by SDS-PAGE (Section 2.5.2). For all other samples, a 1:1 ratio of 2× SDS-sample buffer was added to the sample prior to sonication using a UP50H Ultrasonic Processor (Hielscher). All samples were boiled for 5 min at 95^{o}C and cooled on ice prior to separation by SDS-PAGE. All samples were stored at –20^{o}C for long-term storage.

2.5.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were separated on 10 or 12% (v/v) acrylamide gels by SDS-PAGE, under denaturing and reducing conditions using the Mini Protean III Gel System (Bio-Rad). The resolving gel contained 380mM Tris-HCl pH 8.8, 6-15% (v/v) acrylamide (30% (w/v) acrylamide: 0.8% (w/v) Bis-acrylamide (37.5:1)) (National Diagnostics), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate (APS) and 0.08% N,N,N′,N′-Tetramethylethylenediamine (TEMED). The stacking gel consisted of 125mM Tris-HCl pH 6.8, 5% (v/v) acrylamide (30% (w/v) acrylamide: 0.8% (w/v) Bis-acrylamide (37.5:1)), 0.1% (w/v) SDS, 0.1% (w/v) APS and 0.1% TEMED. The resolving gel was layered with isopropanol during the polymerisation process to ensure a level interface with the stacking gel. Following polymerisation, the IPA was discarded and the top of the gel was gently washed with water. The stacking gel was then layered onto the resolving gel and a lane-forming comb was inserted. When the stacking gel had polymerised, the gels were immobilised in the clamp system, submerged in SDS-PAGE running buffer (25 mM Tris-base, 192 mM glycine, 1% (w/v) SDS). Pre-stained Protein Marker, Broad Range (6-175 kDa) (New England Biolabs) or all blue prestained Precision Plus (10-250 kDa) (Bio-Rad) were used as protein standards. Gels were run at 150 V for approximately 85 min or until the appropriate resolution of the visible pre-stained protein standard was achieved.

2.5.3 Coomassie stained SDS-PAGE gels

For direct visualisation of protein after SDS-PAGE, the gel was stained with coomassie stain (0.1% (w/v) coomassie brilliant blue, 40% (v/v) methanol, 10% (v/v) acetic acid) for 1-2 h or overnight at room temperature, followed by destaining for 2-6 h in 40% (v/v) methanol, 10% (v/v) acetic acid at room temperature. Gels were rinsed in distilled water and scanned using the Odyssey® Infrared Imaging System (LI-COR Biosciences) in the 700nm channel with a focus offset of 0.5mm.

2.5.4 Western blotting

Following electrophoresis, protein from unstained SDS-PAGE gels was transferred to nitrocellulose membrane (Trans-blot transfer medium, 0.2mm pore size, Bio-Rad) using the Mini Trans-blot electrophoretic transfer cell (Bio-Rad). Proteins were transferred for 1 h at 75 V for a single gel or 85 V for two gels in transfer buffer (25 mM Tris-base, 192 mM glycine, 1% (w/v) SDS, 20% (v/v) methanol). The membrane was blocked either for 1 h at room temperature or overnight at 4^{o}C in TBST with 5% (w/v) dried skimmed milk powder. After blocking, the membrane was incubated for 1-2 h at room temperature or overnight at 4^{o}C in primary antibody diluted at the required concentration (see Section 2.5.5) in 5-10 ml blocking solution. The membrane was washed with two 5 min washes in TBST and a final wash in blocking solution, before incubation with the secondary antibody (Section 2.5.5) diluted in 5-10 ml blocking solution. The secondary antibody was incubated at room temperature for 30 min (Alexa Fluor 680 or IRDye 800-conjugates) protected from light. Three additional 5 min washes with TBST were completed before detection. Fluorescent secondary antibodies were detected using the two-colour Odyssey® Infrared Imaging System (LI-COR Biosciences).

2.5.5 Antibodies and stains

A table of antibodies and stains used in this thesis is provided below with their individual uses and required dilutions. Abbreviations: WB; western blot, ICC; immunocytochemistry, HTRF; Homogenised time-resolved fluorescence, EuK; europium cryptate, DSHB; Developmental Studies Hybridoma Bank)

Name	Source	Type	Use and dilution
Primary Antibodies			
TCF4 M03	Commercial (Abnova, M03)	Mouse monoclonal	WB (1:1000)
SC35	Commercial (Abcam, ab11826)	Mouse monoclonal	ICC (1:5000)
β-actin	Commercial (AC-15, Sigma)	Mouse monoclonal	WB (1:1000)
caspase 3	Commercial (Cell Signaling, 9662)	Mouse monoclonal	WB (1:1000)
myc-EuK (conjugated)	Commercial (Cisbio, 61MYCKLA)	Mouse monoclonal	HTRF (1:400)
HA-XL665 (conjugated)	Commercial (Cisbio, 610HAXLA)	Mouse monoclonal	HTRF (1:400)
TCF4_01	In-house (anti-GST-TCF4)	Rabbit polyclonal	WB (1:250)

			ICC (1:400)
TCF4_02	In-house (anti-TRX-TCF4)	Rabbit polyclonal	WB (1:250) ICC (1:400)
α-tubulin	DSHB (Hybridoma, 12G10)	Mouse monoclonal	WB (1:20'000)
myc	DSHB (Hybridoma, 9E10)	Mouse monoclonal	WB (1:400) ICC (1:1000)
Secondary Antibodies			
Alexa Fluor 546	Commercial (Invitrogen)	Goat anti-rabbit	ICC (1:2'000)
Alexa Fluor 680	Commercial (Invitrogen)	Donkey anti- rabbit	WB (1:20'000)
IRDye 800	Commercial (LI-COR)	Donkey anti- mouse	WB (1:20'000)
Stains			
Alexa Fluor 488 phalloidin	Commercial (Invitrogen)	N/A	ICC (165nM)
Hoechst 33342	Commercial (Invitrogen)	N/A	ICC (1μg/ml)

2.6 Protein expression and purification

2.6.1 Induction of protein expression using pET-32 or pGEX-4T system

The TRX-TCF4 and GST-TCF4 recombinant plasmids were expressed in BL21 (DE3) strain of *E. coli*. Cells containing expression plasmids were selected on LB agar containing ampicillin (Section 2.2.1) and a single colony used to inoculate 10ml LB media supplemented ampicillin. This was incubated overnight at 37^{o}C, 200 rpm. 500 μl of the overnight culture was used to inoculate 250ml LB media, which was incubated at 37^{o}C, 200 rpm for approximately 3 h until an OD_{600} of approximately 0.6. Expression of the fusion protein was induced by the addition of 1mM IPTG for a further 3-5 h at 37^{o}C, 200 rpm. Cells were harvested at 12,000 rpm for 20 min. In some cases, the pellet was kept for long-term storage at –20^{o}C. Pre-induction and post-induction aliquots were processed for SDS-PAGE and coomassie stained to verify successful induction (Section 2.5.1, 2.5.2 and 2.5.3).

2.6.2 Purification of the TRX-TCF4 fusion protein

Cell pellets were resuspended in 15 ml sonication buffer (20mM Tris pH 8.0, 100mM NaCl) and sonicated using a Vibra-Cell Ultrasonic Processor (Sonics). Samples were periodically chilled on ice to avoid heat-denaturation of fusion protein. Cells were harvested by centrifugation at 12,000rpm, 15 min at 4^{o}C. Soluble proteins were found

in the supernatant; whereas insoluble proteins were extracted from the pellet by resuspension in 15ml sonication buffer containing 8M Urea, followed by centrifugation at 12,000rpm, 15 min, and 4^{o}C to remove cell debris. The TRX-TCF4 fusion protein in the soluble fraction was purified using immobilised metal affinity chromatography (IMAC) using TALON® Resin (Clontech). A 2ml bed volume was packed into an Econo-Pac® disposable chromatography column (Bio-Rad) and equilibrated with one column fill of sonication buffer. The fusion protein was then passed over the column for 1h. Unbound protein was drained and the affinity resin washed with a column fill of sonication buffer and a frit inserted just above the packed affinity resin. The recombinant protein was eluted from the TALON® column using sonication buffer plus 100mM imidazole and collected in 2ml fractions. The protein was typically present in the 2nd and 3rd fractions. Aliquots of pre-purification fractions and elutions were processed for SDS-PAGE and coomassie stained to verify successful purification (Section 2.5.1, 2.5.2 and 2.5.3).

2.6.3 Purification of GST-TCF4 fusion protein

Solubilisation of the GST-TCF4 fusion protein was performed in sonication buffer as described above for TRX-TCF4 (Section 2.6.2). The GST-TCF4 fusion protein in the soluble fraction was purified using Glutathione Sepharose 4B (Amersham Biosciences). A 0.5ml bed volume was packed into an Econo-Pac® disposable chromatography column (Bio-Rad) and equilibrated with one column fill of sonication buffer. The fusion protein was then passed over the column for 1h. Unbound protein was drained and the affinity resin washed with a column fill of sonication buffer and a frit inserted just above the packed affinity resin. The Glutathione Sepharose 4B column was washed with 10 column volumes of 50 mM Tris pH 8.0 prior to elution in 50 mM Tris pH 8.0 containing + 10 mM reduced glutathione. Elutions were incubated on the column for 5 min and collected in 0.5 ml fractions. Aliquots of pre-purification fractions and elutions were processed for SDS-PAGE and coomassie stained to verify successful purification (Section 2.5.1, 2.5.2 and 2.5.3).

2.6.4 Antibody Production

Rabbit polyclonal antibodies were raised against the TRX-TCF4 and GST-TCF4 fusion proteins obtained using the methods described above (Section 2.6.1-2.6.3). Immunisations were performed by Covalab (UK). Two New Zealand rabbits were immunised for each purified fusion protein using an 88 day protocol with four immunisations.

2.6.5 Production of affinity purification columns

The thioredoxin fusion protein used to produce antisera was desalted into reduction buffer (0.1 M sodium phosphate, 5 mM EDTA) using an Econo-Pac® 10DG desalting column (Bio-Rad) equilibrated with reduction buffer. 3 ml was allowed to enter the column, and was eluted with 4ml reduction buffer. This was reduced with 25 mM DTT at 37°C for 90 min in the dark. The Econo-Pac® 10DG desalting column was equilibrated with coupling buffer (50mM Tris pH 8.5, 5mM EDTA). 2 ml coupling buffer was added to the 4ml reduced fusion protein and 3ml added to the desalting column. The fusion protein was eluted with 4 ml coupling buffer, and the process repeated with the final 3 ml to give a final volume of 8 ml fusion protein in coupling buffer. This was coupled to 2ml packed (4ml slurry) of Sulfolink® Coupling Gel (Pierce) as per manufacturers' instructions, and the resulting column used for the affinity purification of antibodies. Columns were stored at 4°C with the column bed equilibrated in phosphate buffered saline (PBS; 10mM phosphate buffer, 2.7mM KCl, 137mM NaCl, pH 7.4) plus 0.025% (v/v) sodium azide.

2.6.6 Affinity purification of antisera

Affinity chromatography columns were equilibrated with PBS. 5 ml of antiserum was diluted to 20 ml with PBS. The diluted antiserum from the TRX-TCF4 immunisation was passed over a thioredoxin column (kindly provided by Dr. C. Tinsley, DJ Blake Lab, Cardiff University) for 1h at room temperature to pre-absorb against thioredoxin specific antibodies in the antisera. The pre-absorbed antiserum was then repeatedly passed over the TRX-TCF4 affinity chromatography column for 1h. In the case of the antiserum from the GST-TCF4 immunisation, the diluted antiserum was directly passed over the TRX-TCF4 affinity chromatography column for 1h without pre-absorbing. To purify non-specific IgG, 1 ml of pre-immune rabbit serum was diluted to 20 ml with PBS and passed over a Protein A agarose bead column with 2 ml bed

volume (provided by Dr. A. Waite, DJ Blake Lab, Cardiff University) for 1 h. Following antibody binding, columns were washed with 2 column fills of PBS and the bound antibodies eluted with ImmunoPure® IgG elution buffer (Pierce) in 1ml fractions and neutralised by the addition of 50 ml/ml 1 M Tris pH 9.0. Columns were regenerated by thorough washing with at least five column bed volumes of ImmunoPure® IgG elution buffer, followed by a column fill of PBS, and long term storage was maintained in PBS with 0.025% sodium azide at 4°C.

2.7 Proteomics

2.7.1 Cross-linking antibodies to protein A beads

The TCF4_01, TCF4_02 and pre-immune IgG antibodies were coupled to protein A agarose beads (Invitrogen). 2 mg of affinity-purified antibody was coupled to 0.5 ml packed beads (equilibrated with PBS), rotating at room temperature for 2 h. The beads were spun down at 2,000 rpm for 5 min and washed twice with 5 ml 0.1M borate buffer (0.2 M di-sodium tetraborate, 0.2 M boric acid, pH 9.0). The bound antibody was cross-linked to the beads through incubation with 20 mM dimethyl pimelimidate dihydrochloride (DMP) in 9.5 ml borate buffer by rotating for 30 min at room temperature. The beads were spun down by centrifugation at 2,000 rpm for 5 min and the reaction stopped by washing the beads in 5 ml 0.1M ethanolamine, pH 8.0. Unreacted DMP was quenched by incubating the beads in 5 ml 0.1 M ethanolamine for 1 h at room temperature, before centrifugation (as above), washing in 5ml PBS and centrifugation again. To remove uncoupled IgG, the beads were washed with 5 ml of ImmunoPure® IgG elution buffer, followed by two PBS washes and transfer to a 1.5ml microcentrifuge tube. The beads were stored as a 50% slurry in PBS plus 0.05% sodium azide at 4°C.

2.7.2 Immunoprecipitation (IP)

SH-SY5Y cells were seeded at 5×10^6/dish in four 15 cm dishes (BD Bioscience) and grown for 48 h. Cells were washed twice with PBS, scraped into a 15 ml falcon tube (Corning) with 10 ml of PBS, and pelleted for 2 min at 2500 rpm and 4°C. PBS was removed, and the cells were solubilised in 4 ml of ice-cold RIPA buffer for 30 min at 4°C. The sample was homogenised for 10 s using a POLYTRON® PT3100 (Kinematica) and then placed on ice 30 s. Homogenisation was repeated two more

times. The sample was further homogenised by passage through a 23G needle six times. The lysate was divided into four 1.5 ml eppendorfs and clarified by centrifugation for 10 min at 13'000 rpm and 4°C. Samples were pooled into a 15cm falcon tube and a 50 µl sample (input) was taken prior to pre-clearing with 50 µl of packed protein A agarose beads (equilibrated with RIPA) for 1-2 h at 4°C with constant rotation. The beads were pelleted for 2 min at 2500 rpm (4°C) and the pre-cleared lysate was divided into four samples (two controls, two IPs). For TCF4 IP, 5µg of affinity purified TCF4_01 or TCF4_02 was added to the protein lysates. For control IPs, either 5 µg of affinity purified 9E10 antibodies or no antibody was added (no Ab). All samples were incubated overnight at 4°C with constant rotation. The following day, insoluble material was pelleted by centrifugation for 10 min at 13'000 rpm (4°C) and the supernatants were transferred to new 1.5 ml eppendorf tubes. Antibody complexes were captured by the addition of 15 µl of packed protein A agarose beads (equilibrated with RIPA) to each tube and constant rotation for 1 h at 22 rpm (4°C). Beads were spun down by centrifugation and a 50 µl sample was taken (post-IP) prior to washing the beads four times with 1 ml of RIPA buffer. RIPA buffer was removed and 70 µl of 2× SDS-sample buffer was added to the beads. 50 µl of 2× SDS-sample buffer was added to the 'input' and 'post-IP' samples. Samples were heated to 95°C for 5 min before resolution by SDS-PAGE (Section 2.5.2).

2.7.3 Large-scale immunoaffinity purification (IAP)

The following section describes the procedure for IAP with one TCF4 antibody and one control antibody. SH-SY5Y cells were seeded at 5×10^6 cells/dish in sixteen 15 cm dishes (BD Bioscience) and grown for 48 h (approximately 2×10^8 cells after growth). Each dish was washed twice with ice-cold PBS, the cells were scraped and divided into two 50 ml polypropylene tubes (Corning). Tubes were filled to 50 ml with ice-cold PBS and the cells were pelleted for 2 min at 2500 rpm and 4°C before removal of the PBS. At this stage, cells could be stored at -80 °C. Frozen cells were thawed on ice prior to use. Fresh or thawed cells were solubilised in 10 ml of ice-cold RIPA buffer per 50 ml polypropylene tube for 30 min at 4°C. The lysed cells were homogenised for 10 s using a POLYTRON® (Kinematica) and then placed on ice 30 s. Homogenisation was repeated two more times. The sample was further homogenised by passage through a 23G needle six times. The lysates were combined

and clarified by centrifugation for 10 min at 13'000 rpm (4°C) and a 50 µl sample (input) was taken. The lysate was pre-cleared with 300 µl of packed protein A agarose beads (equilibrated with RIPA buffer) in a 50 ml polypropylene tube for 3h at 4°C. The pre-cleared lysate was divided into two samples, one incubated with 50 µl of packed TCF4-protein A agarose beads (equilibrated with RIPA buffer), and the other 50 µl of IgG-protein A agarose beads (equilibrated with RIPA buffer). Samples were incubated under constant rotation for 3 h at 4°C. The following incubation, an unbound fraction was taken, beads were washed four times with 10 ml RIPA buffer followed by transfer to a 1.5 ml microcentrifuge tube. Proteins were eluted with 70ul of SDS-sample buffer, boiled at 95°C for 5 min, and cooled on ice. The eluted sample was pulse centrifuged before resolution on a 4-12% gradient NuPAGE® Novex Bis-Tris gel (Invitrogen), run in 1× MOPS Running Buffer (Invitrogen) using an XCell *Surelock*™ Mini-Cell electrophoresis system (Invitrogen). Gels were run for 90 min at 150V before fixing for 15 min in 50% (v/v) methanol, 7% (v/v) acetic acid, diluted molecular grade water (Millipore). Gels were washed two times 5 min in molecular grade water prior to staining with 20 ml colloidal coomassie blue stain (Gelcode Blue Stain Reagent, Pierce) for 2 h, followed by de-staining with molecular grade water. Coomassie stained gels were scanned using the Odyssey® Infrared Imaging System and bands of interest excised with a clean scalpel blade, transferred to sterile 1.5 ml microcentrifuge tubes and sent for mass spectrometry.

2.7.4 Mass spectrometry

Excised gel plugs were sent to the Functional Genomics and Proteomics Laboratories, School of Biosciences, University of Birmingham to be processed and digested for mass spectrometry. The UltiMate® 3000 HPLC series (Dionex) was used for peptide concentration and separation. Peptides were eluted using a Triversa Nanomate nanospray source (Advion Biosciences) into a LTQ Orbitrap Velos ETD mass spectrometer (ThermoFisher Scientific). The data-dependent scanning acquisition was controlled by Xcalibur 2.1 software. Peptide data was matched against the NCBInr database using Mascot algorithm (Matrix Sciences) and software Proteome Discoverer 1.3 (ThermoFisher Scientific). The Mascot algorithm uses probability based scoring to identify proteins from the database. The cumulative protein score is based on summing the ion scores of the unique peptides identified for

that protein. If a peptide is redundantly identified, only the highest-scoring peptide is used. Scores are reported as $-10 \times \mathrm{Log}_{10}(P)$, where P is the probability that the observed match is a random event. Variable modifications were deamidation (N and Q), oxidation (M) and phosphorylation (S, T and Y).

2.8 Cell biology

2.8.1 Immunocytochemistry

Cells were grown on sterile glass coverslips (22mm x 22mm) in 6-well plates seeded with 5×10^4 COS-7 or 7.5×10^4 SH-SY5Y cells. Cells were grown for 24 h, transfected, and fixed 24 h later. COS-7 cells were transfected with 1 µg of plasmid DNA and 6 µl of FuGENE (Roche). SH-SY5Y were transfected with 3 µg of plasmid DNA and 10 µl of lipofectamine 2000 (Invitrogen). For siRNA treatment, cells were fixed after the 72 h knockdown protocol (Section 2.9.2). Cells were washed twice with PBS and fixed with 1ml of 4% paraformaldehyde (PFA) solubilised in PBS for 15 min at room temperature. The PFA was removed and the cells permeabilised with 2 ml PBS and 0.1% (v/v) Triton X-100 at 4°C for 15 min. Following permeabilisation, the cells were washed three times with PBS for 5 min with shaking. The cells were blocked with 1 ml of 10% (v/v) FCS for 20 min at room temperature. Blocking solution was removed and the primary antibody diluted in 1 ml PBS containing Heoscht stain (1ug/ml) was added to the cells for 1 h at room temperature with shaking. The cells were washed twice with PBS for 5 min and were incubated with the secondary antibody in 1 ml PBS at room temperature with shaking for 30 min. The antibodies used in this study are presented in Section 2.5.5. Following a final set of three 5 min PBS washes, the coverslips were mounted onto glass slides with Aqua Poly/Mount (Polyscience, Inc.). Cells were visualised using a Leica SP5 or Leica DM6000B confocal microscope with a 63x oil immersion objective lens. The images were assembled using ImageJ, Adobe Photoshop and Illustrator, in which images were cropped and adjusted for brightness and contrast but otherwise not manipulated.

2.8.2 Homogeneous time resolved fluorescence (HTRF)

Six well plates were seeded with HEK-293T cells at 5×10^5 cells/well. After 24 h, 1 µg of the appropriate HA-tagged and myc-tagged constructs were transfected into a

6-well using FuGENE 6. Negative control experiments were also performed that consisted of either a transfection with only one of the epitope-tagged constructs or the empty pCMV-HA and pCMV-myc vectors. 24 h after transfection, the cells were washed with ice-cold PBS and lysed in 0.5 ml of chilled NP-40-containing buffer (50mM Tris pH 7.4, 150mM NaCl, 1% (v/v) NP-40, 0.1% (w/v) BSA, 100mM potassium fluoride) for 30min at 4°C and homogenised using a 23 G syringe needle. Lysates were clarified by centrifugation at 13'000 rpm for 10 min (4°C) and 100µl of clarified supernatant was diluted 1:1 in NP-40 buffer containing 0.5µl of anti-HA-europium cryptate (Cisbio) and 0.5µl anti-9E10-XL665 (Cisbio) and incubated in a 96-well microplate (Greiner Bio-one) overnight on a shaker at 4˚C. Three biological replicates and two technical replicates for each experimental condition were read on a dedicated HTRF reader (Berthold Artemis K-101) with a 100ms delay, 100ms integration time and flash number of 50. For each condition, the mean and standard error were calculated from at least three separate experiments.

2.8.3 Luciferase assays

HEK-293T cells were seeded at 10^4 cells/well on 96 well plates and left to grow for 24 h. The pGL3-basic (luciferase reporter), pRL-CMV (*Renilla* reporter) and all reagents used were part of the Dual-Luciferase assay system (Promega). The design strategy for promoter regions cloned into pGL3-basic is described in Section 2.10.2. Each 96-well was transfected with one HA-tagged bHLH transcription factor construct (5-10 ng), one myc-tagged transcription factor (20 ng), a *Renilla* luciferase control (0.5 ng), and a luciferase reporter construct (pLuc-CNTNAP2 or pLuc-NRXN1β) in total volume of 2 µl using FuGENE 6 (Roche) (see table below). Luciferase reporter constructs in the reverse direction (pLuc-CNTNAP2_R and pLuc-NRXN1β _R) were used as negative controls for promoter activity. 24 h post-transfection, cells were washed with 100 µl of PBS and lysed in 20 µl of passive lysis buffer for 15 min at room temperature with slow mixing on an orbital shaker. The fluorescent signal was read on a Berthold LB 96 V plate-reader with automatic injection dispensing 50 µl of LARII reagent followed by 50 µl of "stop and glow" reagent. Each measurement was integrated over 10s with 2s delay between measurements. Results represent the mean and standard error of experiments performed in quadruplicate.

Each transfected 96-well contained one construct of the following categories:

Construct	ng DNA
1. HA-tagged bHLH protein	
HA-TCF4, HA-ATOH1, HA-ASCL1, HA-NEUROD1, HA-ID2	5-10
2. myc-tagged TCF4 or mutant	
myc-TCF4, myc-G358V, myc-D535G, myc-R574P, myc-R580W, myc-A614V	20
3. Promoter	
pLuc-CNTNAP2_F or pLuc-CNTNAP2_R	100
pLuc-NRXN1β _F or pLuc-NRXN1β _R	75
***4. Renilla* luciferase (transfection control)**	
pRL-CMV	0.5

2.8.4 Cell viability and caspase assays

SH-SY5Y cells were seeded at 10^4 cells/well in two 96 well clear bottom plates and left to grow for 24 h. 24 replicate wells were treated with the appropriate siRNA using the 72 h knockdown protocol (Section 2.9.2). Cell viability and caspase 3/7 activation were measured in the same wells using the CellTiter-Fluor™ Cell Viability (Promega) and the Caspase-Glo® 3/7 (Promega) assays. 3 h prior to the start of the assays, 1 µl of staurosporine dissolved in DMSO (100µM stock) or 1 µl of DMSO was added to 24 previously untreated wells as a positive control and negative control respectively. For cell viability, 20 µl of a 5× concentrate of CellTiter-Fluor™ Cell Viability reagent was added to the 100 µl of each 96 well and mixed briefly by orbital shaking. The plates were incubated for 45 min at 37°C and the emitted luminescent signal was measured in 24 wells for each condition using the luminometer module of a FLUOstar OPTIMA microplate reader with a gain of 1000. After the first measurement, 120 µl of 1× Caspase-Glo® 3/7 reagent was added to 12 wells of each condition and mixed briefly by orbital shaking prior to incubation for a further 45 min at 37°C. The emitted fluorescent signal was measured using the fluorometer module of a FLUOstar OPTIMA microplate reader. Readings were taken from the top of the wells with a gain of 4000. All data were normalised to the levels in the mock-transfected cells, and represented as the mean and standard error. *t*-tests were used to compare treatment groups. In addition to the microplate assays, caspase 3 cleavage was assessed by western blotting. SH-SY5Y cells were seeded in 6 well plates and transfected with the appropriate siRNA using the 72 h knockdown protocol

(Section 2.9.2). After treatment, dead cells were collected from the media by centrifugation and lysed 100 µl RIPA buffer with the remaining live cells. The protein content was measured using a BCA assay (Pierce) and 15µg of each lysate was used for western blotting with caspase 3 and β-actin antibodies (Section 2.5.4).

2.9 Gene expression analysis

2.9.1 Short interfering RNA (siRNA) design

27 nucleotide long blunt-ended RNA oligonucleotide duplexes (27mers) were used for all siRNA treatments. To knockdown *TCF4*, siRNAs were designed using predictive algorithms (siDESIGN Center, http://www.thermoscientificbio.com; siMAX^{TM} Design Tool, http://www.eurofinsdna.com). Two non-overlapping sequences were selected according to predicted efficacy scores and homology to regions of interest. For siRNA sequences targeting the TCF4-A isoform, only one sequence (KDA2) was identified by the predictive algorithms due the limited length of the unique sequence in exon 1 therefore the second siRNA was designed manually (KDA1). The 27mer targeting *GAPDH* was designed from a 23 nucleotide sequence known to be effective for knocking down *GAPDH* provided by Dr. C. Tinsley (DJ Blake Lab, Cardiff University). The GAPDH siRNA was used in control experiments to identify transcripts that were altered by non-specific mechanisms related to the presence interfering RNA in cells. Lyophilised siRNA was reconstituted in siMAX Universal Buffer (30mM HEPES, 100mM KCl, 1mM $MgCl_2$; pH 7.3) as a 50mM stock and stored at -80°C. Sequences of siRNA duplexes used in this thesis are provided in the table below:

siRNA name	siRNA binding site	siRNA-mediated effect	siRNA Sequence
GAPKD	GAPDH exon 1-2	GAPDH Knockdown	5'- CGGAGUCAACGGAUUUGGUCGUAUUGG
KD1	TCF4-B exon 10-11	Global TCF4 knockdown	5'-GGGACAGACAGUAUAAUGGCAAAUAGA
KD2	TCF4-B exon 17	Global TCF4 knockdown	5'-AUAAUGACGAUGAGGACCUGACACCAG
KDB1	TCF4-B exon 1	TCF4-B knockdown (long isoform)	5'- GGGACGGACAAAGAGCUGAGUGAUUUA

KDB2	TCF4-B exon 2-3	TCF4-B knockdown (long isoform)	5'-ACUGGCUCAAAUGUAGAAGACAGAAGU
KDA1	TCF4-A exon 1	TCF4-A knockdown (short isoform)	5'-ACTGCGCATACACAATCCCGGGCATGG
KDA2	TCF4-A exon 1	TCF4-A knockdown (short isoform)	5'-GCAACUCUUUGAUGUACUACUAUAAUG

2.9.2 siRNA-mediated knockdown protocol

SH-SY5Y were seeded at 2×10^5 cells per well in 6 well plates. 24h after seeding, cells were transfected with the required siRNA duplexes at a final concentration of 10 nM using 4 µl of Lipofectamine® RNAiMAX (Invitrogen) or mock transfected with just the transfection reagent. For the 72 h knockdown protocol used in Chapter 4, the cells were re-transfected 48 h after the first transfection using the same conditions and cells from each condition were prepared for RNA and protein extraction the following day. Each transfection was performed in quadruplicate and three biological replicates were used for gene expression analysis whilst the remaining sample was used for western blotting. For the 48 h knockdown protocol used in Chapter 5, the cells were re-transfected 24 h after the first transfection using the same conditions and cells from each condition were prepared for RNA and protein extraction the following day. Each transfection was performed in five replicates and four biological replicates were used for gene expression analysis whilst the remaining sample was used for western blotting. In both protocols, cells were lysed in 350 µl of RLT buffer for RNA extraction (Section 2.9.3) and 100 µl of RIPA buffer for protein extraction (Section 2.5.1). The RNA and protein quantification was determined using qPCR (Section 2.9.4) and western blotting (Section 2.5.4) respectively.

2.9.3 RNA extraction and cDNA synthesis

Total RNA was prepared from SH-SY5Y cells using the RNeasy Plus Mini Kit (QIAGEN). Cells were lysed directly using 350 µl RLT buffer and homogenised using a syringe and 23 G needle. RNA quality and concentration was determined using a Nanodrop spectrophotometer (GE Healthcare). After RNA extraction, the samples were DNase I treated with the Ambion® TURBO DNA-free™ DNase I (Invitrogen) and the RNA stored at -80°C. 1st strand cDNA was generated by reverse transcription of 400ng to 1 µg of DNAse-free RNA using the ProtoScript® M-MuLV

First Strand cDNA Synthesis Kit (NEB). Oligo-dT priming was used in most instances although Oligo-dT and random hexamer primers were mixed together when quantifying 18S RNA. Control reactions with no reverse transcriptase (-RT) were performed to examine the levels of residual genomic DNA (gDNA). The –RT sample was prepared using pooled RNA from each condition. The cDNA products were diluted as required (at least 5 times) for quantitative or reverse transcriptase PCR (qPCR and RT-PCR).

2.9.4 Quantitative PCR using SYBR Green

SYBR® green was used to quantify the level of *TCF4* and *GAPDH* transcripts after knockdown treatments. Gene expression was quantified using SensiMix™ SYBR No-ROX (Bioline) and a Qiagen Rotor-gene 3000 real time PCR machine (Qiagen) following the conditions outlined in the table below. Triplicate 25µl qPCR reactions were prepared on a Corbett robot (Corbett Robotics CAS1200/QIAGEN) in a 72-well ring. A typical PCR reaction mixture with cycling parameters is provided below. –RT controls and non-template controls (5 µl of water instead of DNA) were included in each qPCR run. The melt curve analysis was performed from 55°C to 95°C and inspected to ensure only one PCR product was being amplified. The threshold cycle (Ct) values were calculated by the Qiagen Rotor-Gene 3000 software after setting a manual threshold in linear phase of PCR amplification (usually 0.4). Relative quantification was used to determine gene expression and compare treatment groups (Section 2.9.8).

Typical qPCR reaction mixture with and cycling parameters:

Name	Volume (µl)	Program
SensiMix (2x master mix)	12.5	1. 95°C for 10 min
Forward primer (100 µM)	0.75	2. 95°C for 15 s
Reverse primer (100 µM)	0.75	3. 60°C for 15 s
Molecular grade water	6	4. 72°C for 20 s
cDNA (2ng/µl)	5	5. Go to step 2; 40 times

2.9.5 Quantitative PCR using TaqMan assays

TaqMan® gene expression assays were used to validate gene expression changes after TCF4-knockdown in Chapters 4 and 5 (Applied Biosystems). Genes selected for validation had a fold change above or below 1.5 and represented important functional categories from the enrichment analyses. A list of TaqMan probe IDs and their associated genes are provided in Appendix II. 1 or 10ng of cDNA was used as template in 20µl reactions with the TaqMan® Fast Advanced Master Mix depending on transcript abundance (Applied Biosystems). All samples were run in triplicate according to the Fast Advanced Master Mix protocol using the ABI 7900HT Fast Real-Time PCR System (Applied Biosystems). In addition to the samples, a dilution series of template cDNA (0.1ng to 10ng or 1ng to 100ng) was used to determine amplification efficiency of each TaqMan assay (Section 2.9.7). Results were analysed on the RQ Manager 1.2 (Applied Biosystems) using manual Ct determination according to manufacturer's guidelines. Background thresholds were set in the non-amplifying phase of the PCR reaction and the Ct thresholds were set in the linear phase of PCR amplification. Relative quantification was used to determine gene expression and compare treatment groups (Section 2.9.8).

2.9.6 Reverse transcriptase PCR (RT-PCR)

RT-PCR was used to determine the expression of epithelial-to-mesenchymal (EMT) genes in a semi-quantitative manner. 10 ng of cDNA prepared from the RNA (Section 2.9.3) of each knockdown treatment group (Section 2.9.2) was amplified for 30 cycles in a 25µl reaction with REDTaq® DNA Polymerase (Sigma) (Section 2.3.3). The sequences of the gene-specific primer pairs are listed Appendix II. 20µl of the PCR reaction was separated on a 2.5% agarose gel containing ethidium bromide. PCR products were imaged with a GelDoc-It® TS Imaging System (UVP).

2.9.7 Primer optimisation

Primer sequences were designed to amplify short cDNA sequences between 50-150 bp and cross-exon boundaries wherever possible to eliminate gDNA amplification. Primer sequences are provided in Appendix II. Prior to use, primer sets were evaluated at different concentrations (50 – 300 nM) on serial dilutions of template cDNA covering the required range of DNA being measured (usually 1 ng to 100 ng). The amplification efficiency was calculated from the standard curve generated from

plotting the Ct values acquired from qPCR analysis against the log of the corresponding amount of cDNA being measured in nanograms. Primer efficiencies were calculated from the following equation, efficiency = $(10^{-1/\text{slope}} - 1) \times 100$. Only primers with amplification efficiencies between 90% and 110% were used in for relative quantification of gene expression (Section 2.9.8).

2.9.8 Relative quantification of gene expression

Relative quantification was determined using the comparative Ct method (ΔCt). Mean Ct values for each transcript were calculated from two or three technical replicates. The mean Ct value of the gene of interest was subtracted from the mean Ct value of a reference gene (ΔCt). Reference genes (often housekeeping genes) are genes that have a high and stable expression under the conditions tested. For the results in Chapter 4 the chosen reference was 18S rRNA and for results in Chapter 5 the chosen reference was the average expression of the COX6A1 and NDUFA4 genes. The COX6A1 and NDUFA4 genes were chosen from the microarray results obtained in Chapter 4, as both genes were highly expressed in SH-SY5Y cells and did not vary in between treatment groups. Logarithmic ΔCt values were converted to linear gene expression values ($2^{\Delta Ct}$). The linear gene expression values of at least three biological replicates were used to calculate the mean abundance and standard error of each transcript. The abundance of each transcript was reported as a percentage of the control group (usually pooled values from mock treated and GAPDH-knockdown treated cells) that was set as 100%. The standard error values were scaled to percentages for each group. A two-tailed *t*-test was performed on two groups of interest to determine if gene expression was statistically different. Genes were considered differentially expressed between groups if $P < 0.05$.

2.9.9 Toray Microarray

Gene expression profiling was performed on a Toray 3D-gene Human Oligo chip 25k (Toray Industries Inc.) by Central Biotechnology Services (CBS), Cardiff University according to manufacturer's instructions. The Toray 3D-gene Human Oligo chip contains 24,267 probes that are mostly located at the 3' end of transcripts. Three biological replicates from four treatment groups (TCF4 KD1and KD2, GAPDH KD and mock) were sent for analysis. The concentration and quality of the RNA for each

sample was assessed using the RNA 6000 Nano Kit and the Agilent 2100 bioanalyser (Agilent Technologies). These tools provided an RNA Integrity Number (RIN) for each sample prior to microarray analysis. 500 ng of total RNA was amplified and labelled using an Amino Allyl MessageAmp™ II aRNA Amplification kit (Life Technologies) and Cy™5 mono-Reactive Dye (GE Healthcare Life Sciences). Labelled aRNAs were hybridised onto 3D-Gene Human 25k oligo chips version 2.1 (Toray Industries Inc). The chips were stringently washed, and fluorescent signals were scanned with a 3D-GeneTM Scanner 3000 (Toray Industries Inc) and analysed using Toray Extraction and 3D-GeneTM Scanner 3000 software (Toray Industries Inc). Hybridised probe spots with signal intensity greater than the mean intensity plus 2 standard deviations of the background signal were considered to be valid. The background average was subtracted from the signal intensity and multiplied by the normalisation factor (25 divided by the medium signal intensity of all the subtracted background data) to generate the normalised data.

The Toray microarray data provided by CBS was in the form of an excel file containing the normalised, background subtracted, signal intensity values with their corresponding gene and probe annotations. The data was imported manually into Partek Genomics Suite 6.6 (Partek Inc.) with all the probe relevant annotations. As there were only three replicates in each group, probe sets with more than one missing value in any group were removed from the analysis. Data from biological replicates were grouped into treatment categories. The clustering of treatment groups was assessed using principal component analysis (PCA). A one-way ANOVA comparing control (mock and GAPDH) and TCF4 knockdown (KD1, KD2) groups was then performed to identify differentially expressed genes. A false discovery rate (FDR) was used to correct for multiple testing (FDR < 0.01, 1204 differentially expressed genes; FDR < 0.05, 5374 differentially expressed genes). The FDR 0.01 corrected gene list was used for enrichment analysis (Section 2.10.3) although the FDR 0.05 corrected gene list was used to identify additional signaling pathway components. The top 1204 differentially expressed genes (FDR < 0.01) were subjected to hierarchical clustering to confirm the validity of the experimental grouping in the one-way ANOVA. Gene expression changes from the array were validated using TaqMan assays (Section 2.9.5). The microarray data discussed in Chapter 4 has been

deposited in NCBI's Gene Expression Omnibus (GEO) and is accessible through GEO Series accession number GSE48367.

2.9.10 Affymetrix Microarray

Gene expression profiling was performed on a GeneChip® Human Gene 2.0 ST Array (Affymetrix Inc.) by Central Biotechnology Services (CBS), Cardiff University according to manufacturer's instructions. The Human Gene 2.0 ST Array contains 988'528 probes and has a high coverage for each transcript. Probes are designed to bind across multiple exons of each gene, providing a gene-level summary of expression. Four biological replicates from six treatment groups (mock, GAPDH KD, TCF4B KD1, TCF4B KD2, TCF4A KD1, TCF42 KD2) were sent for analysis. The concentration and quality of the RNA for each sample was assessed using the RNA 6000 Nano Kit and the Agilent 2100 bioanalyser (Agilent Technologies). These tools provided an RNA Integrity Number (RIN) for each sample prior to microarray analysis. Sense strand cDNA was prepared using the Ambion WT Expression Kit (Applied Biosystems). Sense strands were fragmented and terminally biotinylated using the GeneChip WT Terminal Labeling Kit (Affymetrix). Labelled cDNA samples were hybridised to the GeneChip Human Gene 2.0 ST Array then washed, stained and scanned according GeneChip Expression Wash, Stain and Scan Protocol. The chips were washed using the Fluidics Station 450 and scanned using the GeneChip Scanner 3000 (Affymetrix).

The Affymetrix microarray data provided by CBS was in the form of .CEL files containing raw signal intensity information. The .CEL file for each sample was imported into Partek Genomics Suite v6.6 (Partek Inc.) with the appropriate probe and transcript annotations (HuGene-2_0-st-v1), downloaded automatically from the Affymetrix website. The robust multiarray averaging (RMA) algorithm was used to summarise the probe-level data to a single value for each probe set (Okoniewski and Miller, 2008). Adjustments were made for probe GC content on pre-background-subtracted values. The exon-level probe sets were summarised to the gene-level using the probe set mean. Data from biological replicates were grouped into treatment categories. The clustering of treatment groups was assessed using principal component analysis (PCA). A one-way ANOVA was used to determine differentially

expressed genes between groups. To identify genes differentially expressed in TCF4-B, a one way ANOVA was performed comparing the control (mock and GAPDH) and TCF4-B knockdown (KDB1 and KDB2) groups. To identify genes differentially expressed in TCF4-A, a one way ANOVA was performed comparing the control (mock and GAPDH) and TCF4-A knockdown (KDA1) groups. The KDA2 knockdown group was omitted from the analysis due to a lack of gene expression generated by this siRNA. A false discovery rate (FDR) of 0.01 was used to correct for multiple testing and generate gene lists for enrichment analysis (Section 2.10.3). The most significant gene expression changes (FDR < 0.01) were subjected to hierarchical clustering to confirm the validity of the experimental grouping in the one-way ANOVA. Gene expression changes from the array were validated using TaqMan assays (Section 2.9.5).

2.10 Bioinformatics

2.10.1 General

DNA and protein sequence homology searches were performed using the BLAST (Basic Local Alignment Search Tool) program. Gene and protein sequences were obtained using the NCBI's CCDS (Consensus Coding Sequence) and Refseq (Reference Sequence) databases and the Ensembl database. Alignment of the nucleotide and protein sequences was performed with ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). DNA to protein translation and molecular weight prediction was carried out using the ExPASy (Expert Protein Analysis System) proteomics server from the Swiss Institute of Bioinformatics (http://www.expasy.org/). Pair-wise linkage disequilibrium between single nucleotide polymorphisms (SNPs) was calculated using the SNAP (SNP Annotation and Proxy Search) tool (http://www.broadinstitute.org/mpg/snap/ldsearchpw.php). Graphics for genome wide association data were obtained using Ricopili tool with the "PGC SCZ GWAS (Sept. 2011)" data set (http://www.broadinstitute.org/mpg/ricopili/). The biological functions of genes were annotated using the BP_FAT option in DAVID v6.7 (Database for Annotation, Visualisation and Integrated Discovery) (Huang da et al., 2009) (http://david.abcc.ncifcrf.gov/home.jsp). For a broader classification of biological processes, GOSlim (Gene Ontology Slim) annotations were used from the AgBase GOSlimViewer (McCarthy et al., 2007) (http://www.agbase.msstate.edu/cgi-

bin/tools/goslimviewer_select.pl). Gene and protein accession numbers were converted to Entrez gene IDs using the batch converter tool in DAVID v6.7. Prediction of phophorylation sites was performed with PhosphoMotif Finder (http://www.hprd.org/PhosphoMotif_finder).

2.10.2 Design of promoter constructs

The CNTNAP2 and NRXN1β promoter regions were identified using a bioinformatics approach. Mouse and human genomic DNA sequences up to 5 kb upstream of the predicted transcriptional starts for the CNTNAP2 and NRXN1β genes were compared the UCSC genome browser. Evolutionarily similar regions that also contained conserved E-box motifs were used to select the putative promoter sequences. The UCSC genome browser was used to verify the epigenetic marks (H4K3me3) indicative of transcriptional start sites. Using these methods a 2 kb and a 600 bp region directly 5' of the CNTNAP2 and NRXN1β transcriptional start sites were identified that would act as putative promoters in the luciferase assays (Section 2.8.3). These genomic regions were amplified from human gDNA, sequence verified and subcloned into the pGL3-basic reporter construct (Promega).

2.10.3 Process network and gene ontology (GO) enrichment Analysis

Process network enrichment analysis was performed in MetaCore^{TM} (GeneGo, Thomson Reuters) using the enrichment analysis workflow. Biological process GO term enrichment analysis was performed using the GOTERM_BP_FAT option in the Database for Annotation, Visualisation and Integrated Discovery (DAVID) v6.7 (Huang da et al., 2009). Entrez Gene IDs were used for all analyses to avoid the redundancy of gene symbols. For enrichment analysis of microarray data, differentially expressed genes from high confidence lists of (FDR 0.01) were exclusively used unless otherwise stated. All gene lists were compared to a background list of genes related to the microarray type. For the Toray microarray, the background list consisted of Entrez Gene IDs from all genes that were detected by the array (18966 IDs). For the Affymetrix microarray, the background list consisted of all the probes on the array that had Entrez Gene IDs, since absolute gene expression values cannot be calculated in Partek Genomics Suite 6.6 for this array type (25957 IDs).

The enrichment analysis categorises the input data into different process network (MetaCore) or biological process (DAVID) ontologies. The significance of the enrichment is evaluated based on the size of the intersection between the selected data and the entire ontology it has been categorised in. The resulting *P* value represents the probability that this intersection has occurred by chance, considering the numbers of genes in the input data, the background list of genes (see above) and the number of genes represented in the entire ontology term. An FDR correction of 0.05 was used to determine significantly enriched ontology terms. The gene lists used in enrichment analyses are summarised below:

List name	N° genes	Description	Array type	Analysis
TCF4KD	1031	All differentially expressed genes in global TCF4 knockdown experiment (FDR 0.01)	Toray (Chapter 4)	MetaCore
TCF4KD_Up	425	All upregulated genes in global TCF4 knockdown experiment (FDR 0.01)	Toray (Chapter 4)	DAVID
TCF4KD_Down	606	All downregulated genes in global TCF4 knockdown experiment (FDR 0.01)	Toray (Chapter 4)	DAVID
KDB	386	All differentially expressed genes in TCF4-B isoform knockdown experiment (FDR 0.01)	Affymetrix (Chapter 5)	MetaCore
KDB_Down	287	All downregulated genes in TCF4-B isoform knockdown experiment (FDR 0.01)	Affymetrix (Chapter 5)	MetaCore
KDA1	1319	All differentially expressed genes in TCF4-A isoform knockdown experiment (FDR 0.01)	Affymetrix (Chapter 5)	MetaCore
KDA1_Down	830	All downregulated genes in TCF4-A isoform knockdown experiment (FDR 0.01)	Affymetrix (Chapter 5)	MetaCore

Chapter 3

Functional Analysis of TCF4 Missense Mutations that Cause Pitt-Hopkins Syndrome

3.1 Introduction

As mentioned in Chapter 1, Pitt-Hopkins syndrome (PTHS) is a rare genetic disorder characterised by heterozygous loss-of-function of the *TCF4* gene (Section 1.11). The most prominent symptoms of the disorder are severe intellectual disability, dysmorphic features and breathing abnormalities although many other symptoms have been described in patients including developmental delay with impaired speech, epilepsy, stereotypic movements constipation and ocular anomalies (Pitt and Hopkins, 1978; Amiel et al., 2007; Zweier et al., 2007).

There are currently 112 clinically diagnosed PTHS patients described in the literature and most have discrete mutations affecting different parts of the *TCF4* gene. The spectrum of mutations includes whole gene deletions (23 patients) and partial gene deletions (10 patients) in addition to frameshift (33 patients), splice-site (6 patients), nonsense (18 patients) and missense mutations (22 patients) (Whalen et al., 2012a). The diversity of mutations found in patients and the lack of a genotype-phenotype correlation for the disorder strongly suggests haploinsufficiency is the pathological mechanism in PTHS.

In addition to *TCF4*, recessive mutations in the genes encoding the neuronal cell adhesion molecules *CNTNAP2* (OMIM 610042; contactin-associated protein like 2) and *NRXN1* (OMIM 614325; neurexin) have also been identified in cases with PTHS symptomatology but without *TCF4* mutations (Zweier et al., 2009). Importantly, *TCF4*, *NRXN1* and *CNTNAP2* have also been implicated in the genetic etiology of several other neuropsychiatric disorders suggesting they are crucial for neurodevelopment (Stefansson et al., 2009; Blake et al., 2010; Penagarikano et al., 2011; Ripke et al., 2011). Rare deletions in *NRXN1* have been

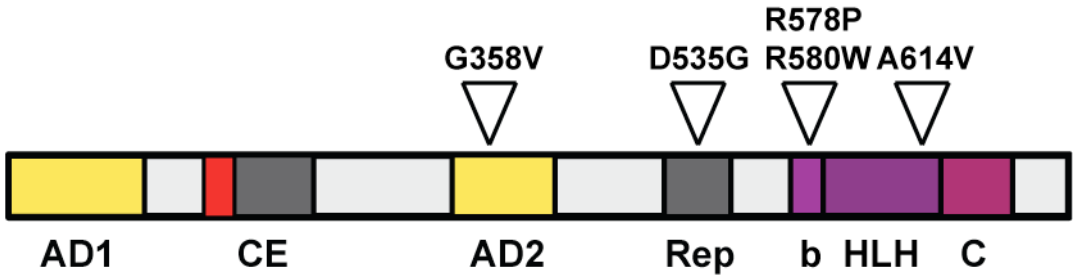
found in patients with schizophrenia and genetic variation in *CNTNAP2* has been linked to a spectrum of neurodevelopmental disorders including schizophrenia, autism and attention deficit hyperactivity disorder (Alarcon et al., 2008; Friedman et al., 2008; Kirov et al., 2009; Elia et al., 2010). The overlap of these genetic loci in at least two separate disorders suggests a shared mechanism involving *TCF4*, *NRXN1* and *CNTNAP2* regulating key neurodevelopmental events (Blake et al., 2010).

At the outset of this work, the functional analysis of PTHS-associated *TCF4* mutations was limited. Characterisation of *TCF4* mutations was restricted to assessing their impact on gene activation using synthetic promoter constructs containing multiple E-boxes. Two studies focused on truncating mutations (G232fsX25, S508fsX5, R385X, G500X) and missense mutations (R576W, R572G, R576Q) concluding that all mutants displayed a similar reduction in activity on these synthetic promoters (Zweier et al., 2007; de Pontual et al., 2009). These studies were the first to report functional deficits related to PTHS-associated mutations although further characterisation is required to fully delineate the role of *TCF4* mutations in PTHS.

To improve our understanding of *TCF4* mutations in PTHS, I examined the effect of five discrete missense mutants using a range of complementary techniques (Zweier et al., 2007; de Pontual et al., 2009) (Figure 3.1). Missense mutations were investigated rather then truncating mutations as they offer additional insight into the different functional domains of the TCF4 protein. Each mutation was assessed for its effect on basic biochemical properties relating to TCF4 function namely: subcellular localisation, transcriptional activity and dimerisation with other bHLH proteins.

The work described in this chapter has been published in *Human Mutation,* December 2012, vol. 33, issue 12, pages 1676-86 (Forrest et al., 2012).

Figure 3.1. TCF4 protein structure and location of PTHS-associated missense mutants characterised in this study. The illustration shows the conceptual organisation of the TCF4 protein. Triangles approximate positions of missense PTHS mutants according to domain architecture. The table below indicates the exonic location of each mutation, the nomenclature used for mutations and the original references where mutations were identified. Amino acid numbers and exons represent the positions on the full length coding sequence of TCF4-B^{+} (671 amino acids including RSRS insertion). Abbreviations: AD1, activation domain 1; AD2, activation domain 2; CE, CE repressor domain; NLS, nuclear localisation signal; Rep, Rep repressor domain; b, basic domain; HLH, helix-loop-helix domain; C, domain C. Illustration not to scale.



Coding exon	Amino acid	Mutation	Mutant name	Reference
13	358	Gly to Val	G358V	(Zweier et al., 2008b)
16	535	Asp to Gly	D535G	(de Pontual et al., 2009)
17	578	Arg to Pro	R578P	(Zweier et al., 2008b)
17	580	Arg to Trp	R580W	(Zweier et al., 2008b)
17	614	Ala to Val	A614V	(de Pontual et al., 2009)

3.2 Results

3.2.1 Basic domain mutants have aberrant nuclear localisation

Of the twelve different *TCF4* missense mutations described in PTHS most cluster in the bHLH domain (Amiel et al., 2007; Zweier et al., 2007; de Pontual et al., 2009; Takano et al., 2010; Whalen et al., 2012a). However some mutations (G358V and D535G) are found in other conserved regions of the protein where their effect on protein function is unknown. To examine the effect of different mutations on the function of TCF4, five distinct mutants were selected (G358V, D535G, R578P, R580W, A614V) which affect different regions of the protein (Figure 3.1). The sub-cellular localisation of wild type TCF4 and the PTHS-associated missense mutants was determined using indirect immunofluorescence of transfected COS-7 cells expressing GFP-tagged proteins (Figure 3.2).

Wild-type TCF4 was detected uniformly throughout the nucleus but was excluded from the nucleoli (Figure 3.2). The mutants G358V and D535G were also expressed in the nucleus with a pattern indistinguishable from the wild type. Unexpectedly, the mutants R578P, R580W and to a lesser extent A614V, were apparently mislocalised to small, spherical punctae that were dispersed throughout the nucleus and reminiscent of nuclear speckles (Figure 3.2). To determine whether these mutants were indeed present in nuclear speckles, transfected cells were co-stained with an anti-SC-35 antibody to detect this nuclear sub-compartment (Mao et al., 2011). The data shows that SC-35 does not co-localise with the PTHS-mutants (Figure 3.2). Similar results were obtained when these mutants were transfected into the SH-SY5Y neuroblastoma cell line and when myc-tagged TCF4 mutants were transfected (data not shown). Therefore the R578P, R580W and A614V mutations affect the nuclear localisation of the protein.

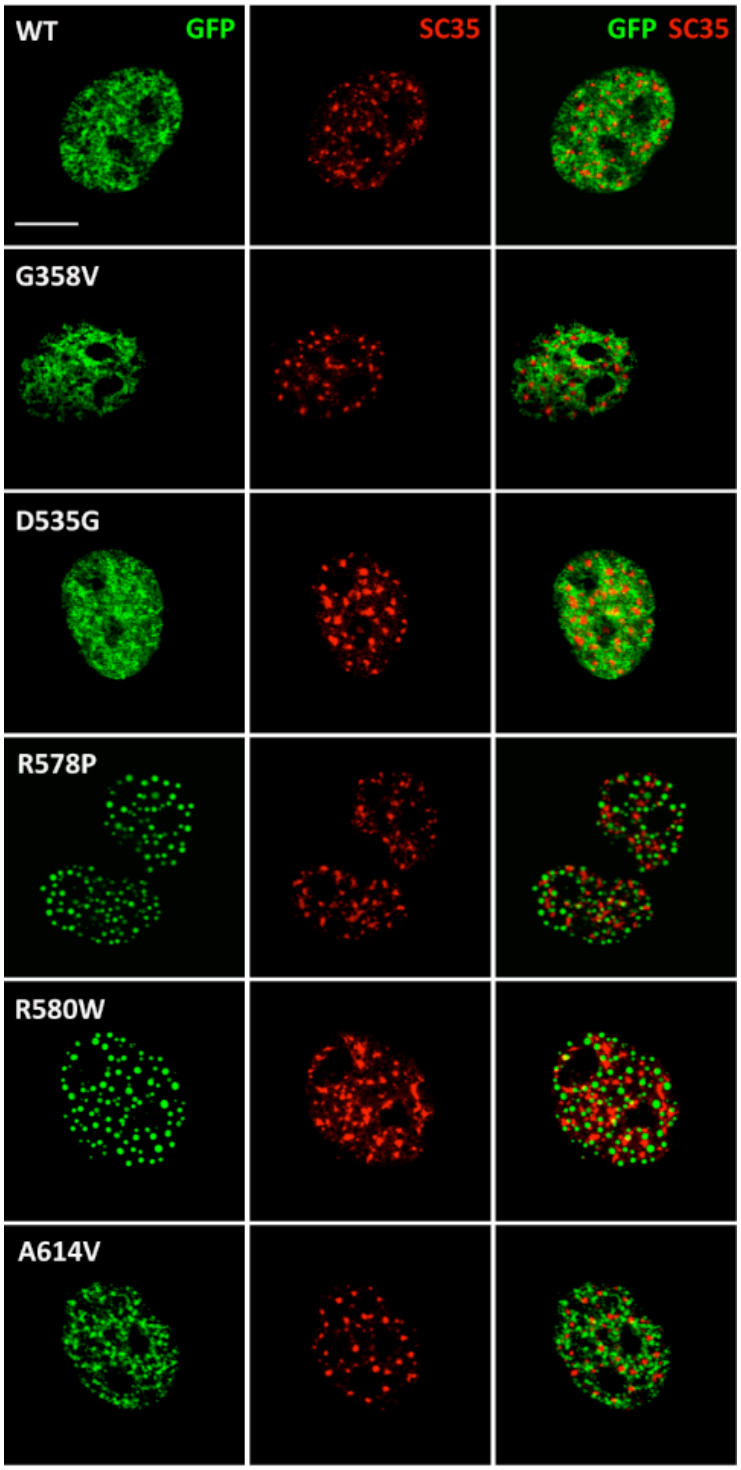


Figure 3.2 Subnuclear localisation of wild type and mutant TCF4.
COS-7 cells transiently expressing GFP-TCF4 or GFP-PTHS mutants were fixed and stained with an anti-SC-35 antibody that detects nuclear speckles (red). Wild-type TCF4 is found in a diffuse pattern throughout the nucleus but is excluded from the nucleolus. By contrast, he mutants R578P, R580W and to a lesser extent A614V form small, spherical punctae differing markedly from the distribution of wild type TCF4 and the mutants G538V and D535G. SC-35 staining does not appear to co-localise with R578P, R580W or A614V mutants indicating that the mutant proteins are not mislocalised to nuclear speckles (merged image). Scale bar, 10μm.

3.2.2 Characterisation of the HTRF system to study bHLH dimerisation

Homo- and hetero-dimerisation are intrinsic properties of bHLH transcription factors (Powell and Jarman, 2008). In common with many transcription factors, TCF4 dimerisation is essential for physiological activity and is considered to be a critical determinant of its biological function. To study this activity in cells, a homogenous time resolved fluorescence (HTRF) assay was developed for TCF4-dependent protein-protein interactions since E-proteins and their binding partners can form stable heterodimers in solution and in the absence of DNA (Longo et al., 2008). This method combines fluorescence resonance energy transfer (FRET) with time-resolved measurements to provide a quantitative output for protein-protein interactions (Degorce et al., 2009). HTRF measures FRET between a donor (europium cryptate) and an acceptor molecule (XL665) only when the two fluorophores are in close proximity (5 to 10 nm) to each other (Figure 3.3). The intensity of the HTRF signal can be used as a measure of protein dimerisation because bHLH transcription factors are in close proximity when they dimerise (Ellenberger et al., 1994). In the following experiments the HTRF signal was measured at 665nm and then normalised to the value ΔF which represents the signal:noise ratio (665nm/620nm). ΔF can be used to compare across experimental conditions and time points making it a useful reference measure.

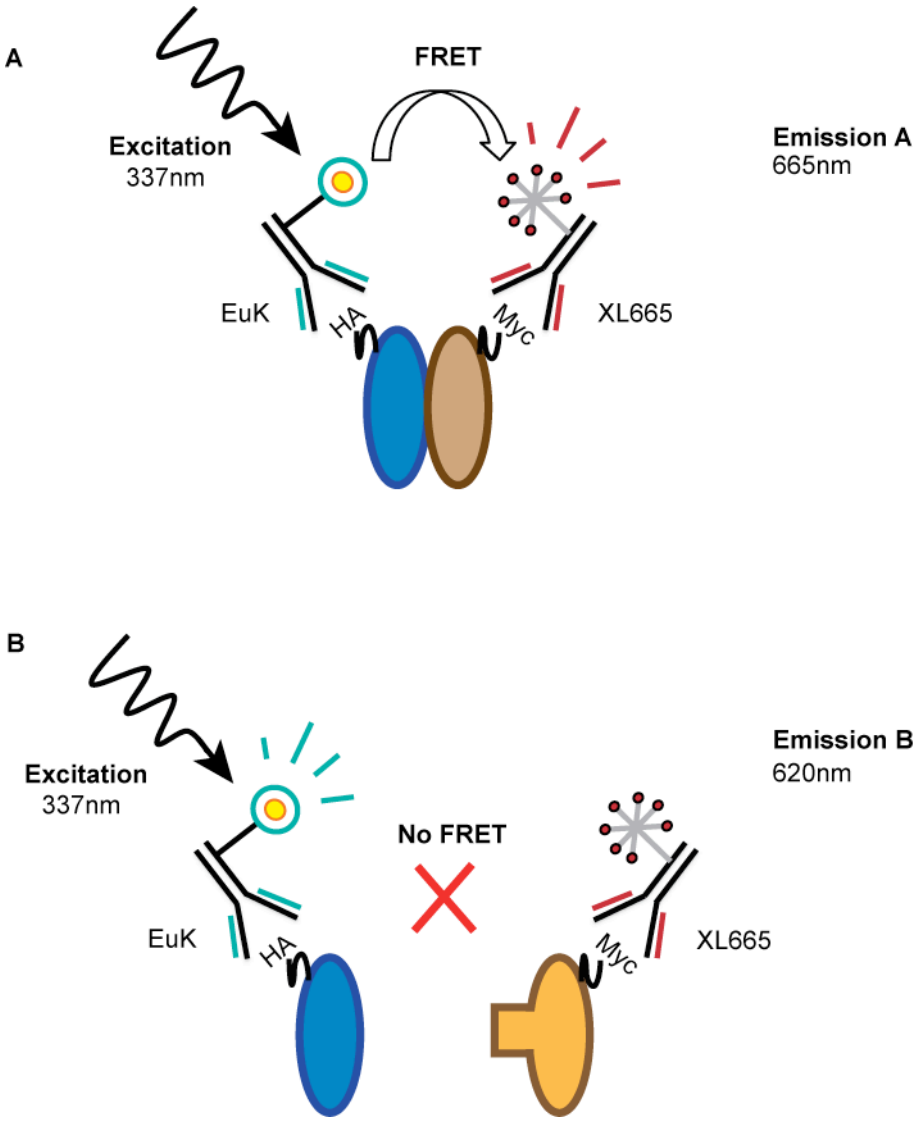


Figure 3.3 Principles of homogenous time-resolved fluorescence (HTRF). HTRF combines standard fluorescence resonance energy transfer (FRET) technology with time-resolved measurement of fluorescence, eliminating short-lived background fluorescence. FRET is based on the transfer of energy between two fluorophores (donor and acceptor) when in close proximity. Conjugated antibodies can be used to direct the donor (europium cryptate, EuK) or acceptor (XL665) fluorophores to epitopes of interest. **A)** Interacting protein pairs that are in close proximity emit a FRET signal at 665 nm when excited. **B)** Non-interacting proteins do not emit a FRET signal (665 nm) when excited and only emit a background signal (620 nm). Protein interactions can therefore be quantitatively measured by the intensity of the FRET emission signal. HTRF data is reported at as the ratio of between the FRET (665 nm) and the background (620 nm) signal (ΔF). Abbreviations: HA, Hemagglutinin-epitope tag; Myc, c-myc-epitope tag.

In order to ensure a FRET signal between wild type TCF4 homo- and hetero- dimers was possible, a series of published TCF4 interactors were cloned into pCMV-HA (HA-TCF4, HA-ASCL1, HA-ATOH1, HA-NEUROD1, HA-ID2) and co-transfected individually with a myc-TCF4 construct in HEK 293T cells. 24 h after transfection, the cells were lysed in NP-40 buffer, and the processed whole cell lysates were subsequently incubated with anti-HA XL665 and anti-myc europium cryptate antibodies. The binding of acceptor and donor fluorophores to each tag resulted in a 665nm emission signal from each condition and demonstrated that a robust signal could be detected for interacting pairs (Figure 3.4). It is noteworthy that although the HA-TCF4, HA-ATOH1 and HA-ASCL1 constructs have a marked difference in their HTRF signal, each one was expressed at a similar level when visualised by western blotting (data not shown). In contrast, HA-NEUROD1 that had the strongest expression, and HA-ID2 that had the lowest expression, generated the highest and lowest HTRF signals respectively. These two observations suggest that both the identity and expression levels of each interactor contribute to the HTRF signal.

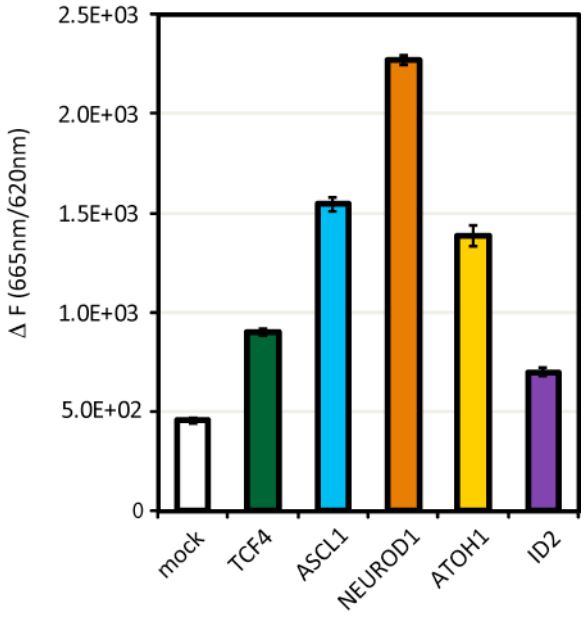


Figure 3.4 Characterisation of wild type TCF4 homo- and hetero- dimers by HTRF
Myc-tagged wild type TCF4 was co-transfected with a series of HA-tagged binding partners (TCF4, ASCL1, NEUROD1, ATOH1, ID2) in HEK-293T cells to assess the strength of the HTRF signal for TCF4 dimerisation. The acceptor and donor fluorophore-conjugated antibodies, which recognise the epitope tags, generated a signal significantly over background (mock) for each interaction, confirming the close proximity of proteins in the dimer. This data substantiates HTRF as a tool to measure bHLH dimerisation. The mock sample consists of mock transfected cell lysates incubated with the acceptor and donor antibodies. Data is reported as the mean of 4 replicates ± standard error of the mean.

3.2.3 Impact of PTHS-associated mutations on homodimerisation

Once the protocol for wild type TCF4 dimerisation was established, DNA for mutant TCF4 sequences were cloned into pCMV-myc vectors and expressed in HEK-293T cells to ensure they generated a viable protein product (Figure 3.5). Although there was variability in the expression of each construct, each of the six transfected vectors was efficiently translated to a protein of the same size. As each of the five mutant sequences all originated from the same sequence and vector, the observed differences in expression are likely to be a direct consequence of the mutations' effect on protein stability.

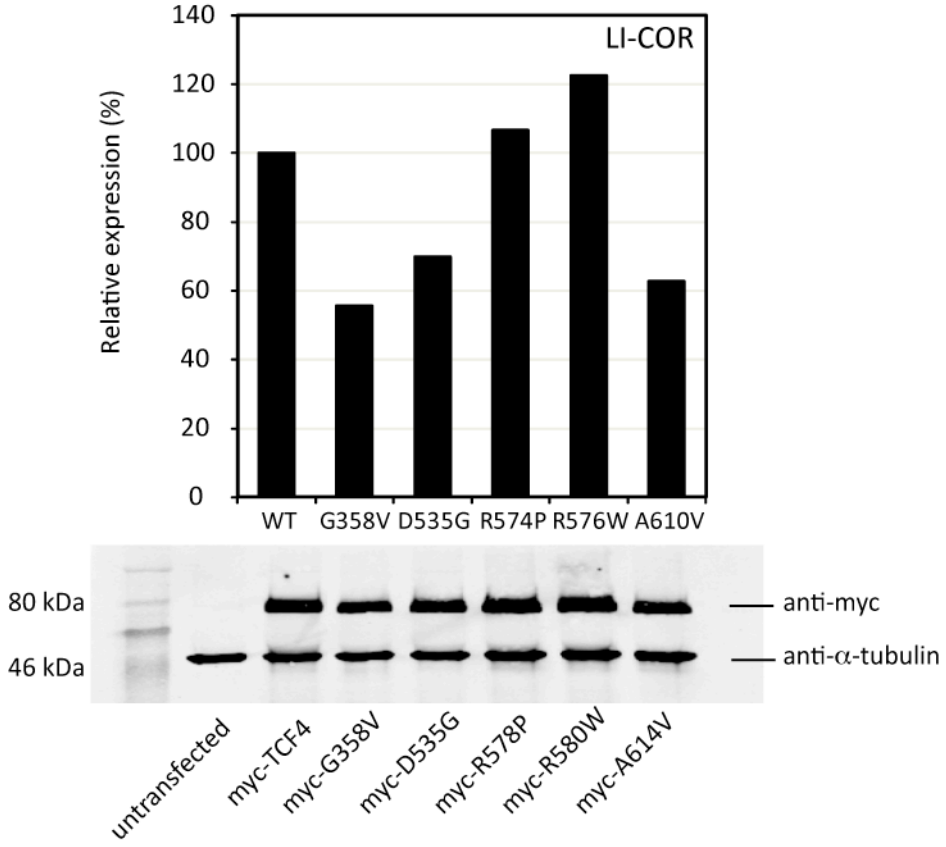


Figure 3.5 Relative expression of wild type and mutant TCF4 constructs.
Wild type TCF4 and its associated PTHS mutants were cloned into pCMV-myc and expressed for 24h in HEK-293T cells. The cells were subsequently lysed and the proteins were separated on a 12% polyacrylamide gel before western blotting. (Top) Relative expression of wild type (WT) and mutant TCF4 constructs. Band intensities for all TCF4 constructs and for α-tubulin were measured with the LI-COR odyssey software. The relative expression of each construct was calculated by normalising TCF4 expression to α-tubulin (see western blot). (Bottom) Western blot of the wild type TCF4 and mutant TCF4 constructs. Detection of α-tubulin was used as a loading control and to calculate relative expression.

To examine the effect of these mutants on homodimerisation with TCF4, HA-tagged wild type protein was co-expressed with myc-tagged TCF4 and the mutants thereof (myc-G358V, myc-D535G, myc-R578P, myc-R580W and myc-A614V) in HEK-293T cells. This experimental design essentially replicates the situation in PTHS patients, with one mutant allele and one wild type allele contributing to the total amount of TCF4 protein. In these experiments, all the mutants generated a reduced FRET signal in comparison to the wild type/wild type interaction (Figure 3.6). Since a robust signal was previously observed from the wild type TCF4 interaction, ΔF was expressed as a percentage of the wild type TCF4 signal for comparative purposes. The largest change in FRET signal was observed with the mutants R578P, R580W and A614V, where the ΔF was reduced to less than 20% of the wild type TCF4 homodimer (Figure 3.6). Importantly, the FRET signal was not abolished by mutations in the bHLH domain indicating that some homodimerisation could still occur between wild type and mutant TCF4. By contrast, the mutations G358V and D535G produced FRET signals that were reduced to approximately 80% of wild type levels (Figure 3.6). Whilst this change is statistically significant, these data suggest that the G358V and D535G mutations do not have a dramatic impact on homodimer formation.

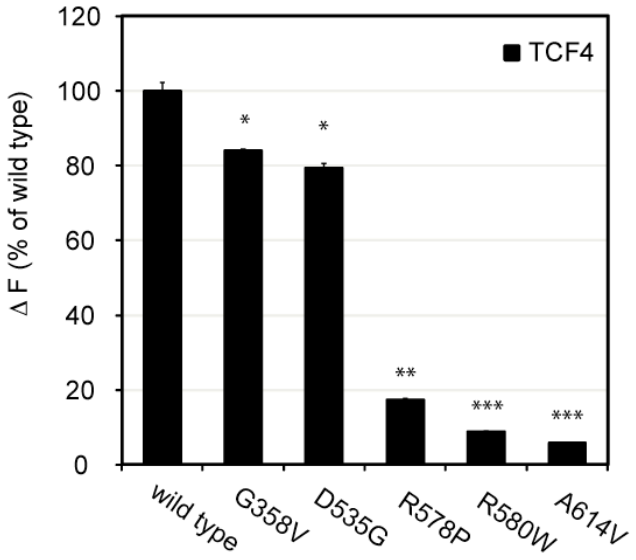


Figure 3.6 HTRF analysis of homodimerisation between TCF4 and PTHS-associated mutants. HEK-293T cells were co-transfected with HA-tagged wild type TCF4 and a panel of myc-tagged PTHS mutants. The association between the protein pairs was measured by HTRF. The 665nm/620nm ratio (ΔF) was recorded for each interaction and expressed as a percentage of wild type activity. Data are expressed as the mean %ΔF ± SEM for three

biological replicates. *P*-values for Student's *t*-test results comparing the PTHS-associated mutants to wild type TCF4 are indicated thus; * $p<0.05$; ** $p<0.001$; *** $p<0.0001$.

3.2.4 Impact of PTHS-associated mutations on heterodimerisation

The HTRF assay was used to examine the effect of the different PTHS-associated missense mutations on the interaction of TCF4 with other bHLH transcription factors, ie. heterodimer formation. Specifically, the association of TCF4 with three proneural proteins ASCL1, ATOH1, NEUROD1 and one inhibitory factor, ID2 was examined. In these experiments quantitative deficits in association for all TCF4 heterodimers tested were detected (Figure 3.7). The damaging effects of these mutations were variable however, dependent both on the identity of the mutant and the binding partner.

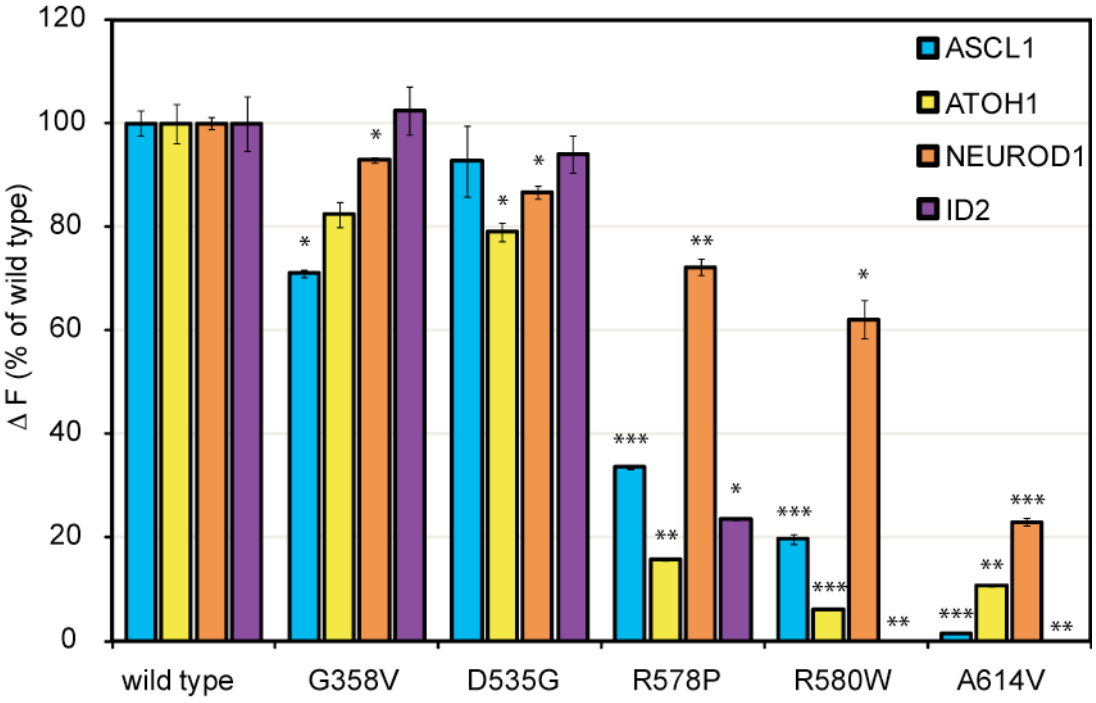


Figure 3.7 HTRF analysis of heterodimer formation between, PTHS-associated mutants and bHLH proteins. HEK-293T cells were co-transfected with HA-tagged bHLH proteins (ATOH1, ASCL1, NEUROD1 or ID2) and a panel of myc-tagged PTHS mutants. The interaction between the two proteins was measured by HTRF and normalised to the wild type signal as previously mentioned (see Figure 3.6). *P*-values for Student's *t*-test results comparing the PTHS-associated mutants to wild type TCF4 are indicated thus; * $p<0.05$; ** $p<0.001$; *** $p<0.0001$.

The association of TCF4 with ATOH1 was largely similar to the effects observed with the TCF4 homodimer (Figure 3.7). Whilst, the mutations G358V and D535G generated a modest change in FRET intensity, mutations affecting the bHLH region of the protein reduced the FRET signal to less than 20% that of wild type TCF4 (Figure 3.7). Similarly, the bHLH mutations caused a large reduction in the FRET signal between TCF4 and ASCL1 (Figure

3.7). However, whilst the G358V mutation resulted in a statistically significant reduction in the FRET signal between the TCF4 mutant and ASCL1, the D535G mutation only reduced the FRET signal by approximately 5% of wild type (Figure 3.7). HTRF analysis of TCF4 binding to NEUROD1 showed that all TCF4 mutations resulted in significant reduction in the FRET signal (Figure 3.7). As before, the more distal mutations produced the largest reduction in FRET signal although in this instance, the two mutations in the bHLH domain R578P and R580W only reduced the FRET signal by approximately 40%. Thus, it can be concluded that each TCF4 mutant can associate with NEUROD1 although the weaker FRET signal suggests that the mutations may reduce the affinity of this interaction. These data are important to consider when interpreting the transactivation of gene expression described below (Section 3.2.5).

In addition to examining the interaction between TCF4 and the proneural genes, the same HTRF assay was used to determine whether TCF4 mutants could interact with ID2, a potent negative regulator of gene expression (Rothschild et al., 2006). Interestingly, the N-terminal mutants essentially behaved like wild type TCF4 and showed almost no difference in FRET signal (Figure 3.7). By contrast, the two most distal mutations, mutations R580W and A614V appeared to abolish the interaction with ID2, whereas the mutant R578P reduced the FRET signal to approximately 20% of wild type (Figure 3.7). These results suggest that residues R580 and A614 are particularly important for ID2 interactions.

In summary, PTHS-mutations have significant effects on the association of TCF4 with a selection of bHLH transcription factors however the mutations differentially affect each interaction.

3.2.5 TCF4 activates the putative CNTNAP2 and NRXN1β promoters

E-proteins such as TCF4 are potent regulators of gene expression that act on E-box-containing promoters. Previous studies on PTHS-mutants have used synthetic multimerised E-box sequences cloned upstream of a reporter to assess TCF4 activity (Zweier et al., 2007; de Pontual et al., 2009). These types of constructs have very little resemblance to *in vivo* promoter structures making comparisons to physiological gene regulation difficult. Sequences flanking E-boxes are crucial to their activity and participate in recruiting specific transcription factor complexes to individual regulatory elements (Castro et al., 2006; Powell and Jarman, 2008). To examine the effect of the PTHS mutations on transcriptional activity,

the mutant's properties were characterised on physiological promoters that may be relevant to PTHS. It was hypothesised that TCF4 may regulate the expression of the *NRXN1* and *CNTNAP2* genes since genetic variation in *TCF4*, *NRXN1* and *CNTNAP2* can result in similar neuropsychiatric disorders including PTHS.

To predict the location of the *NRXN1* and *CNTNAP2* promoters, several bioinformatic tools were used to identify characteristic elements upstream of the transcriptional start site (TSS). The first strategy was to search for highly conserved regions anywhere 5000bp upstream of the TSS, because cross-species conservation is an indicator of functionally important DNA elements. E-box sequences appear frequently within the genome therefore putative promoter sequences were selected if they contained E-box sequences that were conserved between the human and the mouse genome. This approach made it more likely that the promoters contained functional E-boxes. For the *NRXN1* gene, which encodes two main isoforms (*NRXN1α* and *NRXN1β)*, putative sequences upstream of both isoforms were identified. *NRXN1α* is the longer of the two isoforms and is the most affected by copy number variation in schizophrenia patients (Kirov et al., 2009). Although a conserved region was found upstream of the *NRXN1α* TSS, no conserved E-boxes were present in this sequence. In contrast, the sequence directly upstream of *NRXN1β* contained a conserved E-box. Hence, the *NRXN1β* promoter region was chosen for transcriptional assays over the *NRXN1α* promoter region. A region approximately 600bp upstream of the *NRXN1β* gene and a region approximately 2500bp upstream of the *CNTNAP2* gene were selected (Figure 3.8). Each of these sequences was highly conserved across species and contained at least one E-box conserved between the human and the mouse genomes. Putative promoter regions were then overlaid with epigenetic information from the UCSC human genome browser. H3K4me3 methylation sites were specifically selected as these are indicative of active promoter elements. The data from the UCSC genome browser provided additional evidence that these sequences were functional promoter elements (Figure 3.8) however it remained to be determined whether these sequences could be transactivated by TCF4.

To examine the effects of TCF4 on the putative *NRXN1β* and *CNTNAP2* promoters, each region of gDNA was amplified by PCR and cloned bi-directionally in to the pGL3-basic vector to drive the expression of a luciferase reporter. Robust basal transcription was observed only when the putative promoter was in the forward orientation as opposed to the

reverse configuration (Figure 3.9). To determine whether the proneural genes were able to activate these promoters, each transcription factor was expressed alone and in combination with TCF4. Whilst each proneural gene was found to weakly drive luciferase expression, co-transfection of TCF4 strongly activated expression of the forward promoter sequences, consistent with these factors requiring an E-protein for full activation (Figure 3.9). Comparatively, total activation was higher for the *NRXN1β* promoter (a maximum of 1.2 RLU) compared to *CNTNAP2* (a maximum of 0.35 RLU) (Figure 3.9). The luciferase activity for both promoters was dependent on the levels of TCF4 in the transfection (Figure 3.10). Together, these experiments indicate that TCF4 is able to regulate the selected *CNTNAP2* and *NRXN1β* promoter regions *in vitro*.

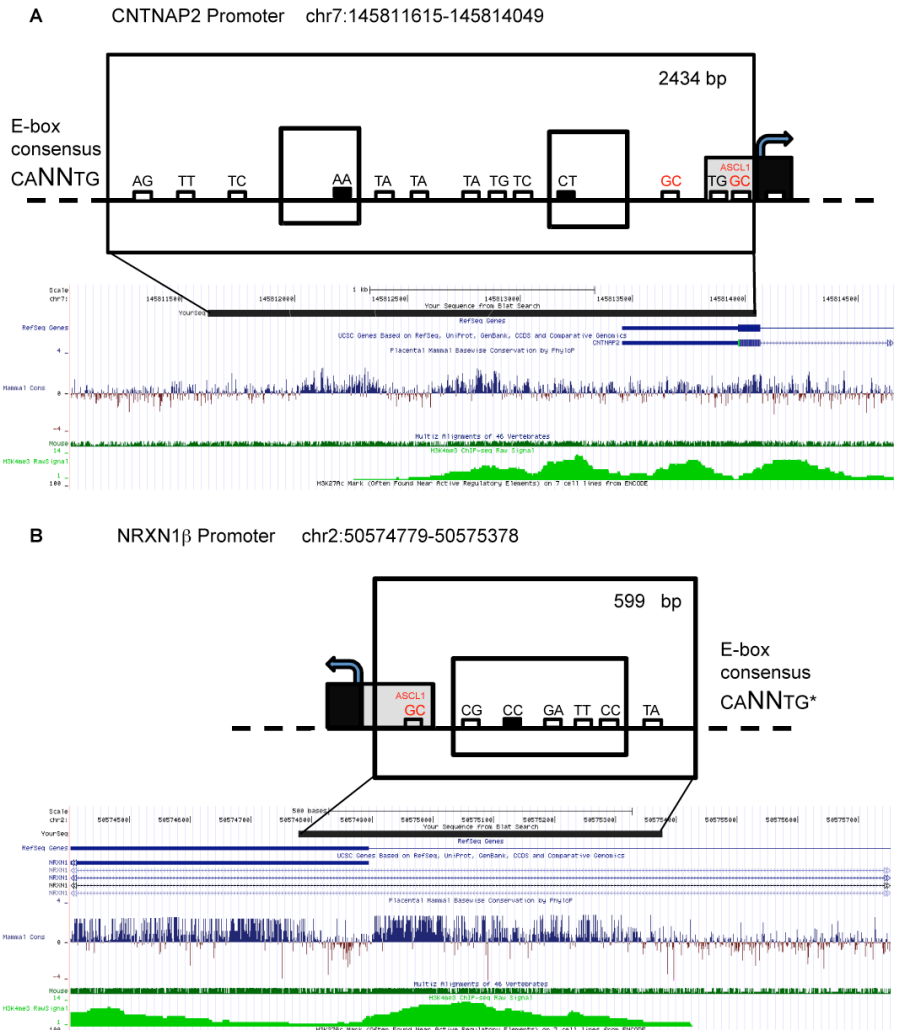


Figure 3.8. Genomic view of the *NRXN1β* and *CNTNAP2* promoters.
Large black boxes delineate the region cloned into the pGL3-basic vectors to create the luciferase reporter constructs. The medium sized black boxes represent regions of homology between human and mouse genomes. The small boxes represent layout of conserved (black) and non-conserved (white) E-boxes (CANNTG) across the promoter. The dinucleotides above each E-box correspond to the variant residues in the E-box consensus. ASCL1 has a preference for GC at the variable position of the E-box consensus as indicated in red (Castro et al., 2011). In addition to ASCL1, E-protein homodimers are known to bind to E-boxes containing CC at the variable position (Henthorn et al., 1990). Two of these E-boxes are located in the *NRXN1β* promoter and may mediate the potent dose-dependent transactivation of this promoter when co-expressed with TCF4 (Figures 3.9 and 3.10). Consensus sequences for the other HLH factors used in this study are poorly characterised and have not been annotated. Arrows indicate the location of start of transcription and beginning of the first coding exon (black) downstream of the 5' untranslated region (grey box). The exact size of each promoter is shown at the top right of the illustration. The chromosomal position of the cloned region is also shown. "Your sequence from Blat search" identifies the promoter for each gene in UCSC genome browser. The screen capture depicts the mammalian conservation plot and the location of H3K4me3 histone modification indicative of active promoters. *For clarity the E-box sequences in the *NRXN1β* promoter are written 5'to 3'.

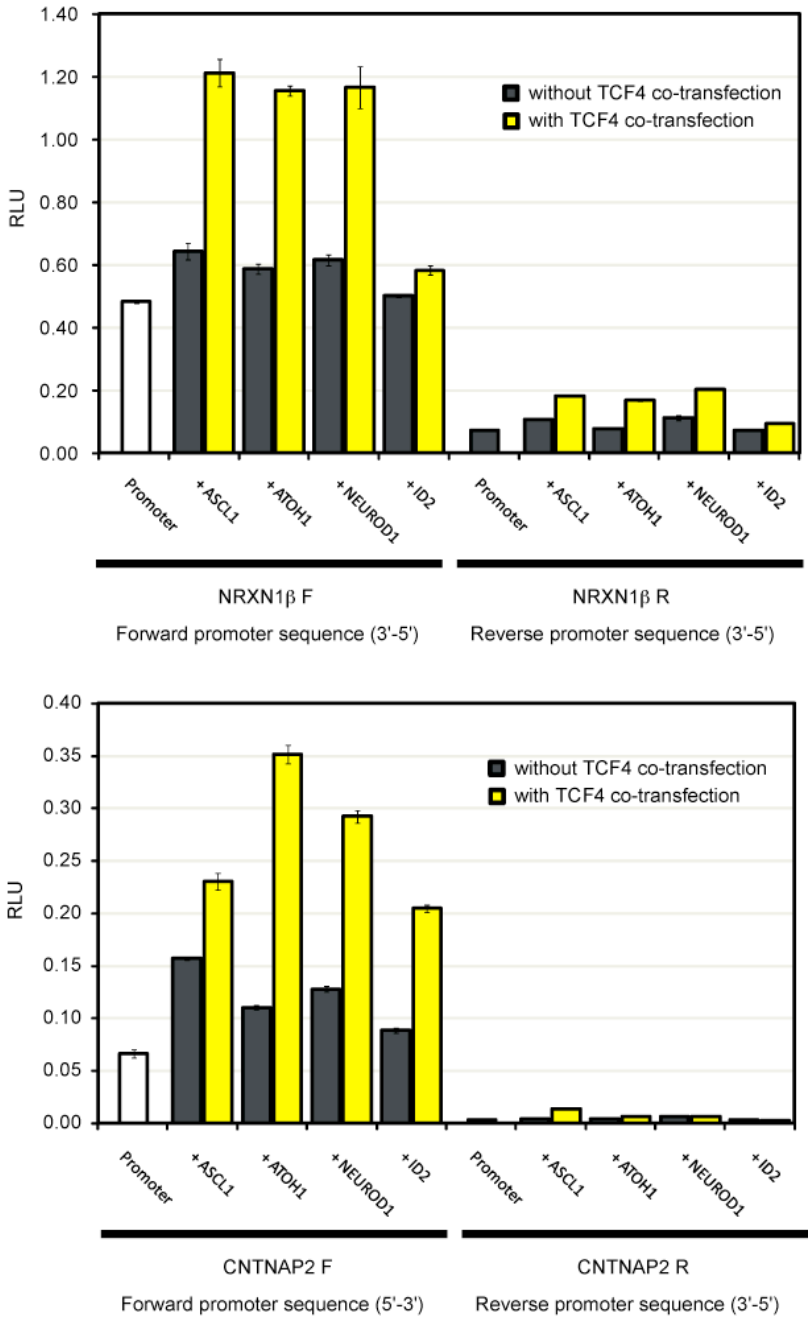


Figure 3.9 Characterisation of the *NRXN1β* and *CNTNAP2* promoters: directionality and activation by bHLH factors. Luciferase assays in HEK-293T cells were used to determine whether bHLH factors co-expressed with TCF4 could activate the *NRXN1β* and *CNTNAP2* promoters. ATOH1, ASCL1, NEUROD1 or ID2 were expressed with either the forward or reverse orientations of the putative *NRXN1β* or *CNTNAP2* promoters in the presence or absence of TCF4. ATOH1, ASCL1 and NEUROD1 enhanced luciferase expression at both promoters. ID2 had little effect on the *NRXN1β* promoter but was able to augment luciferase expression from the *CNTNAP2* promoter by a factor of approximately 1.5 fold. As expected, each bHLH factor resulted in minimal transactivation of the *NRXN1β* and *CNTNAP2* promoters in the reverse configuration. Data are expressed as the mean relative luciferase units (RLU) ± SEM for four biological replicates.

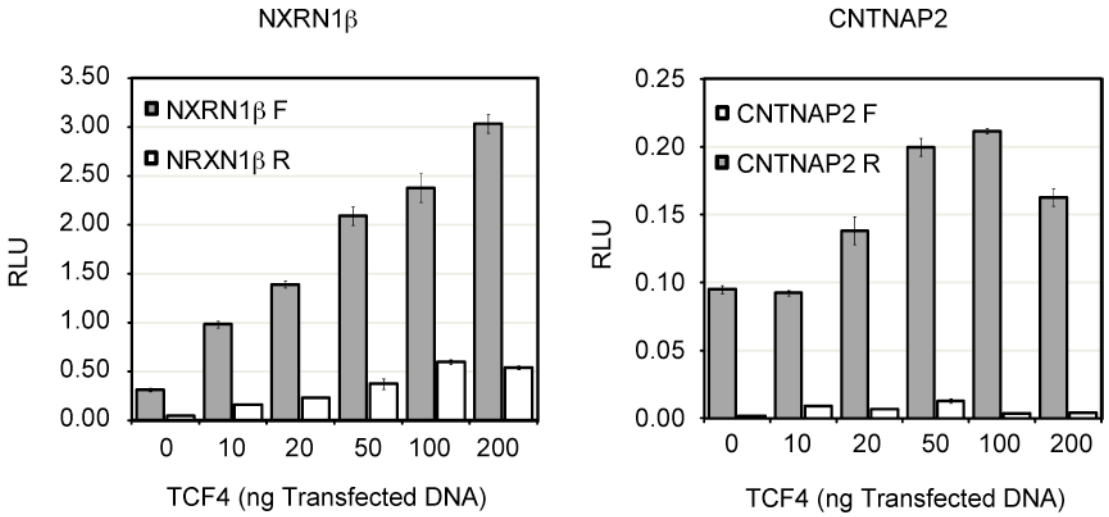


Figure 3.10 TCF4 dose-dependent transactivation of the *NRXN1β* and *CNTNAP2* promoters. Luciferase assays in HEK-293T cells were used to determine whether TCF4 could activate transcription of the putative *NRXN1β* and *CNTNAP2* promoters in a unidirectional manner. Increasing amounts of TCF4 construct were transfected together with the forward (5'-3') and reverse (3'-5') orientations of the putative *NRXN1β* or *CNTNAP2* promoters upstream of a luciferase reporter. Both forward promoter sequences demonstrated TCF4 dose dependence whereas sequences in the opposite orientation were not as sensitive to activation. Data are expressed as the mean relative luciferase units (RLU) ± SEM for four biological replicates.

3.2.6 PTHS mutants have deficits in promoter transactivation that are context-dependent

Having shown that the *CNTNAP2* and *NRXN1β* promoters were responsive to TCF4, the PTHS-associated missense mutants were compared to wild type TCF4 in luciferase assays with homodimer and heterodimer configurations (Figure 3.6 and 3.7). For comparative purposes, each assay was normalised to the activity of wild type TCF4 (expressed as a percentage) for each condition (Figures 3.11 and 3.12). Expressing normalised luciferase activity in this way allows direct comparison between the different co-transfected binding partners and between promoters.

Using the *NRXN1β* promoter, the most damaging effects on transactivation were observed with the DNA-binding mutants R578P and R580W (Figure 3.11). When co-expressed with wild type TCF4, R578P and R580W reduced luciferase activities to approximately 70% compared to the control (Figure 3.11A). This reduction in luciferase activity was statistically significant ($P < 0.001$), whereas the mutants G358V and A614V had no statistically

significant effect on promoter activity. Intriguingly, D535G increased luciferase expression when co-expressed with wild type TCF4.

Co-expression of ATOH1 with each mutant resulted in a significant reduction in luciferase activity for every construct when compared to wild type TCF4 (Figure 3.11B). bHLH mutations had the most damaging effects on promoter transactivation resulting in a reduction of approximately 55% of wild type activity when ATOH1 and R578P were co-expressed (Figure 3.11B). Similar results were also seen with ASCL1, however, G358V appeared to behave as wild type (Figure 3.11C). Co-transfection of NEUROD1 with TCF4 and the PTHS-associated mutants had surprisingly little effect on the *NRXN1β* promoter. The only statistically significant reduction in luciferase activity was observed with the R578P mutant (Figure 3.11D). Importantly, NEUROD1 and TCF4 were already shown to enhance basal expression of the *NRXN1β* promoter, confirming activity at this promoter (Figure 3.9). These data suggest that some of the PTHS-associated missense mutations may retain partial activity at certain promoters highlighting the importance of promoter context in assaying loss of function. It is interesting to note that each TCF4 mutant is able to associate with NEUROD1 (Figure 3.7) potentially explaining the activity of the mutant heterodimers on the *NRXN1β* promoter. Similarly, ID2 had little effect on the activity of TCF4 and its mutants at the *NRXN1β* promoter (Figure 3.11E). Co-transfection of ID2 with G358V or R580W resulted in a modest but statistically significant decrease and increase in luciferase activity respectively (Figure 3.11E).

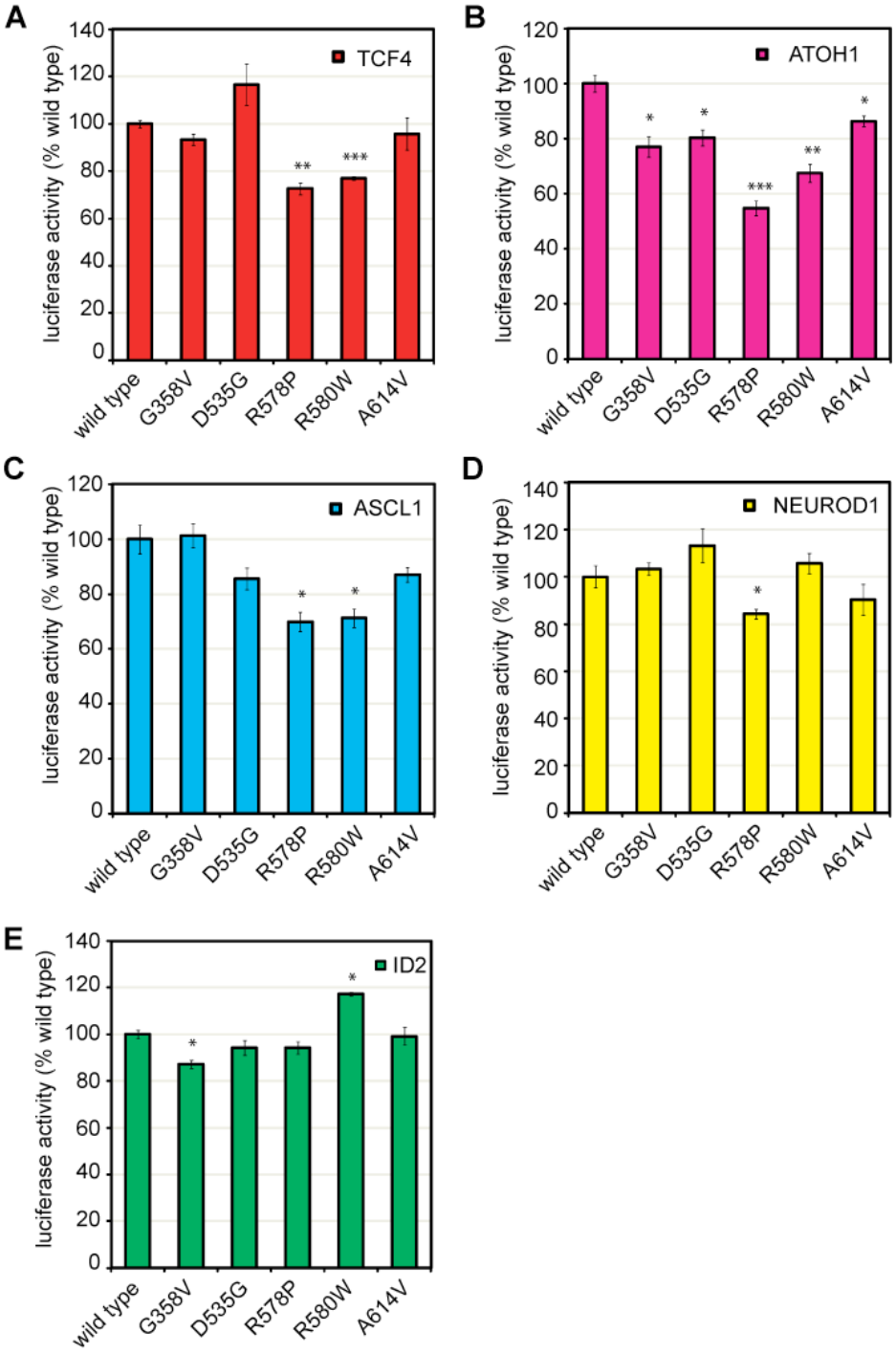


Figure 3.11 Effect of PTHS mutants on *NRXN1β* transactivation.
Luciferase assays were performed to assess the ability of the PTHS mutants to activate transcription from the *NRXN1β* promoter with different bHLH binding partners. Each bHLH protein (TCF4 (A), ATOH1 (B), ASCL1 (C), NEUROD1 (D) or ID2 (E)) was co-transfected with the PTHS mutants and the *NRXN1β* reporter construct. Activity is measured in relative luciferase units (RLU) and is converted to percentage of wild type TCF4 activity. Results are reported as the mean ± SEM of four biological replicates. Student's *t*-test results comparing the mutants to wild type TCF4 are indicated thus, * $p<0.05$; ** $p<0.001$; *** $p<0.0001$.

In addition to the *NRXN1β* promoter, the effects of the different PTHS mutants on the *CNTNAP2* promoter when co-expressed alongside the different bHLH transcription factors were also examined (Figure 3.12). Several important differences between the effects of the PTHS mutants on the two different promoters were observed. D535G increased luciferase expression from the *CNTNAP2* promoter whereas the R578P and R580W mutants reduced luciferase activities to approximately 80% of the wild type homodimer, as was the case with TCF4 activation of the *NRXN1β* promoter (Figure 3.12A). The luciferase activity profile of the PTHS mutants co-expressed with wild type TCF4 or ATOH1 were remarkably similar to those seen with the neurexin promoter (Figure 3.12B). Co-transfection of ATOH1 with R578P and R580W reduced luciferase expression to approximately 40% of wild type TCF4, whereas the other mutants had modest effects on promoter transactivation that was only significant for G358V (Figure 3.12B). Co-expression of ASCL1 with TCF4 and its mutants resulted in a robust, statistically significant reduction in luciferase activity for each construct. These data differ markedly from the previous experiment where ASCL1 and TCF4 were co-transfected with the *NRXN1β* promoter (Figure 3.12C). Co-expression of the A614V mutant with ASCL1 had the most deleterious effect on the *CNTNAP2* promoter activity consistent with the weakest FRET signal indicative of an impaired association between A614V and ASCL1 (Figure 3.7).

Paradoxically, the PTHS-associated TCF4 mutants had very similar effects on the *CNTNAP2* promoter when co-expressed with NEUROD1 or ID2 (Figure 3.12D and E). Whilst G358V and D535G were essentially indistinguishable from wild type TCF4, the DNA-binding domain mutants caused an increase in luciferase activity from the *CNTNAP2* promoter. These data suggest that R578P and R580W function as transcriptional activators at the *CNTNAP2* promoter. By contrast, A614V resulted in a modest reduction in luciferase activity when co-expressed with NEUROD1 whereas it strongly activated the *CNTNAP2* promoter when co-expressed with ID2. As mentioned previously, each of the mutants retained the ability to associate with NEUROD1 in the HTRF assay (Figure 3.7). This data suggests that mutations in the DNA-binding domain of TCF4 can potentially activate transcription when co-expressed with certain bHLH factors at an appropriate promoter. It is also possible that, in certain heterodimer contexts, the DNA-binding mutants remove inhibitory co-factors that participate in the regulation of the *CNTNAP2* promoter. Whilst it is beyond the scope of this

data to speculate about the mechanism for these effects, these experiments highlight clear, context-dependent differences between the PTHS-associated mutants and wild type TCF4.

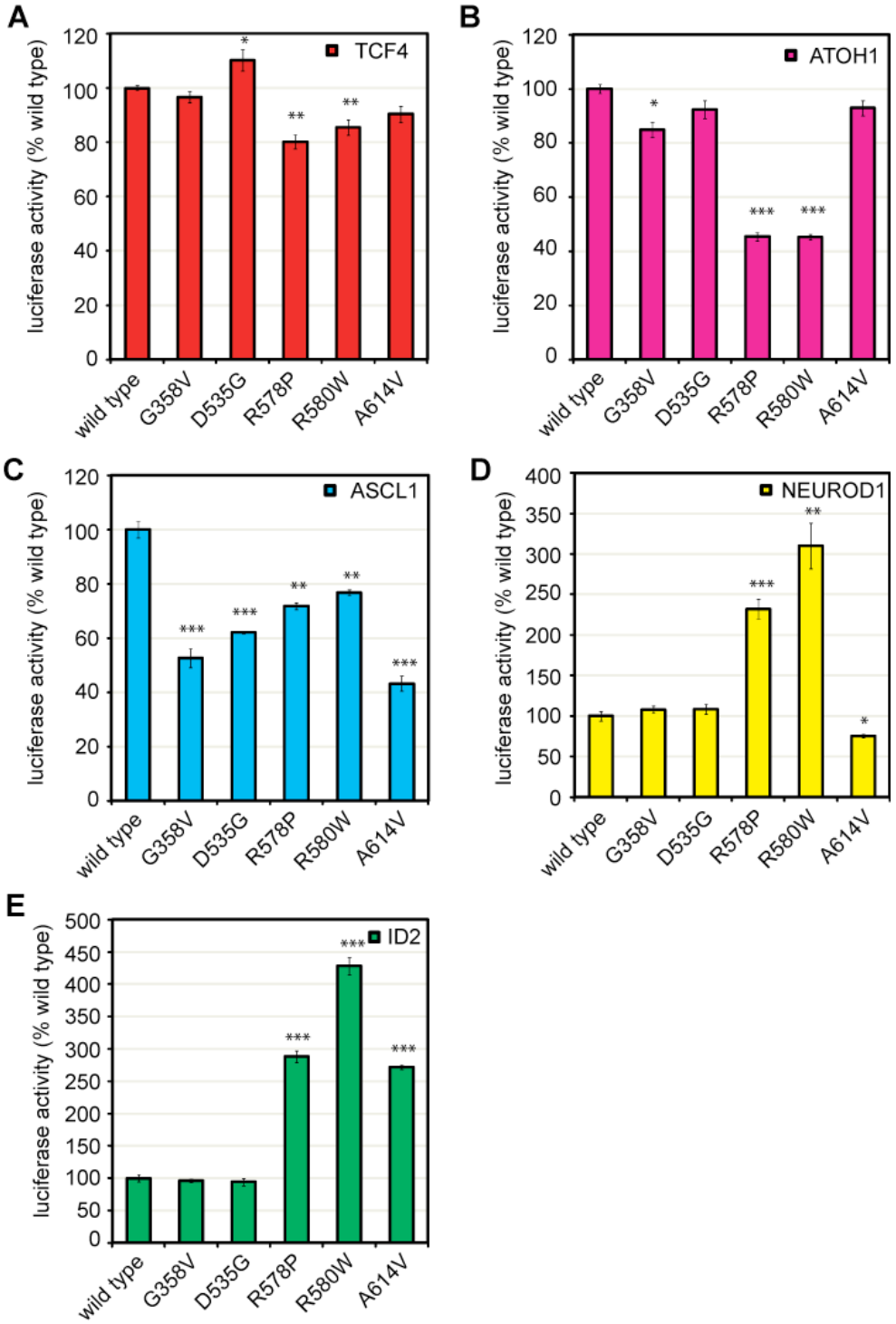


Figure 3.12 Effect of PTHS mutants on *CNTNAP2* transactivation. Luciferase assays were performed to assess the ability of the PTHS mutants to activate transcription from the *CNTNAP2* promoter following the scheme described in Figure 3.11. Each bHLH protein (TCF4 (A), ATOH1 (B), ASCL1 (C), NEUROD1 (D) or ID2 (E)) was co-transfected with PTHS mutants and the *CNTNAP2* reporter construct. Data were analysed as described in Figure 3.11. Student's *t*-test results comparing the mutants to wild type TCF4 are indicated thus, * $p<0.05$; ** $p<0.001$; *** $p<0.0001$.

3.3 Discussion

Autosomal dominant PTHS is caused by mutations in the *TCF4* gene that are thought to lead to haploinsufficiency. Although most PTHS-associated missense mutations alter basic amino acids in the bHLH domain of the TCF4 protein, their impact on the function of the protein has not been extensively studied. To further examine the molecular consequences of PTHS-associated missense mutations on cellular localisation, dimerisation and transactivation, the TCF4 mutants were characterised using several complementary techniques. In addition to the previously described defects in promoter transactivation, *TCF4* mutations were found to affect subcellular localisation and association with other bHLH transcription factors. The data presented in this chapter also provides the first evidence to suggest that TCF4 may regulate the expression of the *NRXN1β* and *CNTNAP2* genes that are mutated in a PTHS-like syndrome, potentially defining an important neurodevelopmental regulatory network. The discussion presented below will summarise the main findings of this chapter and contrast the results to similar work which was published by *Sepp et al.* soon after the completion of these experiments (Sepp et al., 2012a).

3.3.1 TCF4 localisation

Unexpectedly, the mutants R578P, R580W and to a lesser extent A614V were found to alter the subnuclear localisation of the mutant protein. These mutants formed discrete punctae within the nucleus that clearly differed from wild type TCF4 (Figure 3.2). Although these punctae resembled nuclear speckles their failure to co-localise with the nuclear speckle marker SC35 indicated that they must reside in a different compartment in the nucleus. Interestingly, a recently described missense mutation (R244C) in the DNA-binding domain of the forkhead box G1 gene (*FOXG1*), that causes a variant of Rett syndrome, also results in nuclear mislocalisation of the mutant protein; in this case the mutant formed speckles that co-localised with SC35 (Le Guen et al., 2011).

In contrast, Sepp and colleagues did not find that DNA-binding mutants formed nuclear speckle-like structures in transfected cells. This may be explained by methodological differences between the two studies. Sepp *et al*. used an anti-TCF4 antibody to detect the protein (recombinant and endogenous) in transfected cells whilst here, EGFP-tagged constructs were used to differentiate between endogenous and transfected TCF4 (Sepp et al.,

2012a). To confirm that the mislocalisation observed with the EGFP-tagged constructs was due to mutations in TCF4 and not due to an artefact of the EGFP tag, experiments were repeated with the equivalent myc-TCF4 mutants used for HTRF. Immunostaining with anti-myc antibodies resulted in exactly the same mislocalisation observed in the EGFP-tagged constructs (data not shown). Interestingly, a commercial anti-TCF4 antibody (M03, Abnova) was not able to detect the speckle-like punctae in cells transfected with the DNA-binding mutants R578P, R580W (data not shown).

3.3.2 TCF4 mutations affect protein:protein interactions

Dimerising proteins interact with close proximity and should therefore emit a FRET signal when labeled with appropriate fluorophores and excited (Figure 3.3). The HTRF technique was therefore used to quantitatively assay dimerisation in pairs of interacting bHLH proteins (Figure 3.4). A clear signal was obtained from co-transfecting TCF4 with known binding partners therefore the effects of PTHS-mutations on these interactions was measured using HTRF

The HTRF technique was used to demonstrate that mutants can potentially impair protein:protein interactions, either as a homodimer (with wild type TCF4) or through heterodimerisation with proneural or inhibitory bHLH transcription factors (Figure 3.6 and 3.7). FRET was detectable for each pair of interacting wild type proteins (Figure 3.4) however mutations in the bHLH domain were associated with a significant reduction in FRET intensity suggesting a strong effect on protein:protein interactions (Figure 3.6 and 3.7).

The effect of each mutant could be ranked according to its impact on FRET signal, as the HTRF technique provides a quantitative measure of association between proteins. Within the series of mutations modeled, the A614V mutation, located in the HLH domain, had the most deleterious effect on protein:protein interactions. These observations are consistent with the HLH domain being a major determinant of homo- or hetero- dimerisation between bHLH family members (Voronova and Baltimore, 1990). These results differed from mutations N-terminal to the bHLH domain that had a marginal impact on protein association but remained significantly different to wild type in certain heterodimer configurations (Figure 3.6 and 3.7).

In comparison, Sepp *et al*. found that most *TCF4* mutants did not have appreciable effects on homo- and heterodimerisation (Sepp et al., 2012a). These observations may be attributable to the nature of the technique used to qualify the interactions. The HTRF technique used herein provides a quantitative measure of protein:protein association by an indirect measurement of FRET between two fluorophores (Wang et al., 2006a). The data from these HTRF experiments demonstrated that TCF4 mutations impair but do not abolish protein:protein interactions in both homo- or heterodimer configurations. Sepp *et al.* used a nuclear redirection assay to infer homo- and heterodimerisation of wild type and mutant TCF4 in live cells, which were then fixed and imaged. The assay is based on the "piggy-back mechanism" which is used by certain bHLH proteins without a nuclear localisation signal (NLS) to relocate to the nucleus. In this assay, mutants lacking their N-terminal region (including the NLS) were fused to EGFP and co-transfected with a wild type bHLH protein containing a NLS (ASCL1 or NEUROD2) to establish whether mutants could be trafficked to the nucleus by associating (or "piggy-backing") with a wild type protein. Using this assay, Sepp and colleagues concluded that all PTHS-associated mutants could be redirected to the nucleus, implying that they were all capable of homo- and hetero- dimerisation. The data in this chapter is in broad agreement with this conclusion since most TCF4 mutations (homo- or hetero- dimer) produced a FRET signal (with the exception of ID2 that was not tested by Sepp). However, the quantitative nature of HTRF meant differences were still identified between TCF4 mutants which would be much more difficult to assess using qualitative nuclear redirection assays.

3.3.3 PTHS mutations affect transcriptional activity in a context-dependent manner

To date, three previously published studies have examined how PTHS-associated TCF4 mutants affect transcriptional activity. These studies used multimerised E-box-containing reporter constructs from the pTα enhancer of the Herpes simplex thymidine kinase promoter (Zweier et al., 2007), the Delta1 enhancer (de Pontual et al., 2009) and the μE5 immunoglobulin enhancer (Sepp et al., 2012a). These studies found that most TCF4 frameshift and truncation mutations, as well as missense mutations in the basic domain, attenuate or abolish transcriptional activity when co-expressed with ASCL1 (Zweier et al., 2007; de Pontual et al., 2009; Sepp et al., 2012a). Similarly, the data presented here demonstrates that *TCF4* missense mutations attenuate gene expression of a reporter construct driven by the putative *NRXN1β* and *CNTNAP2* promoters when co-expressed with an

accessory bHLH transcription factor (Figures 3.11 and 3.12). However, it must be highlighted that the deficits in transactivation reported here are more subtle then in previously published data, which is an aspect discussed below (de Pontual et al., 2009; Sepp et al., 2012a).

The bHLH domain is an essential determinant of DNA-binding specificity and dimerisation (Davis et al., 1990; Voronova and Baltimore, 1990). In agreement with these previous studies, the data indicates that missense mutations (R578P and R580W) in the basic region of TCF4 attenuate transcriptional activity and are the most damaging mutants modeled (Zweier et al., 2007; Sepp et al., 2012a). This effect is reproducible with the majority of conditions tested and on both the *NRXN1β* and *CNTNAP2* promoters. The remaining mutants (G358V, D535G, A614V) had a weaker effect on luciferase expression, although each mutant was shown to impair transcription in a particular context. These general conclusions are broadly similar to those of Sepp and colleagues, however some differences are apparent which may be important in understanding TCF4 function.

The most salient difference is that promoter context is particularly important to the activating properties of bHLH dimers. In the study published by Sepp and colleagues, the effect of PTHS mutants on transcriptional activation is measured using a synthetic construct composed of 12 E-boxes (CACCTG). Using ASCL1 as a co-activator, Sepp and colleague found that most mutations in the basic, DNA-binding region of the bHLH domain abolished transcription whereas the other mutations have little impact on this function. The data in this chapter shows that the DNA-binding domain mutants show a significant reduction in activity at the *NRXN1β* promoter when co-expressed with ASCL1, however none of the mutants completely abolished transcriptional activity (Figure 3.11C). Residual transcriptional activity in the DNA-binding mutants is also observed in previously published studies that use slightly different synthetic promoters (Zweier et al., 2007; de Pontual et al., 2009). The absolute effect of mutations in the DNA-binding domain on transcriptional activity is therefore highly dependent on promoter sequence and structure. Interestingly, at the *CNTNAP2* promoter all PTHS-associated TCF4 mutants were found to significantly reduce promoter transactivation when co-expressed with ASCL1 (Figure 3.12C), including mutants G358V, D535G and A614V that were unaffected in Sepp's experiments. This is one of the only contexts where mutations outside the basic domain have a substantial effect on transcriptional activity and this may be related to the presence of an ASCL1 favoured E-box adjacent to the

transcriptional start site of the *CNTNAP2* gene (Figure 3.8). The TCF4 mutations may be expected to have similar damaging effects in physiological conditions as each mutation is assumed to be the cause of PTHS in patients. Thus it may be concluded that synthetic promoters are useful to interpret the general properties of a mutation, but they cannot be used to infer deficits *in vivo* which are likely to be much more subtle due to the presence of multiple interacting proteins and complex promoter structures.

The transactivation experiments also demonstrate that some mutations can elicit opposing or variable effects depending on the co-expressed binding partner. For example, R578P and R580W increased luciferase expression from the *CNTNAP2* promoter when co-expressed with NEUROD1 (Figure 3.11D) but had little effect on the *NRXN1β* promoter. This effect may reflect the highly context-specific functions of bHLH factors (Castro et al., 2006; Powell and Jarman, 2008). The gene expression analysis performed on TCF4-knockdown cells in chapters 4 and 6 would suggest that TCF4 can activate or repress transcription and thus may interact with co-activators and co-regulators as has already been proposed (Zhang et al., 2004; Bayly et al., 2006). Consequently, TCF4 mutants may differentially affect the ability of homo and hetero- dimers to recruit regulatory complexes to promoters, depending on the surrounding DNA sequence. In this view, the R578P/NEUROD1 and R580W/NEUROD1 heterodimers may cause aberrant activation of the *CNTNAP2* promoter by recruiting additional co-activating complexes that do not interact on the *NRXN1β* promoter (Figures 3.11D and 3.12D). The R578P/NEUROD1 and R580W/NEUROD1 heterodimers may also sequester repressive complexes specifically affecting the *CNTNAP2* promoter leading to an increased activation (Figures 3.11D and 3.12D). The differential recruitment of co-regulators may also explain the opposing effects observed between the R578P/ID2 and R580W/ID2 heterodimers on the *NRXN1β* and *CNTNAP2* promoters (Figures 3.11E and 3.12E). This data underscores the potentially complex regulatory mechanisms governing gene regulation by bHLH transcription factors and their context dependence. The exact regulatory mechanisms operating on these promoters may be interesting to investigate further if the *NRXN1β* and *CNTNAP2* promoter regions used in this study can be validated as physiological TCF4 binding sites.

Although it is clear that the mutants R578P, R580W and A614V affect the well characterised bHLH domain, it is less certain how the other mutants alter TCF4 function. In addition to the

bHLH domain, TCF4 contains at least two identifiable activation domains as well as two repressor domains (CE repressor and Rep domain) defined by their function in transcriptional assays (Section 1.5.2). The G358V mutation resides in AD2 whereas D535G is located in the Rep domain (Figure 3.1). As a result of their location, these mutations may interfere with the regulation of TCF4 though the AD and Rep domains of the protein. Consistent with the context dependence of these mutants, the Rep domain has already been proposed to behave in a manner that is dependent upon the composition of the dimer (Markus et al., 2002a). Interestingly, the D535G appears to have an augmented activity as a homodimer with the wild type protein when compared to TCF4 and the other mutants in luciferase assays. This relatively weak increase in activity was found with both the *NRXN1β* and *CNTNAP2* promoters even though the D535G reduced the propensity of the mutant to form homodimers in the HTRF assay (Figure 3.6). Similarly, Sepp *et al*., also found that that D535G mutant was more active than wild type TCF4 in luciferase assays using a synthetic promoter construct. Furthermore, using electrophoretic mobility shift analysis (EMSA) *Sepp et al*., also found that D535G had a preference for homodimerisation over heterodimer formation with ASCL1 (Sepp et al., 2012a). Thus, the D535G mutation may augment transcriptional activity of the mutant protein at certain promoters possibly through de-repression (Figure 3.1). These data suggest that PTHS-mutations can dysregulate gene expression by upregulating and downregulating target genes.

In summary, the results suggest that mutations R578P and R580W and other DNA binding mutants have the most pervasive effects on TCF4 function and are the most damaging *in vitro*. This may be the reason why these are the most common missense mutations in PTHS. The A614V mutation seems to specifically disrupt homo- and hetero- dimerisation over other aspects of function whilst the impact of the G358V and D535G mutations are more subtle and highly dependent on particular contexts. The data presented in this chapter define a molecular phenotype for five PTHS-associated missense mutations. In addition, the transactivation experiments on the *CNTNAP2* and *NRXN1β* promoters raise the intriguing possibility these genes may be regulated by TCF4, defining a regulatory network in PTHS. These data provide an important insight into the function of TCF4 and construct a conceptual framework to understand the role of TCF4 in PTHS and related neurodevelopmental disorders.

Chapter 4

Knockdown of *TCF4* and Gene Expression Profiling

4.1 Introduction

TCF4 is associated with an increased risk of schizophrenia – a common psychiatric disorder affecting approximately 1% of the population (Section 1.11). The largest genome-wide association studies to date have repeatedly implicated common DNA variation in the TCF4 gene to schizophrenia susceptibility. *TCF4* was initially associated to schizophrenia in large meta-analysis of GWAS with 12,945 schizophrenia cases 34,591 controls ($P = 4.1\times10^{-9}$; OR 1.23) (Stefansson et al., 2009). Independent studies have since found additional evidence that variants near *TCF4* confer risk to schizophrenia (Ripke et al., 2011; Steinberg et al., 2011; Aberg et al., 2013) (Figure 1.5). Although the functional implications of these common DNA variants are often unclear, many SNPs are found in regulatory DNA elements that can affect gene expression (Kavanagh et al., 2013). Consistent with this hypothesis *TCF4* expression levels in the frontal cortex and the thalamus of healthy patients were significantly associated to *cis*-acting SNPs proximal to the *TCF4* gene (Kim et al., 2012). Variation in SNPs at regulatory elements around the *TCF4* gene may therefore differentially regulate *TCF4* expression and lead to subtle changes in critical neurodevelopmental pathways. Support for this hypothesis is already emerging as increased *TCF4* expression has been identified in the peripheral blood of psychotic patients (Wirgenes et al., 2012). It is therefore plausible that altered expression of *TCF4* leads to an unbalanced transcriptional program that contributes to the etiology of schizophrenia and PTHS.

To date, the genome-wide effects of TCF4 on neuronal transcription are unknown. TCF4 function and its transcriptional program have been comprehensively investigated in the context of the immune system but comparable understanding of its neural function is lacking (de Pooter and Kee, 2010; Reizis, 2010). In order to improve our understanding of TCF4 function and the cellular networks it regulates, a neuronal cell line expressing high levels of endogenous TCF4 was used to investigate the TCF4-regulated transcriptional program. RNA

interference (RNAi) was used to knockdown *TCF4* transcripts in SH-SY5Y cells and the effect on the neuronal transcriptome was measured using microarrays combined with pathway analysis and manual curation of the data. Together, these experiments demonstrate that knockdown of *TCF4* has a direct impact on global transcription and is associated with robust gene expression changes in multiple, convergent signaling pathways. This data offers unique insight into gene expression networks potentially dysregulated in schizophrenia and PTHS and establishes a framework to understand *TCF4* function the brain.

The work described in this chapter has been published in *PLoS One*, August 2013, vol. 8, issue 8, e73169 (Forrest et al., 2013).

4.2 Results

4.2.1 Knockdown of TCF4 in SH-SY5Y cells

The silencing of particular mRNA targets can be achieved through RNAi, a mechanism by which double stranded RNA (dsRNA) triggers the sequence-specific suppression of mRNA through the RNA-induced silencing complex (RISC) (Fire et al., 1998). This technique is now widely used for studying gene function and can be combined with microarrays to study genome wide transcriptional changes.

The *TCF4* gene contains at least 20 coding exons and has complex splicing architecture resulting in multiple transcripts with alternative 5' exons (Sepp et al., 2011). In order to knockdown all published *TCF4* transcripts, bioinformatics tools were used to design siRNAs matching the sequence of core exons shared by all transcripts. High-scoring siRNA duplexes complementary to exon 12 (TCF4 KD1) and exon 19 (KD2) of the *TCF4* sequence were selected for their predicted efficacy. To control for non-specific effects of siRNA treatment, a siRNA duplex against *GAPDH* transcripts was also designed and used along side a mock transfection control in the experiments. In each case, blunt-ended RNA duplexes 27 nucleotides in length (27-mers) were preferred to conventional 21-mer siRNAs as they can be up to 100-fold more potent and can be used at lower concentrations (Kim et al., 2005). Knockdown of *GAPDH* was predicted to be a useful control as it is an enzyme that participates in central metabolism and is not predicted to cause acute gene expression changes if depleted (Appendix III). In addition, the GAPDH siRNA can control for gene expression changes induced by the interferon and the RNAi responses that cells activate non-specifically in the presence of dsRNA. The mock transfected and GAPDH siRNA transfected control groups can therefore be used for background correction in these experiments.

After designing the siRNAs, each duplex was transfected individually in SH-SY5Y cells to test for their sequence-specific efficacy. The knockdown protocol performed in biological triplicates and was scheduled over 72h, to ensure a strong, reproducible depletion. This protocol validated the use of the SH-SY5Y model as each siRNA was efficiently transfected and achieved a robust knockdown of the target gene (Figure 4.1A). Knockdown efficiency was evaluated by quantitative PCR (qPCR) for both *GAPDH* and *TCF4*. The *TCF4*-silencing duplexes reduced transcript levels to 20% of mock (TCF4 KD1) and 38% of mock (KD2)

(Figure 4.1A). Knockdown of *GAPDH* reduced transcript levels reduced to 15% of mock-transfected cells (data not shown).

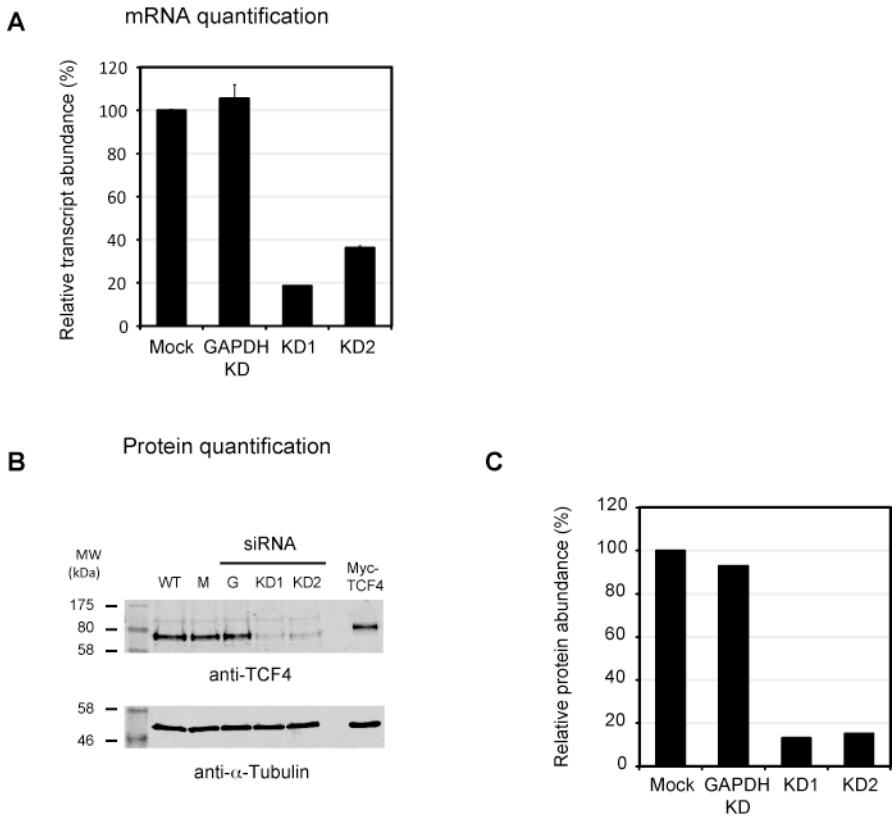


Figure 4.1 siRNA-mediated knockdown of *TCF4* in SH-SY5Y cells. SH-SY5Y cells were transfected with siRNA oligonucleotides targeting TCF4 (KD1 and KD2) or GAPDH. After 72h, RNA and protein were extracted to assess the knockdown efficiency. (A) Primers complimentary to a constitutive exon present in all TCF4 isoforms (exon 13) were used to measure overall transcript abundance by qPCR in three biological replicates. *TCF4* knockdown efficiency was similar with both KD1 and KD2 whereas the control GAPDH KD did not affect *TCF4* transcript levels. (B) Western blots of protein lysates prepared from siRNA-treated SH-SY5Y cells demonstrated that KD1 and KD2 reduced TCF4 levels whilst wild type (WT), GAPDH knockdown (G), and mock (M) treatment cells had no apparent effect. α-tubulin was used as a loading control and for normalization. (C) LI-COR quantitation of TCF4 protein levels in siRNA-treated cells. The levels of TCF4 in each experiment were quantified and normalised to α-tubulin. In agreement with the qPCR results, siRNA treatment reduced TCF4 levels to approximately 20% of mock treated cells.

Importantly, siRNAs targeting *GAPDH* and mock transfection had no significant effect on the level of the *TCF4* transcripts (Figure 4.1). In addition to qPCR, quantification of the knockdown treatment was also verified by semi-quantitative western blotting. A clear

reduction in TCF4 protein was visible for the TCF4 knockdown samples KD1 and KD2 compared to controls (Figure 4.1B). Knockdown of *TCF4* with KD1 and KD2 resulted in a greater than 80% reduction in proteins levels with both duplexes (Figure 4.1C). Although similarly reduced, KD2 was less effective than KD1 at knocking down the TCF4 protein reflecting the qPCR data. As a final assessment, the SH-SY5Y cells were imaged by confocal microscopy to observe any overt treatment effects on cellular morphology. The TCF4-depleted cells had reduced expression of TCF4 in their nuclei compared to controls although no obvious morphological differences were observed (Figure 4.2)

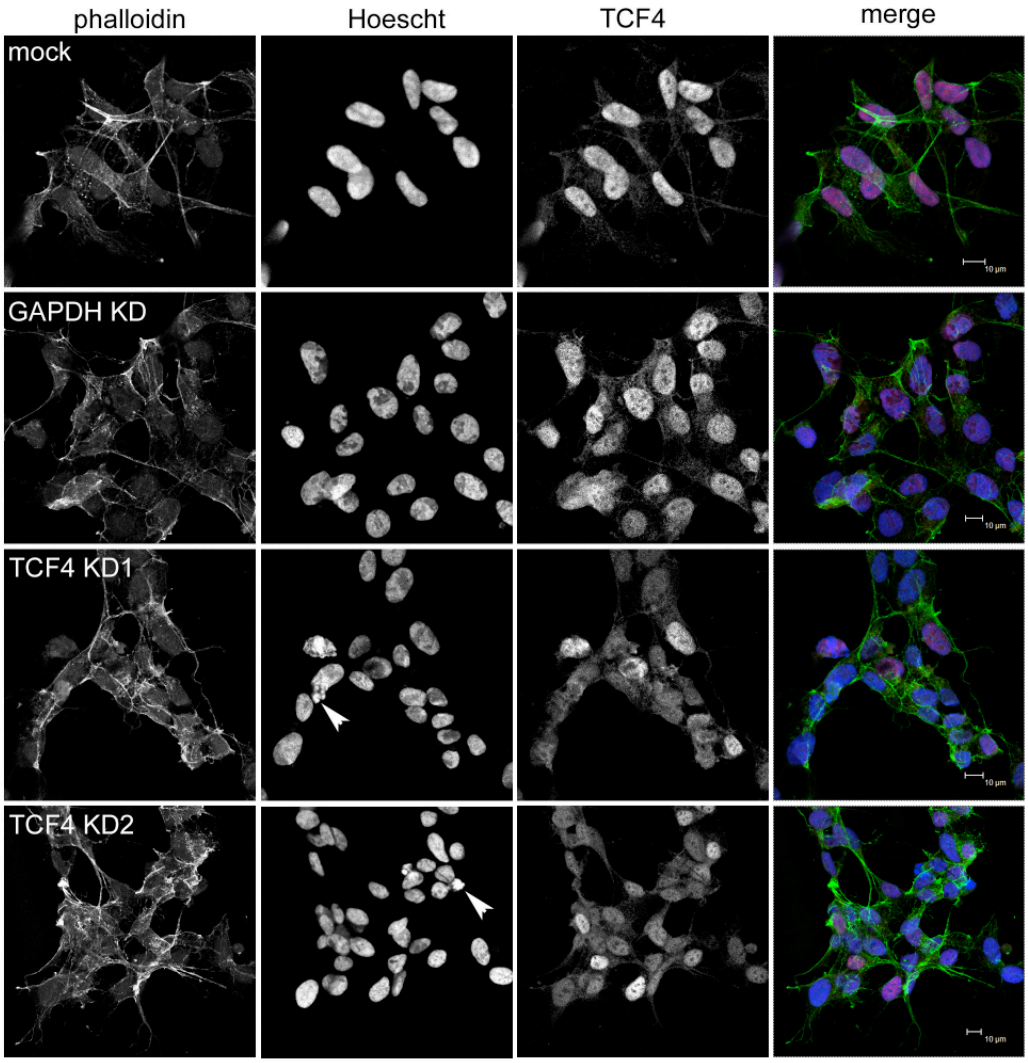


Figure 4.2 Cellular imaging of siRNA-treated SH-SY5Y cells. SH-SY5Y cells were treated with siRNAs for 72h as described above. Fixed cells were stained with Alexa Fluor 488 phalloidin (F-actin, green), Hoescht-33342 (nuclei, blue) and with an anti-TCF4 polyclonal antibody (red). Although there is a marked reduction in TCF4 immunoreactivity in TCF4 KD cells, no apparent morphological differences are observed between treatment groups after 3 days knockdown. A few condensed pyknotic nuclei, indicative of apoptosis were observed after TCF4 KD treatment (arrow heads).

4.2.2 Data processing and quality control of microarray experiments

After successful validation of the gene-specific knockdown, the quality of the RNA from each sample was measured using an Agilent Bioanalyzer. Each RNA sample had a RNA integrity number (RIN) of 10, the maximum score for this instrument, indicating the extracted RNA was of high quality. To ensure that any of the reported gene expression changes were reproducible, three of the previously validated biological replicates from each group were used for microarray analysis, a total of 12 samples. Each of the samples were converted to cDNA, amplified, labeled and hybridised to a Toray microarray for genome-wide transcript analysis (Section 2.9.9).

The processed microarray data received from CBS consisted of an Excel spreadsheet with background-corrected signal intensities for each gene annotated on the Toray microarray platform. The Toray microarray has a unique probe for each gene that is designed to capture the majority of transcripts described in public databases (approximately 25,000 probes). The annotated gene information and signal intensity data for each of the 12 samples (3 mock, 3 GAPDH, 3 TCF4 KD1, 3 TCF4 KD2) was imported manually into Partek Genomics Suite 6.6, log transformed to fit normal distribution and quantile normalised so that the distribution ranges of each sample was similar. As there were relatively few replicates in each condition (n=3), probes with more than 1 missing values across all groups were ignored.
The success of the experimental treatments was measured using principle component analysis (Figure 4.3.2). This graphical display summarises the microarray data for each sample onto three principle components and demonstrates how the experimental samples relate to one another. The quality of the experiment can be assessed by the clustering biological replicates within a treatment group and the segregation of samples between treatments. It is interesting to note that although each treatment group clusters in a distinct area of the graph (indicating distinct treatment effects), the replicates in the KD2 treatment group (light green) are much closer to the control samples compared to the KD1 (dark green) (Figure 4.3.2).

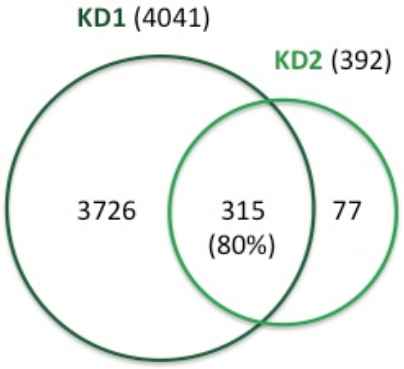


Figure 4.3.1. Comparison of gene expression signatures between TCF4 knockdown groups. The microarray data for TCF4 KD1 and TCF4 KD2 was used to generate two lists of transcripts that were differentially expressed compared to mock treated cells. The lists were generated by performing a one-way ANOVA between the TCF4 knockdown group (KD1 or KD2) and the mock treated group. The lists were compared to determine the number of shared gene expression changes (illustrated in the overlap). The total and shared number of differentially expressed transcripts are indicated. The size of each circle is representative of the number of gene expression changes but is not to scale.

The similarity between the TCF4 KD1 and TCF4 KD2 treatments were subsequently compared in terms of their effect on gene expression compared to mock treated cells. A one-way ANOVA comparing the signal intensities from the TCF4 KD-treated groups to the mock treated group was performed to establish two list of differentially expressed; one for TCF4 KD1 and one for TCF4 KD2. A false discovery rate (FDR) 0.01 was applied to each list to remain in agreement with downstream analysis (see section 4.2.3). After FDR-correction, comparison of the differentially expressed genes in TCF4 KD1 and TCF4 KD2 revealed an overlap of 69%, indicating similarity in the effects of each treatment on gene expression (Figure 4.3.1).

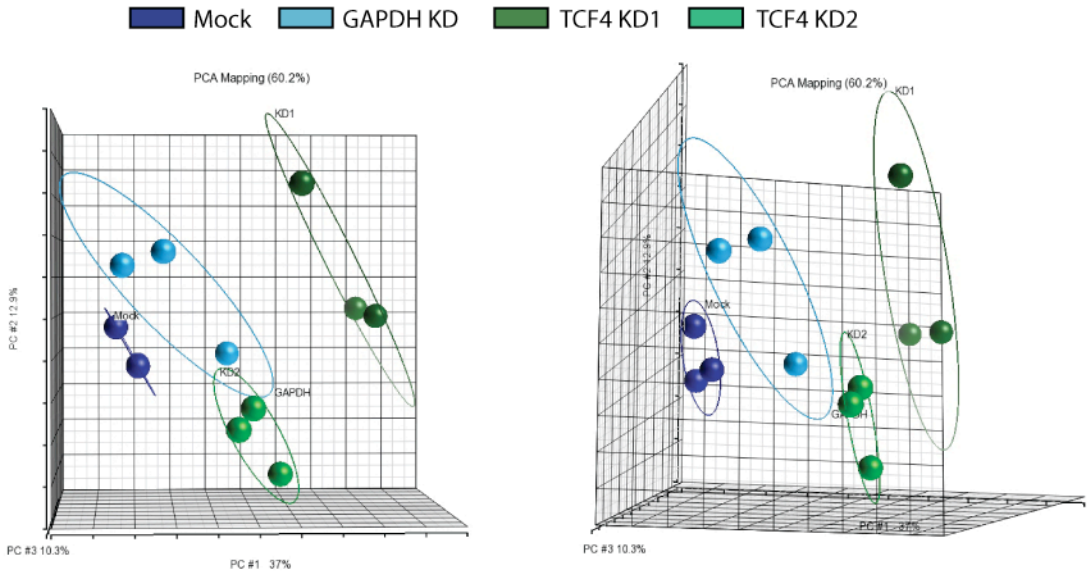


Figure 4.3.2 Principle component analysis of microarray samples. The genome-wide expression data from each sample is reduced to three principle components that capture 60.2% of the variation in the data. The principle components (PC) are plotted against 3 axes: PC1; X-axis, PC2; Y-axis, PC3; Z-axis. The percentage variance captured by each PC is displayed next to each axis label (PC1; 37%, PC2; 12.9%, PC3; 10.3%). Two different viewpoints of the graph are presented to fully appreciate the distribution of data points in dimensional space. For clarity, treatment groups have been colour-coded (see key above graph: Mock; dark blue, GAPDH KD; light blue, TCF4 KD1; dark green, TCF4 KD2; light green)

4.2.3 Overview of differentially expressed genes in TCF4 knockdown cells

In order to identify genes differentially expressed between the TCF4 KD groups and the control groups a one-way ANOVA was performed on the pooled control (mock and GAPDH KD; 6 samples) and pooled TCF4 (TCF4 KD1 and KD2; 6 samples) knockdown groups. To correct for multiple testing, a false discovery rate (FDR) based approach was applied which is appropriate for microarray analysis (Reiner et al., 2003). Decreasing the threshold of FDR correction decreases the number of significant genes in the ANOVA analysis. In our analysis, an FDR correction of 0.05 generated a list of 5376 differentially expressed genes and this list was reduced to 1204 genes when the correction was reduced to 0.01. To select the most robust gene expression changes and reduce the number of genes in the list, an FDR correction of 0.01 was used for all downstream enrichment analysis, as a moderately sized gene lists (between 500 and 2000 genes) are recommended for an informative enrichment analysis (Huang da et al., 2009).

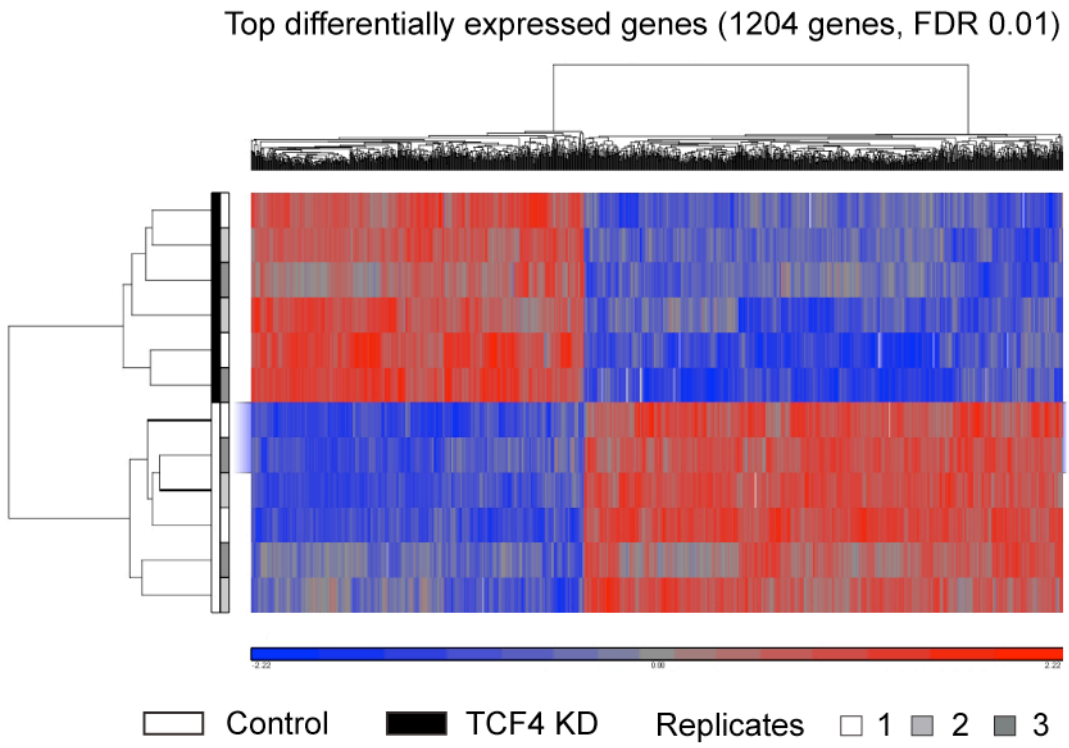


Figure 4.4 Hierarchical clustering of the top differentially expressed genes. The top 1204 differentially expressed genes (FDR 0.01) cluster by treatment group as evident from the dendrogram to left of the heatmap. Samples from the control treatments (mock and GAPDH KD) and TCF4 KD treatment (KD1 and KD2) are shown to cluster into defined groups (white and black respectively). The top section of the heat map displays differential expression relative to controls: up-regulated genes are shown in red while down-regulated genes are coloured blue.

Hierarchical clustering was used to visualise the top gene expression changes in the FDR 0.01 list. The top gene expression changes clustered with the experimental treatment indicating control and TCF4 KD groups had distinct gene expression signatures (Figure 4.4). This clustering analysis also displays that there is a higher portion of downregulated (710) compared to upregulated genes (494) (Figure 4.4). The top 40 up- and downregulated genes ranked by fold change are illustrated in Figure 4.5A. Initial examination of the most robust gene expression changes in TCF4-knockdown cells (Figure 4.5A) appeared to suggest a role for TCF4 in apoptosis or inflammasome function (up-regulation of *CASP1* and *CASP4*), cell signaling (down-regulation of *IGF2*, *BMP7* and *LEFTY1*) and neurodevelopment (down-regulation of *NEUROG2*, *ASCL1* and *MEF2C*). Furthermore, several of the major gene expression changes in TCF4-knockdown cells involved transcription factors including ASCL1 and NEUROG2 that interact directly with TCF4 at E-boxes (Zweier et al., 2007; Forrest et al., 2012; Li et al., 2012b; Sepp et al., 2012b). Finally, a number of imprinted

genes, *IGF2*, *H19* and *CDKN1C* were prominent amongst the most significantly downregulated genes in TCF4-knockdown cells (Figure 4.5A).

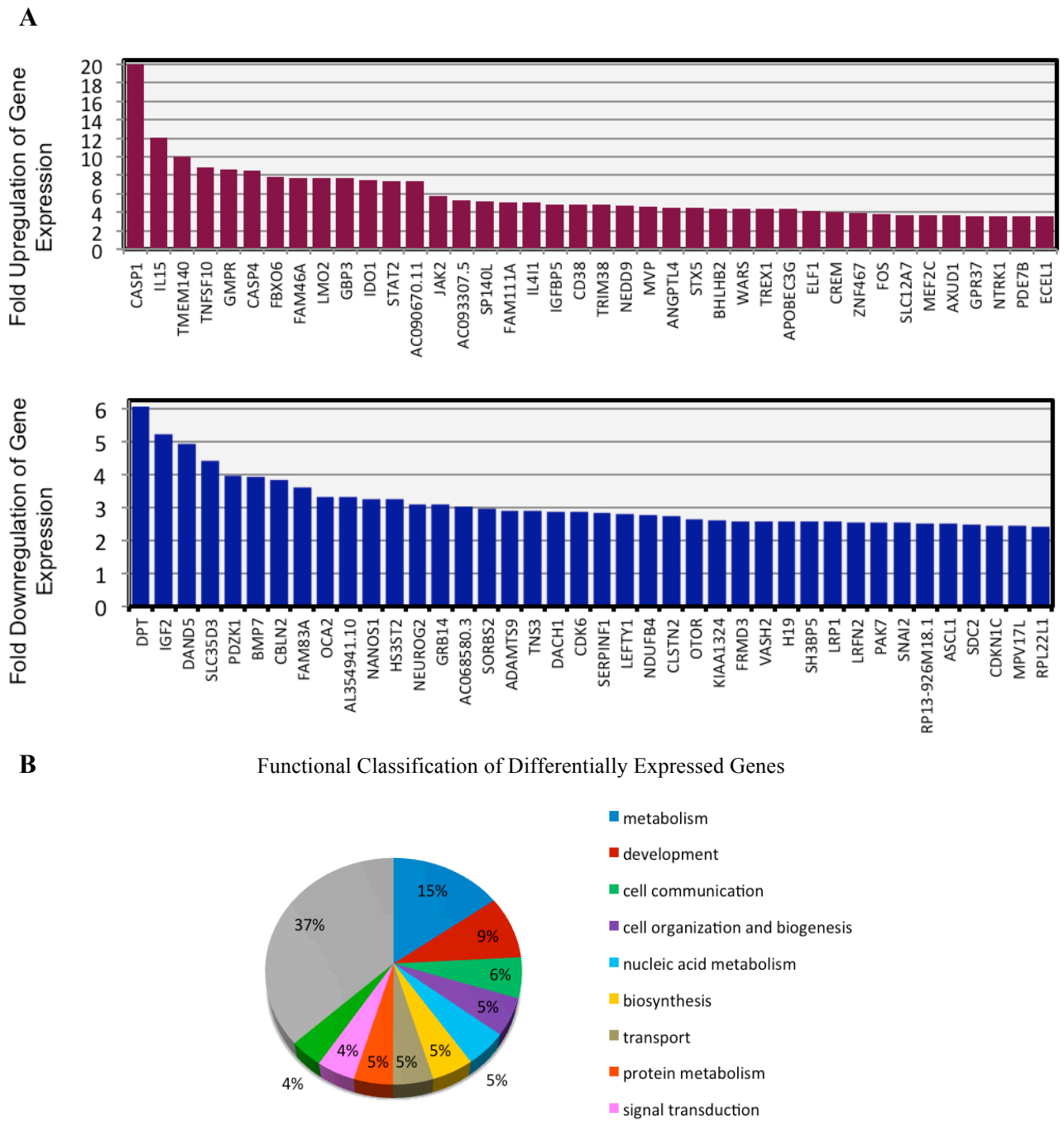


Figure 4.5 Overview of differentially expressed genes after *TCF4* knockdown
(A) Top 40 upregulated (upper panel) and downregulated (lower panel) genes after FDR correction (0.01) ranked by fold change. (B) Functional characterization of all differentially expressed genes (upregulated and downregulated) (FDR 0.01) using GOSLIM annotations. All annotation categories representing less than 4% of the gene list were grouped and labeled as "other".

To confirm the gene expression changes detected on the microarray, qPCR was used to independently validate the results. 5 up- (*FAS, NTRK1, CASP8, NOTCH1, CASP1*) and 5 down- regulated genes (*IGF2, CDKN1C, NEUROG2, BMP7, CDK6*) were selected that had a fold change above or below 1.5. The genes were also designated on the basis of their known importance to developmental processes and their recurrent appearance in the downstream enrichment analyses (Tables 4.2 and 4.3). Transcript abundance was measured from the same RNA samples used on the microarray, allowing a direct comparison of the inferred changes. In addition, TaqMan probes were selected to match the exons targeted on the microarray. The qPCR analysis confirmed that each of the 10 selected genes were differentially expressed and each had similar a fold change to that detected on the microarray (Table 4.1).

Table 4.1 qPCR validation of differentially expressed genes. Ten differentially expressed genes belonging to various functional categories were chosen for qPCR validation. All the genes selected for validation had similar fold changes by qPCR to that seen on the microarray. In both cases, the presented *P* values represent the statistical significance (Student's *t*-test) between the pooled control (mock treated, GAPDH KD) and TCF4 KD (KD1, KD2) groups.

Gene	Microarray		qPCR	
	Fold change	*P* value	Fold change	*P* value
FAS	1.9	2.88E-05	2.9	3.58E-05
NTRK1	3.8	1.53E-05	4.1	3.58E-05
CASP8	3.0	1.38E-05	10.4	2.40E-03
NOTCH1	2.3	1.02E-04	2.9	6.03E-05
CASP1	15.7	6.13E-05	9.8	8.06E-03
IGF2	-5.2	1.52E-06	-4.2	2.51E-05
CDKN1C	-2.4	2.36E-06	-2.3	1.02E-03
NEUROG2	-3.0	1.50E-05	-3.6	8.30E-07
BMP7	-3.7	1.34E-04	-2.4	5.46E-05
CDK6	-2.9	2.87E-07	-2.6	3.71E-06

4.2.4 Enrichment analysis of differentially expressed genes in TCF4-depleted cells

Initial functional annotations associated with the 1204 differentially expressed genes in TCF4-knockdown cells were found to span a range of biological functions that cluster around the broad themes of metabolism, development and cell signaling (Figure 4.5B). To ascertain whether any biological functions were enriched in the high confidence gene list, all genes with Entrez gene IDs were selected for enrichment analysis to avoid the redundancy associated to using gene names. Of the 1204 genes that passed FDR correction of 0.01, 1031 had corresponding Entrez IDs (425 upregulated, 606 downregulated). The 1031 genes were subsequently tested for GO annotation enrichment using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang da et al., 2009). Functional annotations from up- and downregulated genes were compared separately to a background list of genes consisting of all expressed genes detected on the microarray. This analysis identified several biological processes that were significantly enriched (FDR 0.05) in TCF4 knockdown cells. These terms included non-coding RNA metabolic processes, apoptosis and regulation of NF-κB signaling (Table 4.2). Importantly, up- and down- regulated genes seemed to cluster by distinct functional annotations (Table 4.2). For example, regulation of the NF-κB and apoptotic signaling cascades and were significantly enriched in the upregulated genes whereas annotations relating to non-coding RNA metabolism, ribosome biogenesis and protein folding were downregulated as a consequence of TCF4 knockdown. Although not all the GO annotations listed in the downregulated gene category passed multiple test correction, they were nominally significant by DAVID's EASE score ($P < 0.05$) and functionally linked. In this context it is important to note that DAVID recommends using p-values simply as a ranking tool without the need for statistical cut offs to be set, as they may be arbitrary (Huang da et al., 2009).

Table 4.2 GO term enrichment analysis using DAVID. The Entrez Gene IDs of upregulated and downregulated genes were analyzed separately for biological process enrichment using DAVID. Enriched GO terms are ranked in order of decreasing significance. The genes in each GO category are listed as wells as the corrected (FDR) and uncorrected *P* values. For clarity redundant terms were removed. *Pass FDR correction of 0.05.

	GO annotation	Genes	*P value*	FDR
Upregulated genes (423 DAVID IDs)				
1	GO:0043122~regulation of I-kappaB kinase/NF-kappaB cascade	*CD40 TFG, TNFSF10, CASP1, TBK1, PLK2, TRIM38, NOD1, RHOC, HTR2B, WLS, CARD8, SHISA5, TRADD, ZDHHC17*	4.87E-07	0.001*
2	GO:0042981~regulation of apoptosis	*MEF2C, LOC100289713, NGFRAP1, ARHGEF6, NOTCH1, SH3GLB1, TNFSF13B, B4GALT1, TNFSF10, IDO1, BNIP3, APH1A, NOD1, ADAM17, CARD8, NGFR, SOCS3, KALRN, TRADD, SOX9, MSX2, AEN, CADM1, BARD1, CASP4, CREB1, CD38, CASP1, AKT1, CASP3, MCL1, JAK2, DPF2, TXNIP, FAS, TIA1, ANGPTL4, EYA1, YWHAZ, NTRK1, SIRT1, BCL2L13*	1.10E-05	0.019*
3	GO:0009967~positive regulation of signal transduction	*CD40, BMPR1A, TFG, HIF1A, FKBP8, TNFSF10, CASP1, JAK2, TBK1, PLK2, TRIM38, NOD1, HTR2B, ADAM17, JAG1, RHOC, WLS, RICTOR, SMAD4, SHISA5, TRADD, ZDHHC17*	1.18E-05	0.021*
4	GO:0043065~positive regulation of apoptosis	*LOC100289713, NGFRAP1, ARHGEF6, NOTCH1, SH3GLB1, B4GALT1, TNFSF10, BNIP3, APH1A, NOD1, NGFR, KALRN, TRADD, MSX2, AEN, CADM1, BARD1, CASP4, CD38, CASP1, AKT1, CASP3, JAK2, DPF2, FAS, TXNIP, TIA1, BCL2L13*	1.28E-05	0.022*
5	GO:0006915~apoptosis	*MEF2C, LOC100289713, NGFRAP1, ARHGEF6, SH3GLB1, FKBP8, TNFSF10, BNIP3, CSRNP1, APH1A, NOD1, CSRNP2, CARD8, NGFR, KALRN, TRADD, AEN, CADM1, CASP4, RNF34, DAPL1, CASP1, AKT1, CASP3, JAK2, MCL1, DPF2, FAS, ELMO1, TIA1, SHISA5, GULP1, SIRT1, BCL2L13*	1.29E-05	0.022*
6	GO:0043067~regulation of programmed cell death	*MEF2C, LOC100289713, NGFRAP1, ARHGEF6, NOTCH1, SH3GLB1, TNFSF13B, B4GALT1, TNFSF10, IDO1, BNIP3, APH1A, NOD1, ADAM17, CARD8, NGFR, SOCS3, KALRN, TRADD, SOX9, MSX2, AEN, CADM1, BARD1, CASP4, CREB1, CD38, CASP1, AKT1, CASP3, MCL1, JAK2, DPF2, TXNIP, FAS, TIA1, ANGPTL4, EYA1, YWHAZ, NTRK1, SIRT1, BCL2L13*	1.29E-05	0.023*
7	GO:0043068~positive regulation of programmed cell death	*LOC100289713, NGFRAP1, ARHGEF6, NOTCH1, SH3GLB1, B4GALT1, TNFSF10, BNIP3, APH1A, NOD1, NGFR, KALRN, TRADD, MSX2, AEN, CADM1, BARD1, CASP4, CD38, CASP1, AKT1, CASP3, JAK2, DPF2, FAS, TXNIP, TIA1, BCL2L13*	1.39E-05	0.024*

Continued…

Downregulated genes (598 DAVID IDs)				
1	GO:0034660~ncRNA metabolic process	*CARS2, FARSB, POP7, TSEN2, WDR12, EXOSC1, DUS3L, TARSL2, ADAT2, RPL35A, PDCD11, PIWIL1, IMP4, QTRTD1, TRMT10C, PUS3, EXOSC7, FTSJ1, MKI67IP, RPL7, DIMT1*	2.42E-05	0.041*
2	GO:0022613~ribonucleoprotein complex biogenesis	*WDR12, GEMIN6, RRS1, EXOSC1, NCBP1, SURF6, RPL35A, TSR1, IMP4, PDCD11, GEMIN5, EXOSC7, FTSJ1, NUFIP1, MRTO4, RPL7, BYSL, DIMT1*	3.72E-05	0.064
3	GO:0042254~ribosome biogenesis	*WDR12, RRS1, EXOSC1, SURF6, RPL35A, TSR1, IMP4, PDCD11, FTSJ1, EXOSC7, MRTO4, RPL7, BYSL, DIMT1*	8.01E-05	0.14
4	GO:0034470~ncRNA processing	*WDR12, TSEN2, POP7, EXOSC1, DUS3L, ADAT2, RPL35A, IMP4, PDCD11, QTRTD1, TRMT10C, PUS3, EXOSC7, FTSJ1, RPL7, DIMT1, SMAD2*	1.94E-04	0.33
5	GO:0006457~protein folding	*BAG2, DNAJC12, SEC63, RUVBL2, FKBP7, GRPEL1, APCS, SACS, PFDN6, HSPBP1, PPIF, PPID, CCT6A, URI1, PPIH*	7.06E-04	1.20
6	GO:0016072~rRNA metabolic process	*WDR12, EXOSC1, EXOSC7, MKI67IP, FTSJ1, RPL35A, PDCD11, RPL7, IMP4, DIMT1*	2.72E-03	4.55
7	GO:0007005~mitochondrion organization	*MRPL12, TIMM8A, TIMM13, PPIF, OXA1L, PTCD2, BID, GRPEL1, TIMM9, RNASEH1, FAXC, MYC*	4.45E-03	7.36

To determine whether the apoptotic cell death pathway was indeed activated in TCF4-depleted cells, cell viability and caspase activation after treatment with the different siRNAs was examined (Figure 4.6). TCF4-knockdown was associated with significantly reduced cell viability compared to the control groups ($P = 2.8x10^{-16}$). Reduced cell viability was also associated with increased caspase 3/7 activity ($P = 1.3x10^{-3}$) in TCF4-depleted cells (Figure 4.6A). Caspase-3 cleavage, which occurs in cells undergoing apoptosis, was also detected by western blotting in TCF4-depleted cells (Figure 4.6B). Interestingly, cells treated with TCF4 KD2 showed more evidence of apoptotic cell death and reduced viability compared to TCF4 KD1. In control experiments, acute staurosporine-treatment was associated with a robust increase in apoptotic cell death in untransfected SH-SY5Y cells. Although no gross differences in cellular morphology were evident between siRNA-treated cells and controls, condensed pyknotic nuclei were frequently observed in TCF4-depleted cells (Figure 4.2).

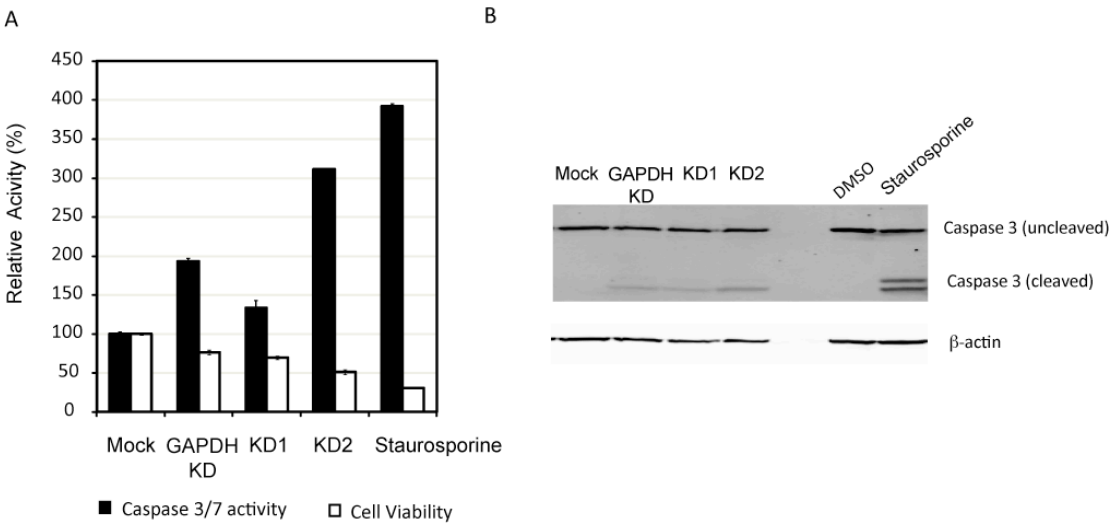


Figure 4.6 Knockdown of TCF4 induces apoptosis in SH-SY5Y cells. Cells were treated with siRNAs for 72h after which cell viability and caspase activity were measured (A). In addition to the siRNA treatment groups, untransfected SH-SY5Y cells were exposed to staurosporine (1μM in DMSO) and vehicle for 3h to induce apoptosis. Knockdown of TCF4 leads to a significant reduction in cell viability ($P = 2.8x10^{-16}$, white). Furthermore, TCF4-knockdown is also associated with an increase in caspase 3/7 activity compared to controls ($P = 1.3x10^{-3}$, black bars). Although GAPDH knockdown is associated with reduced cell viability and elevated caspase 3/7 activity compared to mock-treated cells, both assays showed statistical significant differences between the control groups (mock and GAPDH) compared to TCF4-knockdown (TCF4 KD1 and KD2) supporting the microarray data. As expected, staurosporine treatment also reduced cell viability and increased caspase 3/7 activity in untransfected cells. Western blot analysis of caspase 3 processing after 72h knockdown shows that caspase 3 cleavage products are detected in siRNA treated cells (B). β-actin was used as a loading control for all treatment groups.

4.2.5 Differential expression of genes in the TGF-β signaling and EMT pathways in TCF4-knockdown cells

To refine the analysis of downstream gene expression changes due to TCF4 knockdown, the MetaCore™ (GeneGo) analytical suite was utilised which provides a manually curated database of "process networks" which detail more specific biological processes than GO annotations. In this analysis, the full list of up- and downregulated genes with Entrez IDs (1031 genes) were interrogated to reveal the concerted function of the top differentially expressed genes. This analysis identified a number of cellular processes in TCF4-knockdown cells including TGF-β signaling, epithelial to mesenchymal transition (EMT), hedgehog signaling, apoptosis and neurogenesis (Table 4.3). These data corroborate the findings using DAVID, as process networks relating to apoptosis ("Death Domain receptors and caspases in apoptosis" and "Apoptosis stimulation by external signals") appeared in the list (Table 4.3).

Novel terms relating to particular signal transduction pathways and developmental processes were identified using MetaCore (Table 4.3). Specifically, three process networks passed stringent FDR correction (FDR 0.05). In the signal transduction category, process networks associated with "TGF-β, GDF and activin signaling" and "BMP and GDF signaling" were found to be over-represented. These terms refer to signaling through the TGF-β superfamily of ligands (Massague, 2012). TGF-β ligands operate through the activation of transmembrane serine-threonine receptor kinases that phosphorylate SMAD proteins to coordinate cell-type specific gene expression. MetaCore analysis revealed that several elements of the TGF-β signaling pathway where differentially expressed after TCF4 knockdown. Differentially expressed genes included two ligands of the TGF-β family (*INHBA* and *BMP7*), the *BMPR1A* receptor, and several of the SMAD transcription factors (*SMAD2*, *SMAD4*, *SMAD7*), demonstrating that each level of the pathway was affected (Table 4.4).

Table 4.3 Process network enrichment in MetaCore. The Entrez gene IDs derived from the high-confidence list of 1031 differentially expressed genes in TCF4-knockdown cells were analyzed for enrichment using MetaCore™. Each term is presented with its functional category, network and corresponding *P* value. *Pass FDR correction of 0.05.

Category	Process Network	Genes	*P* value	FDR
Signal Transduction	TGF-β, GDF and Activin signaling	*PTPRK, PTGER2, CREB1, GATA3, CCND1, INHBA, SMAD4, FOS, SP1, SMAD2, HIF1A, ATF2, SMAD7, LEFTY1, MYC, TOB1, NOTCH1, CDK6*	3.54E-04	0.044*
Signal Transduction	BMP and GDF signaling	*CREB1, MSX2, BMPR1A, AKT1, BMP7, SMAD4, TLE1, ATF2, SMAD7, MYC, NODAL, SOX9, TOB1, CDK6*	8.02E-04	0.044*
Development	Regulation of epithelial-to-mesenchymal transition	*NOTCH1, CREB1, ACTA2, JAK2, BMP7, TRADD, SMAD4, FOS, SP1, SMAD2, HIF1A, ATF2, ABBP1, JAG1, SMAD7, ADAM17, SNAI2, IGF2, TJP1, SOX9*	8.38E-04	0.044*
Development	Regulation of telomere length	*TINF2, MAX, TEP1, SP1, HNRNPC, MYC, PTGES3, CDK6*	5.35E-03	0.190
Development	Hedgehog signaling	*NOTCH1, CREB1, HESX1, CCND1, SIRT1, INHBA, AKT1, BMP7, RBX1, SMAD4, ASCL1, ZIC1, SP1, ROCK2, PBX1. CDKN1C, HES1, JAG1, MYC, ADAM17, SOX9*	6.06E-03	0.190
Apoptosis	Death Domain receptors and caspases in apoptosis	*NGFRAP1, NGFR, TRADD, CD40, TIMP3, TNFSF10, CASP1, TRAF4, NOD1, PDCD5, CASP3, BIRC8, FAS, CASP4, CARD8*	7.26E-03	0.190
Apoptosis	Apoptosis stimulation by external signals	*NGFRAP1, JAK2, BID, NGFR, NTRK1, TRADD, SMAD4, TNFSF10, FOS, SMAD2, ADAM17, CASP3, FAS, BID*	9.66E-03	0.217
Development	Neurogenesis in general	*NOTCH1, CREB1, HESX1, GFRA3, INHBA, RCAN1, ASCL1, ZIC1, PBX1, NEUROG2, HES1, JAG1, ADAM17, CHRM3, MEF2C, SERPINI1*	1.24E-02	0.237
DNA damage	Checkpoint	*YWHAQ, CCND1, TLK2, BRIP1, CCND3, RUVBL2, ATF2, ATF3, RAD1, BARD1, MYC, YWHAZ, CDK6*	1.51E-02	0.237
Cardiac development	BMP and TGF-β signaling	*MSX2, ISL1, BMPR1A, BMP7, SMAD4, PDLIM3, MYH7, SMAD2, SMAD7, NODAL, SNAI2, SOX9, MEF2C*	1.51E-02	0.237

Table 4.4 Gene expression changes associated with TGF-β and Notch signaling pathways in TCF4-knockdown cells. Gene expression changes in manually curated signaling pathways in TCF4-knockdown cells. The data are presented as statistically significant (FDR 0.01) gene expression changes (fold change) for the genes in the TGF-β and Notch signaling pathways. *Pass FDR correction of 0.05.

Gene	Fold Change	*P* value	Description
TGF-β signaling pathway			
BMP7	-3.94	1.34E-04	BMP Ligand
BMPR1A	2.46	1.81E-05	Tyrosine kinase receptor (BMP)
SMAD2	-2.15	4.38E-04	R-SMAD (TGF/Nodal)
SMAD4	1.49	4.71E-04	co-SMAD
SMAD7	1.38	3.97E-04	I-SMAD
SMAD6	*1.23*	1.04E-02*	I-SMAD
SMAD1	*1.28*	9.59E-03*	R-SMAD (BMPR1)
NODAL	-1.25	2.27E-04	Ligand (Nodal)
INHBA	-2.38	2.68E-07	Ligand (Activin A)
LEFTY1	-2.81	3.37E-04	Ligand (Inhibitor of Nodal)
Notch signaling pathway			
NOTCH1	2.40	1.02E-04	Ligand receptor
NOTCH2	*-1.43*	8.48E-03*	Ligand receptor
JAG1	2.59	4.61E-04	Ligand receptor (inhibitor of Notch)
ADAM17	1.44	3.83E-04	metallopeptidase (Notch-cleavage)
APH1A	1.64	1.57E-04	γ-secretase complex (Notch-cleavage)
HES1	1.65	4.67E-05	Notch effector
HES7	*-1.54*	6.40E-03*	Notch effector

The third term that remained statistically significant after FDR correction was in the development category and relates to EMT (Table 4.3). EMT is a developmental process whereby cells loose their adhesive properties and become more motile. EMT is essential for neural tube formation and is thought to be a key step regulating cancer cell metastasis (Acloque et al., 2009). Two important EMT regulators, *SNAI2* and *DEC1* (*BHLHB2* in Figure 4.5A) are differentially expressed in response to TCF4 depletion. *SNAI2* promotes EMT and is downregulated in TCF4-knockdown cells whereas *DEC1* is upregulated in TCF4-knockdown cells (Figure 4.5A). It is also noteworthy that differentially expressed genes associated to the EMT category are also of general importance to development as they include elements of the Notch, BMP and IGF signaling pathways (Tables 4.3, 4.4 and 4.5). RT-PCR was used to gain further support for the differential expression of EMT genes in

TCF4-depleted cells (Figure 4.7). These data confirm the gene expression changes detected on the microarray and show that in addition to *SNAI2*, *SNA11* is also downregulated in TCF4-depleted cells. Importantly, expression of the closely related class I bHLH gene *TCF3* (*E12*/*E47*) was unaltered in TCF4-depleted cells suggesting the alterations in the EMT gene expression pathway occurred independently of E47 activity.

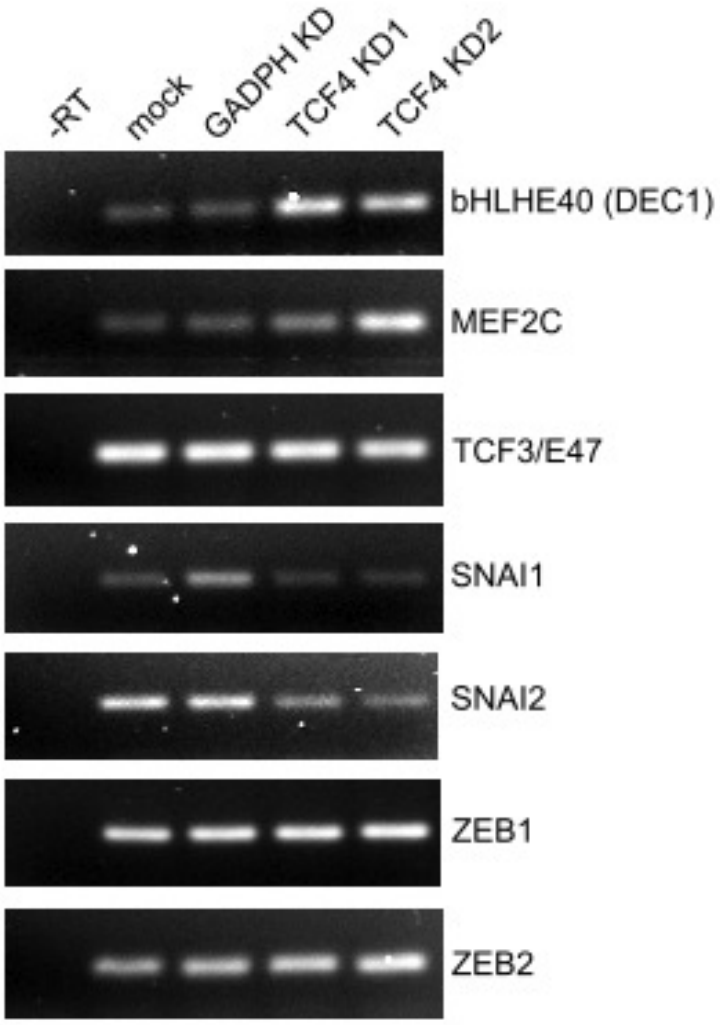


Figure 4.7 Semiquantitative RT-PCR analysis of EMT-regulating transcription factors in TCF4-knockdown cells. RT-PCR was used to confirm differential expression of several transcription factors that drive EMT. *SNAI1* and *SNAI2* transcripts are down-regulated in TCF4-knockdown cells whereas *BHLHE40* (*DEC1*) and *MEF2C* transcripts are up-regulated. These data also show that *TCF4* homologue *TCF3* (*E47*), a known regulator of EMT, is unaltered in TCF4-depleted cells. Note that alterations in *SNAI1* were not evident on the microarray possibly due to its low expression in SH-SY5Y cells.

The IGF-signaling pathway also appears to be altered in TCF4-depleted SH-SY5Y cells (Fernandez and Torres-Aleman, 2012). Specifically, *IGF2* and *GRB14* are among the most significantly downregulated genes whereas *IGFBP5* and the genes encoding other IGF-binding proteins are upregulated in our dataset (Figure 4.5 and Table 4.5). These alterations in gene expression are also evident in downstream elements of the IGF-signaling pathway and include down-regulation of several kinases and adaptor molecules such as *AKT1* and *RPS6KB1* that encodes the p70 ribosomal S6 kinase (Table 4.5).

Table 4.5 Gene expression changes associated with the IGF signaling pathway in TCF4-knockdown cells. Gene expression changes in the manually curated IGF signaling pathway in TCF4-knockdown cells. The data are presented as statistically significant (FDR 0.01) gene expression changes sorted by fold change for the genes in the IGF signaling pathway (Fernandez and Torres-Aleman, 2012). *Pass FDR correction of 0.05.

Gene	Fold change	*P* value	Description
IGF2	-5.22	1.52E-06	IGF ligand
IGFBP5	4.90	2.35E-05	IGF carrier protein
IGFBP3	1.28	1.17E-02*	IGF carrier protein
IGFBP4	2.59	3.09E-03*	IGF carrier protein
GRB14	-3.08	1.79E-05	Receptor-bound protein (inhibitor)
GRB7	-1.22	1.95E-04	Receptor-bound protein (inhibitor)
PIK3CG	-1.75	9.67E-03*	Signal transduction (kinase)
PIK3CA	1.58	6.16E-04*	Signal transduction (kinase)
PIK3C2B	-1.45	5.42E-03*	Signal transduction (kinase)
PIK3CD	-1.30	7.37E-03*	Signal transduction (kinase)
PDPK1 (PDK1)	-1.98	8.88E-05	AKT inhibition (kinase)
AKT1	1.36	3.53E-04	Signal transduction (kinase)
YWHAG	-1.97	7.64E-05	Regulatory co-factor
YWHAZ	1.26	3.63E-04	Regulatory co-factor
FOXO1A	1.60	3.75E-03*	Transcription factor
FOXO3B	-1.72	3.65E-03*	Transcription factor
RPS6KB1 (S6K1)	-1.28	3.83E-03*	Ribosomal subunit kinase
RPS6KC1	1.78	4.15E-03*	Ribosomal subunit kinase
RPS6KA3	1.47	6.83E-03*	Ribosomal subunit kinase
RPS6KA1	-1.32	2.57E-03*	Ribosomal subunit kinase

4.2.6 Dysregulation of neurogenic and neurological disease genes in TCF4-depleted cells

Human genetic studies implicate *TCF4* in neurological and neuropsychiatric disorders. In accordance with these findings, our data supports a role for TCF4 in regulating genes important for neurodevelopment. The proneural genes *ASCL1* and *NEUROG2* were found to be downregulated in TCF4-depleted cells (Figure 4.5A). In addition, a number of genes implicated in rare Mendelian forms of mental retardation were differentially expressed in TCF4 knockdown cells (Table 4.6). Interestingly, these differentially expressed genes are mutated in disorders that resemble Rett syndrome and have phenotypic similarities to PTHS. For example, *MEF2C* is found in the top 40 most significantly upregulated genes in TCF4-knockdown cells (Figure 4.5A). Furthermore, *UBE3A* (mutated in AS) is downregulated in TCF4 knockdown cells whereas the MWS gene and EMT regulator *ZEB2*, is upregulated (Table 4.6).

Table 4.6 Dsyregulation of several genes associated with intellectual disability in TCF4-knockdown cells. Data presented are the fold change (FC) and the corrected *P* value (FDR 0.01) derived from the microarray. *Pass FDR correction of 0.05, **pass FDR correction of 0.1. Abbreviations: PTHS, Pitt-Hopkins syndrome; AS, Angelman Syndrome; MWS, Mowat-Wilson syndrome, AD; autosomal dominant.

Gene	Disease	OMIM	Inheritance	FC	*P* value	Reference
TCF4	PTHS	610954	AD	-3.10	2.02 E-05	(Peippo and Ignatius, 2012)
MEF2C	Syndromic MR	613443	AD	3.73	1.37 E-05	(Zweier and Rauch, 2012)
UBE3A	AS	105830	Loss or mutation of maternal allele	-1.39	1.56 E-03*	(Mabb et al., 2011)
ZEB2	MWS	235730	AD	1.48	3.60 E-03*	(Garavelli and Mainardi, 2007)
FMR1	FXS	300624	X-linked	1.94	1.51 E-02**	(Bhakar et al., 2012)

4.3 Discussion

It is now well established that rare, highly penetrant *TCF4* alleles are associated with neurodevelopmental phenotypes whereas common variants are associated with disorders such as schizophrenia (Amiel et al., 2007; Brockschmidt et al., 2007; Zweier et al., 2007; Stefansson et al., 2009; Talkowski et al., 2012). Recent cognitive and imaging studies have also shown that TCF4 is important for normal brain function (Blake et al., 2010; Navarrete et al., 2013). However, knowledge of TCF4-regulated genes and pathways in the brain is comparatively sparse. Using genome-wide transcriptional profiling of SH-SY5Y cells, experiments in this chapter demonstrate that silencing TCF4 affects the expression of genes involved in cell signaling, cell survival and neurodevelopment. By deconstructing the top gene expression changes, it is apparent that genes controlling apoptosis tend to be upregulated whilst genes that support processes involving non-coding RNA metabolism are downregulated. Process network enrichment analysis also suggests that the TGF-β, NOTCH and IGF signaling pathways may converge on the EMT pathway in TCF4-depleted cells.

4.3.1 GO enrichment analysis implicates TCF4 in apoptosis and non-coding RNA metabolism

Several bioinformatics tools were used to search for processes and pathways that may be altered in TCF4-depleted cells. Initial results from GO term enrichment analysis using DAVID showed that apoptosis and NF-κB signaling were statistically significant processes among the upregulated genes. Shared terms in GO categories between NF-κB signaling and cell death categories show that many of the genes in each pathway are the same, indicating some degree of convergence between NF-κB signaling and apoptosis (Table 4.2). In neurons, NF-κB regulates the expression of genes participating in seemingly diverse aspects of neurodevelopment, learning and memory (Gutierrez and Davies, 2011). NF-κB signaling has also been implicated in enhancing neuronal apoptosis associated with neurodegenerative disease, brain injury and inflammatory conditions (Qin et al., 2007). Similarly in neurons, caspase activation is known to be important for axon pruning and synapse elimination and can mediate some of the chronic neuropathological events associated with brain injury or neurodegeneration (Hyman and Yuan, 2012). Since acute knockdown of TCF4 over a period of 72h reduced cell viability and increased caspase 3/7 activity, up-regulation of some caspases and components of the NF-κB pathway in TCF4-depleted cells may induce cell

death (Figure 4.6). It is therefore possible that TCF4 may regulate the expression of some caspase genes and other components of the pro-apoptotic signaling pathway in SH-SY5Y cells.

4.3.2 TCF4-knockdown affects genes in the TGF-β signaling pathway

To gain further insight in to TCF4-regulated pathways, MetaCore was utilised to identify network processes that were altered in TCF4-depleted cells. Enrichment analysis on the top 1031 differentially expressed genes, demonstrated that several components of the TGF-β signaling pathway are affected. Specifically, robust down-regulation of the TGF-β superfamily of ligands (*BMP7*, *NODAL*, *LEFTY1* and *INHBA*) was observed, as well as altered expression of several downstream components of the TGF-β signaling cascade including *BMPR1A,* and some of the SMAD transcription factors (Table 4.4). TGF-β signaling regulates many aspects of cell proliferation, differentiation, migration and apoptosis (Massague, 2012). In the nervous system, TGF-β signaling regulates neural crest formation and in also required for neurogenesis, neurite outgrowth and synaptogenesis (Panchision et al., 2001; Bond et al., 2012). Interestingly, *Bmp7* regulates the survival and proliferation of neural progenitor cells in the developing neocortex of mice and maintains *Ngn2* (the murine orthologue of *NEUROG2*) in ventricular and subventricular zones of the cortex (Segklia et al., 2012). Notably, both *BMP7* (FC, -3.93) and *NEUROG2* (FC, -3.10) are robustly downregulated in TCF4-depleted cells suggesting that TCF4 may coordinately regulate each gene (Figure 4.5A).

4.3.3 TCF4-knockdown affects EMT regulators

In addition to driving developmental programs, TGF-β signaling is also involved in EMT. Process network enrichment identified genes in the EMT pathway to be differentially expressed in TCF4-knockdown cells compared to controls (Table 4.3). This pathway governs the transition between epithelial and mesenchymal phenotypes known to be important during development and in cancer metastasis (Nieto, 2011). In this process, TGF-β has been shown to activate *SNAI1* through SMAD3/4 signaling that results in repression of critical epithelial cell genes such as *CDH1* (E-cadherin) allowing activated cells to switch to the mesenchymal phenotype (Massague, 2012). There are a variety of EMT markers that are commonly used to detect an induction of the epithelial or mesenchymal phenotype. Epithelial cells can be

identified by a variety of cell-cell and cell-basal lamina junction proteins including *CDH1* (E-cadherin), *TJP1* (*ZO-1*), *DSP* (desmoplakin, desmosomes) or *LAMA1* (laminin-1, basal lamina). Mesenchymal cells tend to be recognised by markers such as *CDH2* (N-cadherin), *VIM* (Vimentin) or *FN1* (Fibronectin) (Kalluri and Weinberg, 2009). A selection of these markers are shown to be differentially expressed on the array in TCF4-knockdown cells including *FN1*, *TJP1*, *LAMA-1* and *CDH2* (Table 4.3 and data not shown).

In addition to components of the TGF-β signaling pathway and cell adhesion markers, transcriptional regulators of the EMT pathway *SNAI1/2*, *ZEB2*, *BHLHB2 (DEC1*) and *TWIST1* were differentially expressed in TCF4-knockdown cells (Figure 4.5A, Figure 4.7, and data not shown). SNAI1 and 2 are transcriptional repressors that bind to E-boxes in the promoter regions of genes including *CDH1* (Cobaleda et al., 2007). ZEB2 is an E-box-binding homeobox protein that induces EMT by repressing a variety of genes involved in epithelial cell–cell junctions. *ZEB2* has also been implicated in interneuron subtype specification in the developing brain (McKinsey et al., 2013). DEC1 is another transcriptional repressor that regulates EMT in pancreatic cancer cells in response to TGF-β stimulation (Wu et al., 2012). TWIST1 is repressor of E-protein activity and an important mediator of EMT (Spicer et al., 1996; Yang et al., 2004). Together this data establishes that knockdown of TCF4 regulates a complex assortment of genes related to EMT, consistent with findings in other cell types where E-proteins are shown to have an important role in this pathway (Bolos et al., 2003; Moreno-Bueno et al., 2006; Sobrado et al., 2009).

The function of the EMT has been well characterised in embryonic development and cancer however its function in neuronal cells is unclear. Mechanistically, aspects of EMT can be likened to neural stem and progenitor cell delamination, a critical step in early neurogenesis (Pacary et al., 2012). During cortical development, neural stem cells have a distinctive polarity and are anchored to the basal membrane of the ventricular zone, forming the neuroepithelium. During neurogenesis, neural precusors delaminate from the basal lamina and migrate into the cortex to form new neurons in a process similar to EMT. Repression of N-cadherin (*CDH2*), a marker of EMT, promotes this detachment from the neuroepithelium and regulates the subsequent migration of neural progenitors (Jossin and Cooper, 2011; Rousso et al., 2012). These early steps in neurogenesis are initiated by NEUROG2 that in turn activates *FOXP2* and *FOXP4* to repress N-cadherin expression (Heng et al., 2008;

Rousso et al., 2012) and allows cells to adopt a migratory character. Interestingly, NEUROG2 (FC -3.10, FDR0.01) and FOXP2 are downregulated (FC -1.2, FDR0.05) in the TCF4-knockdown data set, and N-cadherin is upregulated (FC 1.8, FDR0.05), consistent with molecular characteristics of cells remaining in the neuroepitheium. This observation suggests that if an analogous situation were to occur in neuronal stem cells, *TCF4* expression may promote delamination and neurogenesis in an EMT-like process. This mechanism has already been proposed for snail family members *SCRT1* and *SCRT2*, advocating a role for EMT regulators in early neurogenesis and migration (Itoh et al., 2013).

4.3.4 Gene expression changes in the IGF signaling pathway

In addition to changes in gene expression associated with the TGF-β and NODAL signaling, TCF4-knockdown also alters the expression of components of the IGF signaling pathways (Table 4.5). TCF4-knockdown is associated with a 5.2 fold down-regulation of *IGF2* and up-regulation of the genes encoding IGF binding proteins, IGFBP3, -4 and -5 (Figure 4.5A and Table 4.5). In the brain, IGF2 is required for memory consolidation and enhancement (Chen et al., 2011), adult hippocampal neurogenesis (Bracko et al., 2012), synapse formation and dendritic spine maturation (Schmeisser et al., 2012). Since *IGF2* has a role in learning and memory and PTHS patients have profound intellectual disability (Van Balkom et al., 2012), TCF4 regulation of *IGF2* expression may be a determinant of cognitive dysfunction. IGF1 treatment has been shown to reverse some of the neurophysiological abnormalities in a mouse model of Rett Syndrome lacking methyl CpG-binding protein 2 (*Mecp2*, (Tropea et al., 2009)). Because IGF1 and IGF2 activate the same receptor, IGF2 may have some utility in reversing some of the cognitive deficits in PTHS patients. Another component of the IGF pathway *RPS6KB1* (encoding a member of the ribosomal protein S6 kinase family), is also downregulated in TCF4-depleted cells (Table 4.5). *RPS6KB1* may be particularly important in the context of neurodevelopment because genetic ablation of this gene rescues multiple physiological and behavioural phenotypes in a mouse model of fragile X syndrome, caused by aberrant synaptic translation (Bhattacharya et al., 2012).

4.3.5 Altered expression of neurodevelopmental genes in TCF4-knockdown cells

ASCL1 and *NEUROG2* are important neurogenic bHLH transcription factors that interact directly with TCF4 and appear among the top downregulated genes in TCF-depleted cells (Figure 4.5A). Since both of these proneural genes are downregulated, this may indicate that

as well as regulating proneural activity through protein:protein interactions, TCF4 may also regulate proneural gene expression. In addition to *ASCL1* and *NEUROG2*, other neurodevelopmental transcriptional regulators such as, *MEF2C* (syndromic mental retardation) and *ZEB2* (MWS) were also differentially expressed in TCF4-depleted cells.

PTHS, MWS and AS are sometimes classified as Rett-like syndromes because of their similar clinical presentation and genetics (Armani et al., 2012; Peippo and Ignatius, 2012). Furthermore, haploinsufficiency of *MEF2C* also results in a form of severe mental retardation, with absent speech, hypotonia and epilepsy (Le Meur et al., 2010; Zweier and Rauch, 2012). Importantly, genes for each of these phenotypically similar disorders (*MEF2C*, *ZEB2* and *UBE3A*) were differentially expressed in TCF4-depleted cells (Table 4.6). From a mechanistic perspective it is interesting to note that, TCF4, ZEB2 and MEF2C can all regulate E-box activity in the promoters of certain genes (Molkentin and Olson, 1996; Remacle et al., 1999). These data, allied with the phenotypic overlap between these disorders, suggest that each of these genes may participate in a similar neurodevelopmental transcriptional pathway. Thus, alterations in the activity or levels of TCF4 as seen in PTHS, and possibly and schizophrenia, may be associated with dysregulation of several transcription factors that control neurodevelopmental gene expression programs at E-box containing promoters.

Chapter 5

Proteomic Analysis of TCF4 Isoforms

5.1 Introduction

The *TCF4* locus is particularly complex, encoding multiple transcripts with different N-termini and at least two major protein isoforms (Figure 1.2, Figure 5.1, Table 5.1). The evidence to support the existence of most protein isoforms is largely based upon conceptual translation of different transcripts generated by RT-PCR (Sepp et al., 2011). The lack of isoform specific antibodies or antibodies with epitopes in specific exons has impeded the confirmation of their existence in cells and tissues.

The transcriptional architecture of the *TCF4* gene has already been extensively characterised in humans (Sepp et al., 2011). Expression of TCF4 is widespread in different tissues and is particularly high in the brain where most of the alternatively spliced transcripts exist. The vast array of alternatively splice forms in humans consequently raises questions about TCF4 protein diversity and function in the brain. Originally, two highly related protein isoforms were proposed to exist in cells (TCF4-B and TCF4-A) although three distinct protein isoforms have been detected in Neuro2A cell extracts (Skerjanc et al., 1996; Sepp et al., 2011).

The TCF4-B isoform is considered the full-length canonical protein and contains the two conserved activation domains, AD1 and AD2 (Figure 5.1 and Appendix IV). In contrast, TCF4-A is a truncated isoform that contains an N-terminus devoid of the AD1 and the nuclear localization signal (NLS) (Figure 5.1 and Appendix IV). It is widely thought that TCF4 can interact with a number bHLH proteins that regulate its DNA binding properties and transcriptional activity (Sections 1.4 and 1.9.1). E-proteins have also been shown to interact with chromatin modifying proteins through the AD1 and AD2 domains (Section 1.5.1). Although the N-terminal AD1 and AD2 domains are highly conserved among E-proteins, the interactions with chromatin

modifiers have only been extensively demonstrated for TCF3 (E2A) and TCF12 (HEB) and there is a paucity of evidence for these interactions in neuronal cells (Bayly et al., 2004; Zhang et al., 2004; Guo et al., 2009; Denis et al., 2012).

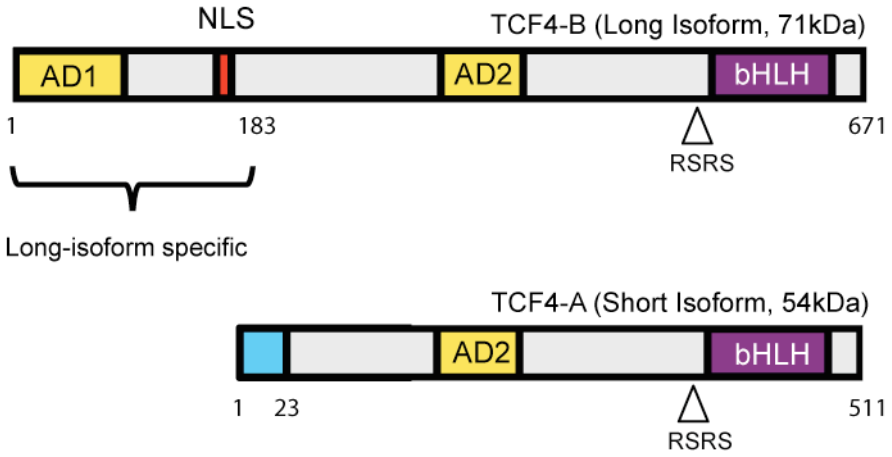


Figure 5.1 Structure of the TCF4-B and TCF4-A Isoforms. This illustration shows TCF4 isoforms and the organization of their functional domains. TCF4-B contains a set of 183 N-terminal amino acids that is specific to long isoforms. TCF4-A lacks the N-terminal sequences encoding the AD1 domain and NLS but contains a unique 5' exon (blue) that allows it to be distinguished from other short isoforms. For clarity, repressor domains were omitted from the illustration. Abbreviations: AD1, activation domain 1; AD2, activation domain 2 (putative leucine zipper); NLS, nuclear localization signal; RSRS, RSRS amino acid insertion sequence; bHLH, basic helix-loop-helix domain. Illustration not to scale.

To gain further insight into the function of TCF4 in SH-SY5Y cells, immunoaffinity purification (IAP) and mass spectrometry (MS) were used to isolate and identify TCF4 and its associated proteins. To efficiently purify TCF4 from SH-SY5Y cells, polyclonal anti-TCF4 antibodies were generated. Immunoaffinity purified TCF4 isoforms were unequivocally identified by MS. Finally, MS data was used to determine TCF4 phosphorylation sites and identify potentially novel interacting proteins that co-purified with TCF4.

Table 5.1 List of TCF4 isoforms in the NCBI database ordered by predicted molecular weight. The NCBI protein database entry for TCF4 is provided along with the corresponding names from the Sepp *et al.* 2012 publication. Schematics of transcript structure are depicted in Figure 1.1. Isoforms that contain unique peptide sequences are indicated. The presence of unique peptide sequences allows isoforms to be unambiguously identified by MS.

TCF4 Isoform	NCBI Reference Sequence	Number of amino acids	Predicted MW (kDa)	Sepp *et al.* 2012	Unique peptides
Isoform c	NP_001230155	773	83.4	TCF4-J	yes
Isoform e	NP_001230157	677	72.4	-	yes
Isoform a	NP_001077431	671	71.8	TCF4-B+	no
Isoform b	NP_003190	667	71.3	TCF4-B-	no
Isoform f	NP_001230159	664	71.2	TCF4-E	yes
Isoform d	NP_001230156	647	69.1	TCF4-C	no
Isoform g	NP_001230160	625	67.0	TCF4-F	yes
Isoform h	NP_001230161	600	64.6	TCF4-G	yes
Isoform i	NP_001230162	537	57.4	TCF4-D	no
Isoform j	NP_001230163	511	54.6	TCF4-A+	yes
Isoform l	NP_001230165	507	54.2	TCF4-I	yes
isoform k	NP_001230164	507	54.1	TCF4-A-	yes

5.2 Results

5.2.1 TCF4 antibody design and characterization

There are three E-protein paralogues in humans each with at least two distinct protein isoforms – TCF3 (E2A), TCF4 (E2-2), TCF12 (HEB). In order to generate a specific TCF4 antibody, care was taken to find amino acids unique to TCF4. A multiple sequence alignment was used to determine a suitable amino acid sequence that was present in all human TCF4 isoforms and that was sufficiently divergent to other E-proteins. A sequence corresponding to amino acids 361-554 in the full length TCF4-B sequence was identified. This 193 amino acid region had approximately 40% sequence similarity to TCF3 and 60% sequence similarity to HEB (data not shown). The nucleotide sequence corresponding to the selected amino acids (exons 14-18 in full length TCF4-B mRNA) was cloned into two different bacterial expression vectors in order to generate thioredoxin (TRX-TCF4) and GST (GST-TCF4) fusion proteins. Each protein was synthesised in bacterial cells and affinity purified before being injected into rabbits. The rabbits were immunised four times over a period of 88 days to produce a robust immune response. Following the immunization protocol, serum was extracted from the rabbits and the TCF4-specific antibodies were immunoaffinity purified using the cognate antigen linked to a column.

The purified TCF4 antibodies isolated from the GST-TCF4 (TCF4_01) and TRX-TCF4 (TCF4_02) immunizations were subsequently characterised for their ability to detect the TCF4 protein. To determine whether the TCF4 polyclonal antibodies could detect denatured TCF4 epitopes, the antibodies were tested by western blotting using extracts from SH-SY5Y and HEK-293T cells (Figure 5.2). Both antibodies were capable of detecting over-expressed TCF4 and endogenous TCF4 isoforms of approximately 71kDa and 54kDa (Figure 5.2A, lane 1 and 2). Similar results were obtained with both polyclonal antibodies, so only the data for TCF4_01 is presented. To ensure the bands detected in the SH-SY5Y cell extract were TCF4-specific, an extract from SH-SY5Y knockdown cells (Section 4.2.1) was included on the blot (Figure 5.2A, lane 3). Both bands from this extract were depleted confirming that the two bands on the blot were TCF4 isoforms. Identical blots were also incubated with the commercial anti-TCF4 antibody (Abnova, M03) used in Chapter 4 (Figure 4.1B).

This experiment demonstrated that Abnova, M03 only detected full length TCF4 (71kDa).

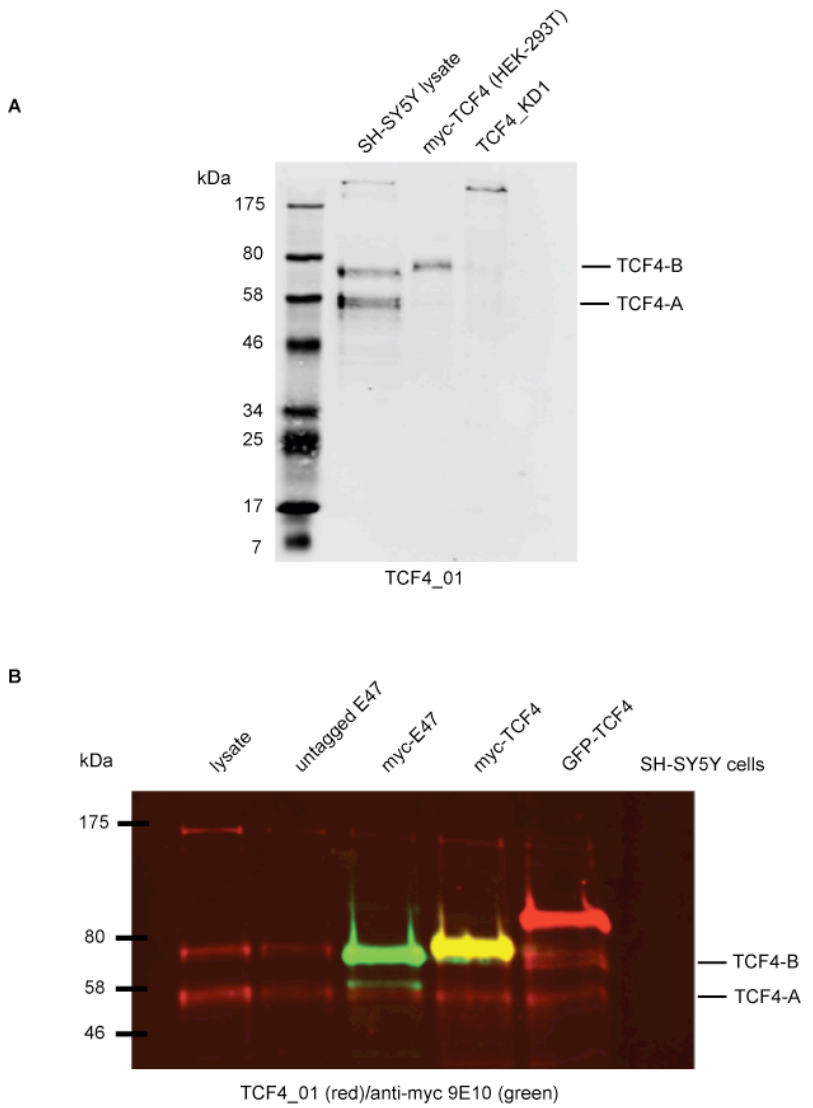


Figure 5.2 Characterization of anti-TCF4 polyclonal antibodies
The TCF4 polyclonal antibody TCF4_01 was used to detect over-expressed or endogenous TCF4 in different cells lines (A). HEK-293T cells were transfected with 1µg of myc-tagged TCF4 constructs and cellular lysates were prepared 24h after transfection. SH-SY5Y cells were either left untreated (lane 1) or transfected with siRNAs targeted against all TCF4 transcripts before preparing cellular lysates (lanes 1 and 3 respectively). (B) Two-channel western blotting imaging of SH-SY5Y cellular lysates either left untreated (lane 1) or transfected with 1ug of various E47 (TCF3) or TCF4 constructs for 24h (lanes 2-5). Western blots were incubated with anti-myc (9E10, green channel) and anti-TCF4 polyclonal antibodies (TCF4_01, red channel). TCF4 antibodies detect endogenous and overexpressed TCF4 (myc-TCF4, GFP-TCF4) and do not detect overexpressed E47 (pcDNA-E47, myc-E47). Anti-myc antibodies are able to detect the overexpressed myc-E47 and myc-TCF4 proteins. The myc-TCF4 construct can be recognised both the anti-myc and anti-TCF4 antibodies (yellow).

Western blotting was also used to determine whether the anti-TCF4 antibodies cross-reacted with other E-proteins. Another western blot was performed with SH-SY5Y cell extracts from cells overexpressing TCF4 and the TCF4 paralogue E47 (TCF3). The TCF4 antibodies were able to detect over-expressed TCF4 and did not detect the E47 protein, even when highly over-expressed (red channel, Figure 5.2B). Anti-myc antibodies were used to confirm the expression of each myc-tagged fusion protein in transfected cells (green channel, lanes 3 and 4, Figure 5.2B). These experiment indicated that the TCF4 polyclonal antibodies specifically detect TCF4 and not E47 on western blots.

To further validate the antibody, immunocytochemistry was used to detect native TCF4 in untransfected SH-SY5Y cells (Figure 4.2). Each TCF4 polyclonal antibody confirmed the nuclear localization of the TCF4 that had been established by the over-expression of TCF4 constructs (Figure 3.2). The nuclear signal was diminished in siRNA treated cells demonstrating the specificity of the antibodies in these experiments (Figure 4.2).

5.2.2 Immunoprecipitation of TCF4 and characterization of isoforms in SH-SY5Y cells

Immunoaffinity purification combined with mass spectrometry (IAP-MS) were used to confirm the identity of the two TCF4 isoforms previously detected by western blotting (71kDa and 55kDa) in SH-SY5Y cells (Figure 5.2). Initially, small scale immunoprecipitations (IPs) were used to enrich each TCF4 isoform. Proteins were solubilised from SH-SY5Y cells in RIPA buffer prior to incubation with TCF4_01 and TCF4_02. Following overnight incubation with the primary antibody, immune complexes were captured on protein A beads. The beads were washed stringently with RIPA buffer to eliminate non-specific binding of proteins to the protein A beads. The antibody-TCF4 complexes were denatured in SDS loading buffer and the samples were resolved by western blotting (Figure 5.3). This experiment demonstrated the TCF4_01 and TCF4_02 antibodies efficiently immunoprecipitated TCF4 from SH-SY5Y cells. The IP protocol efficiently depleted TCF4 from the cellular extracts (lanes 4 and 5) and the TCF4-antibody complexes were strongly enriched on the protein A beads (lanes 8 and 9) (Figure 5.3). Importantly, TCF4 did not precipitate on

the protein A beads or with a non-specific antibody (anti-myc), as no TCF4 protein was detected in the control IPs (lanes 6 and 7).

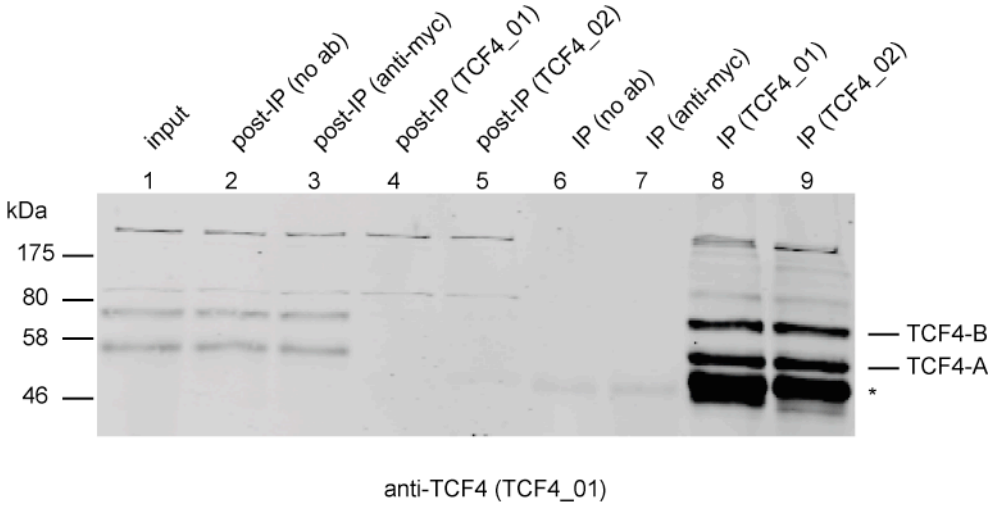


Figure 5.3 Characterization of TCF4 polyclonal antibodies by immunoprecipitation. Western blotting was used to show that the TCF4 polyclonal antibodies could immunoprecipitate TCF4 from SH-SY5Y cells. Cellular extracts were incubated with TCF4 antibodies (TCF4_01 and TCF4_02) overnight and TCF4-antibody complexes were subsequently captured on protein A agarose beads. The eluted TCF4 complexes were resolved by SDS-PAGE and detected by western blotting (IP, lanes 8 and 9). A mock immunoprecipitation (IP) with no antibody (no ab) and a negative control IP with an irrelevant antibody (anti-myc) was also performed to ensure TCF4 did not precipitate non-specifically (IP, lanes 6 and 7). The cellular extracts were also sampled before (Input, lane 1) and after the IPs (post-IP, lanes 2-5) to demonstrate the extent of TCF4 depletion in each experiment. Asterisk shows the position of the IgG heavy chain.

Having established that both antibodies were suitable for IP, a large-scale IAP experiment was performed with cross-linked antibody beads (TCF4_01) to enrich sufficient amounts of TCF4 for MS. The immunoprecipitated complexes were separated on a 4-12% polyacrylamide gel that was subsequently stained with colloidal Coomassie blue (Figure 5.4). Several bands were visible after staining, including proteins potentially corresponding to the approximate molecular mass of the two TCF4 isoforms previously identified. Seven gel plugs were excised and sent for analysis by MS (bands 1-7, Figure 5.4).

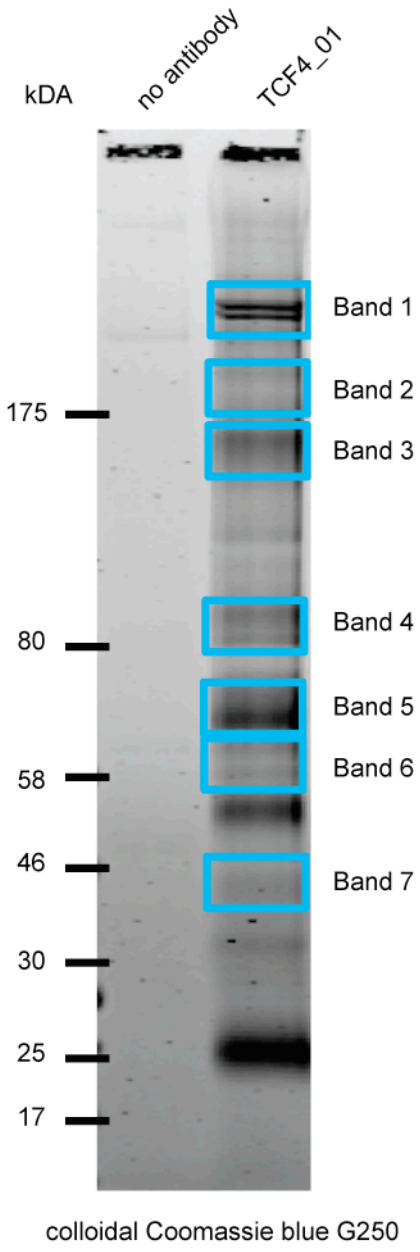


Figure 5.4 Immunoaffinity purification of TCF4 for isoform detection by mass spectrometry. TCF4 was immunoaffinity purified from SH-SY5Y cell extracts using protein A beads cross-linked to TCF4_01. A mock IAP was also performed with no antibody. TCF4 was eluted from the beads and resolved on a 4-12% gradient gel by SDS-PAGE. Gel slices were excised as indicated on the image (bands 1-7) and analysed by mass spectrometry to identify TCF4 isoforms. Approximate molecular weight is provided on the image in kilodaltons (kDa).

A total of 497 TCF4 peptides were identified by MS. Although TCF4 peptides were present in multiple bands, the highest peptide count was found in band 4 (142 peptides) spanning 70% of TCF4-B protein (Table 5.2, Figure 5.4). To search for specific TCF4 isoforms present in the data, unique peptides were defined from TCF4 protein sequences in the NCBI database (12 isoforms, Table 5.1). Eight of the twelve TCF4 isoforms in the NCBI database had at least one unique peptide, meaning they could be unambiguously identified from the MS data. Unique peptides are mainly found in the first 5' exon of each isoform, such as with TCF4-A (blue exon, Figure 5.1), therefore there are only a very limited number of unique peptides that can be generated. Manual examination of the MS data identified two unique peptides for the TCF4-A isoform (NCBI isoform j/k) and one unique peptide for the isoform l (Table 5.2). Most of the TCF4-A associated peptides were found in bands 5 and 6 of the polyacrylamide gel (Figure 5.4).

The MS data also confirmed the presence of long TCF4 isoforms. Peptides derived from the N-terminal fragment of TCF4-B (amino acids 1-183) that are absent in the TCF4-A isoform were readily detected in the data (Figure 5.1, Table 5.2). The TCF4-B isoform does not contain a unique 5' exon therefore it cannot be unambiguously distinguished from other long isoforms of similar molecular weight using MS (isoforms d-f, Table 5.1). However, the MS data did not contain evidence for the long isoforms e and f that do have unique peptide sequences.

In addition to each N-terminal isoform, there was extensive evidence of specific peptides relating to the presence (+) and absence (-) of the RSRS amino acid insertion sequence (Table 5.2). The RSRS insertion is present in a region of the protein shared across all isoforms (coding exon 17 of TCF4-B). Both TCF4-A and TCF4-B have previously been reported to contain the + and – RSRS isoforms (Sepp et al., 2011).

Table 5.2 Identification of TCF4 and isoform-specific peptides in SH-SY5Y cells. Protein complexes obtained from IAP of TCF4 from SH-SY5Y cells were analysed by MS to characterise TCF4 isoforms (Figure 5.4, bands 1-7). TCF4 peptides were identified in all bands (a total of 497 peptides) with the highest abundance of TCF4 peptides being identified in band 4 (protein score 1640, 70% coverage of full length TCF4-B protein). Non-unique and unique TCF4 peptides are presented. Band 4 contained evidence for the TCF4-B isoform, covering the first 183 amino acids of the TCF4-B sequence that are absent in TCF4-A. However, the TCF4-B protein sequence does not contain any unique peptides and shares much of its sequence with other long isoforms. Unique isoform-specific peptides relating to the + and – RSRS isoforms and well as the short isoforms TCF4-A (Isoform j/k) and Isoform l were identified in the data. The characteristics and nomenclature of particular isoforms is summarised in Table 5.1.

Isoform	Peptide sequence	Total # peptides	Highest peptide abundance
Isoforms without unique peptide sequences			
TCF4 (all isoforms)	*C-terminal peptides (amino acids 184-671)* NGGQASSSPNYEGPLHSLQSR GSGAAGSSQTGDALGK LDDAIHVLR SGTNHYSTSSCTPPANGTDSIMANR LLILHQAVAVILSLEQQVR ALASIYSPDHTNNSFSSNPSTPVGSPPSLSAGTAVWSR GMPPGLQGQSVSSGSSEIK GMPPGLQGQSVSSGSSEIKSDDEGDENLQDTK LSYPSHSSADINSSLPPMSTFHR GSHSLLPNQVPVPQLPVQSATSPDLNPPQDPYR VSSEPPPLSLAGPHPGMGDASNHMGQM MVQLHLK DINEAFKELGR DINEAFK HSLMVGTHREDGVALR	497	Band 4
TCF4-B (and other long isoforms)	*N-terminal peptides (amino acids 1-183)* NGPTSLASGHFTGSNVEDR DLGSHDNLSPPFVNSR NYGDGTPYDHMTSR ESNLQGCHQQSLLGGDMDMGNPGTLSPTKPGSQYYQYSSNNPR SSSGSWGNGGHPSPSR ELSDLLDFSAMFSPPVSSGK KVPPGLPSSVYAPSASTADYNR VPPGLPSSVYAPSASTADYNR		
Isoforms with unique peptide sequences			
+ RSRS	SSNNDDEDLTPEQK	29	Band 6
– RSRS	SITSNNDDEDLTPEQK	51	Bands 4/5
TCF4-A (Isoform j/k)	MYCAYTIPGMGGNSLMYYYNGK	4	Band 6
	AVYAPSASTADYNR	13	Bands 5/6
TCF4 (Isoform l)	AEVYAPSASTADYNR	1	Band 5

5.2.3 Phosphoproteomic analysis of TCF4

MS is also able to give precise information about amino acid modifications such as phosphorylation and acetylation. Many of these fixed modifications can be inferred from peptide mass data derived from the Orbitrap experiments (Taus et al., 2011). For example, phosphorylation increases the mass of a peptide by approximately 80Da (Larsen et al., 2001). Phosphorylation is a particularly interesting post-translational modification to investigate because it is known to modify the DNA binding activity of bHLH proteins (Johnson et al., 1996). The TCF4 peptide information was therefore filtered to obtain a list of all the phosphorylated peptides. In total, ten phosphoserine and three phosphothreonine residues were consistently observed in the data (Table 5.3).

Table 5.3 TCF4 phosphopeptides identified by mass spectrometry. TCF4 peptides generated by IAP-MS were filtered to obtain a list of phosphopeptides. The phosphorylated residue within each peptide sequence is indicated with a lower case letter. The position of the phosphorylated residues was determined relative to full length TCF4-B protein sequence (671 amino acids, Figure 5.5).

Peptide sequence	TCF4-B residue
MAALGTDKELSDLLDFSAMF**s**PPVSSGK	S27
SSSGSWGNGGHP**s**PSR	S66
DLGSHDNL**s**PPFVNSR	S92
ESNLQGCHQQSLLGGDMDMGNPGTL**s**PTKPGSQYYQYSSNNPR	S140
SGTNHYST**s**SCTPPANGTDSIMANR	S294
GSGAAGS**s**QTGDALGK	S318
ALASIYSPDHTNNSFSSNPSTPVG**s**PPSLSAGTAVWSR	S351
NGGQASS**s**PNYEGPLHSLQSR	S372
GMPPGLQGQSVSSGS**s**EIKSDDEGDENLQDTK	S511
GMPPGLQGQSVSSGSSEIK**s**DDEGDENLQDTK	S515
ESNLQGCHQQSLLGGDMDMGNPGTLSP**t**KPGSQYYQYSSNNPR	T142
SGTNHYSTSSC**t**PPANGTDSIMANR	T297
SI**t**SNNDDEDLTPEQK	T545

The phosphopeptides were subsequently aligned to the TCF4 protein sequence to establish the positioning of the phosphorylated residues in relation to TCF4's functional domains (Figure 5.5). These sites were distributed across the coding sequence including three sites within AD1 (S27, S66, S92) and two within AD2

(S351, S372) (Figure 5.5, Table 5.3). Interestingly, the amino acid T545 was only phosphorylated in the – (minus) isoform that does not contain the RSRS sequence immediately adjacent to the bHLH domain.

Having identified a number of potential TCF4 phosphorylation sites in SH-SY5Y cells, the full length TCF4 sequence was screened for known kinase phosphorylation motifs using PhosphoMotif Finder (Section 2.10.1). This database defines motifs based on manual curation of published experimental data and provides a list of putative kinases that are capable of phosphorylating a given residue within an amino acid motif. The database predicts phosphorylation sites for an array of kinases that collectively regulate diverse processes such as growth and differentiation, cell cycle regulation, DNA repair and synaptic plasticity (Table 5.4) (Unger et al., 2004; Knippschild et al., 2005; O'Driscoll and Jeggo, 2006; Hur and Zhou, 2010; Kim and Choi, 2010; Rebola et al., 2010; Lim and Kaldis, 2013).

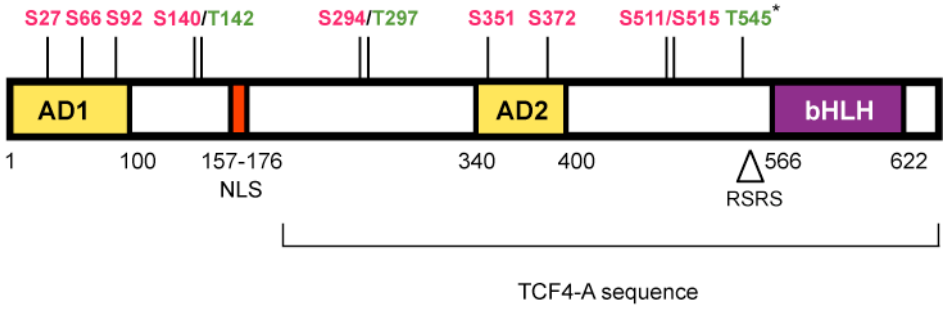


Figure 5.5 TCF4 phosphorylation sites in SH-SY5Y cells. This illustration depicts the location of TCF4 phosphorylation sites identified by MS. Serine (S) and threonine (T) phosphorylation sites are presented with their amino acid positions relative to the full length TCF4-B^{+} sequence. The approximate amino acid positions of functional domains are indicated as well as the region of the protein shared with TCF4-A. The triangle demonstrates the approximate location of the RSRS insertion sequence. Asterisk: The T545 phosphorylation site is only present in the – isoforms that lack the RSRS insertion sequence. For clarity, repressor domains were omitted from the illustration. Abbreviations: AD1, activation domain 1; AD2, activation domain 2; NLS, nuclear localization signal; RSRS, RSRS amino acid insertion sequence; bHLH, basic helix-loop-helix domain. Illustration not to scale.

Table 5.4 Phosphorylated residues in the TCF4 amino acid sequence and their predicted kinases. The location of phosphorylated residues in TCF4 are indicated. The numbering of amino acid residues and exons are in relation to the full length coding sequence of TCF4-B+ (671 amino acids). The residues were predicted to be phosphorylated by specific kinases using PhosphoMotif Finder. The T545 phosphorylation site was only observed on unique peptides corresponding to TCF4 isoforms that lack the RSRS insertion. Abbreviations: S, Serine; T, Threonine; AD1, activation domain 1; AD2, activation domain 2; GSK3B, glycogen synthase kinase 3B; ERK1, extracellular-signal-regulated kinase 1; ERK2, extracellular-signal-regulated kinase 2; MAPKAPK2, mitogen-activated protein kinase-activated protein kinase 2; CDK1, cyclin-dependent kinase 1; CDK2, cyclin-dependent kinase 2; CDK4, cyclin-dependent kinase 4; CDK5, cyclin-dependent kinase 5; CDK6, cyclin-dependent kinase 6; PRKDC , DNA-activated protein kinase; CKI, casein kinase I; CKI, casein kinase II; GRK1, G protein-coupled receptor kinase 1; ATM, ataxia telangiectasia mutated; PKA, cAMP-dependent protein kinase,; PKC , protein kinase C.

Phospho Residue	Exon	Domain	Predicted phosphorylating kinase
S27	2	AD1	GSK3B, ERK1, ERK2, CDK5, MAPKAPK2
S66	3	AD1	GSK3B, ERK1, ERK2, CDK5, PRKDC, CDK1, CDK2, CDK4, CDK5, CDK6
S92	4	AD1	GSK3B, ERK1, ERK2, CDK5, CKI
S140	6	-	GSK3B, ERK1, ERK2, CKII, CDK1, CDK2, CDK4, CDK5, CDK6
T142	6	-	PRKDC
S294	10	-	GRK1
T297	10	-	GSK3B, ERK1, ERK2, CDK5
S318	11	-	PRKDC, ATM
S351	12	AD2	GSK3B, ERK1, ERK2, CDK5, CKI
S372	13	AD2	GSK3B, ERK1, ERK2, CDK5
S511	16	-	MAPKAPK2
S515	17	-	CKI, CKII
T545	18	– isoform	PKA, PKC

5.2.4 Proteomic analysis of TCF4 binding proteins

In addition to providing important information concerning post-translational modification of TCF4, IAP-MS can be used to identify TCF4-associated proteins. E-proteins regulate cell differentiation through interacting with an assortment of tissue-specific bHLH transcription factors. Furthermore, the E-proteins have been demonstrated to interact with co-activators and co-repressors that bind to their N-terminal activation domain (Massari et al., 1999; Zhang et al., 2004). These interactions have been essential in understanding the function of TCF4 within the cell however the repertoire of proteins that specifically interact with TCF4 is sparse.

In order to further characterise the TCF4 interactome, IAP-MS was used to isolate and identify proteins that bind to TCF4 in SH-SY5Y cells. In this experiment, TCF4 complexes were isolated using two different polyclonal antibodies (TCF4_01 and TCF4_02) to minimise the discovery of false positive interactions due cross-reactivity of the antibodies. In addition, one negative control antibody was used that consisted of IgG purified from pre-immunised rabbit serum. Each of the three antibodies was cross-linked separately to protein A beads and subsequently incubated with protein extracts from SH-SY5Y cells (Section 2.7.1). The beads were stringently washed and the captured TCF4 complexes were eluted in SDS-containing buffer. Resolving the TCF4-associated protein complexes on a gradient gel confirmed that a range of proteins had been immunoprecipitated with each antibody (Figure 5.6). The distribution of immunoprecipitated proteins resolved by SDS-PAGE seemed to suggest the reactivity profiles of the antibodies were different (Figure 5.6). Multiple (equivalent) regions of the gel were excised and processed for MS as described above (bands 1-8, Figure 5.6). As an additional negative control, a mock IAP-MS experiment was performed with no antibody to try and identify proteins that bind to the agarose gel matrix; this is commonly known as the bead proteome. Protein complexes in this experiment were eluted in RapiGest reagent and analysed directly by MS. This experiment generated a background list of proteins that could be used in downstream data processing to filter out non-specific interactions (Table 5.5).

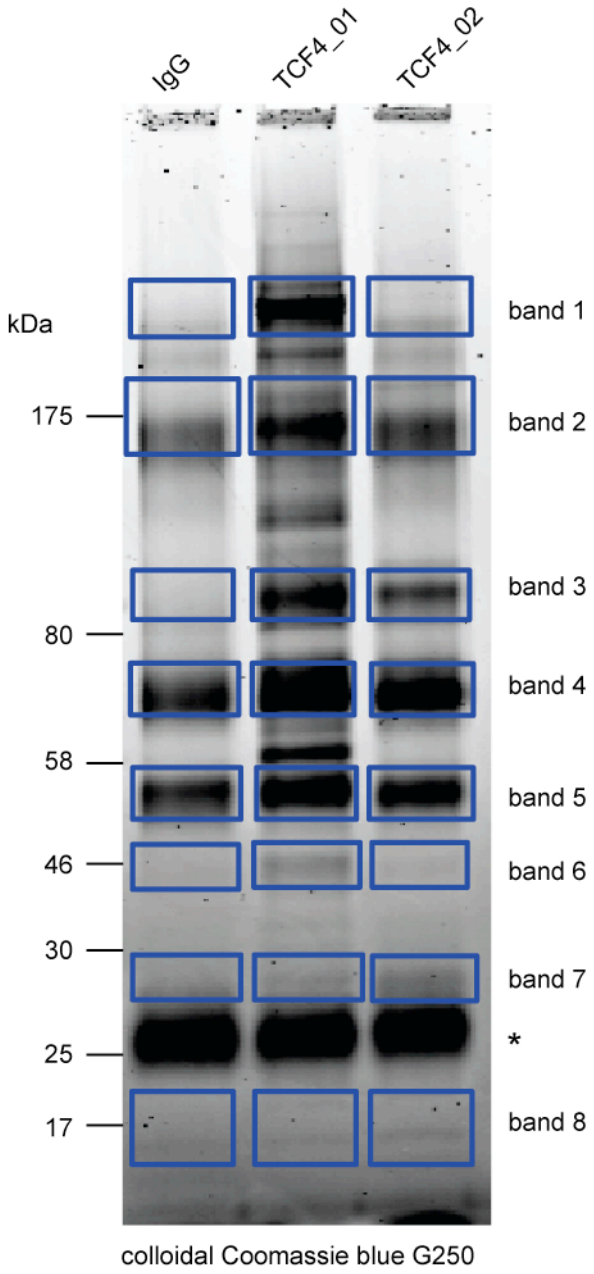


Figure 5.6 Co-immunoprecipitation of TCF4 binding proteins for analysis by mass spectrometry. Protein A beads were cross-linked with either TCF4 polyclonal antibodies (TCF4_01, TCF4_02) or pre-immune IgG (IgG) and used to co-immunoprecipitate TCF4 complexes from SH-SY5Y cell extracts. TCF4 complexes were eluted from the beads and resolved on a 4-12% gradient gel by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Gel slices were excised as indicated on the image (IgG; bands A1-A8, TCF4_01; bands B1-B8, TCF4_02; bands C1-C8) and analysed by mass spectrometry to detect potential TCF4-interacting proteins. Approximate molecular weight is provided on the image in kilodaltons (kDa).

Table 5.5 SH-SY5Y agarose bead proteome. This is a list of proteins that bind non-specifically to protein A agarose beads and represent a common source of false-positive interactions in IAP-MS experiments. The protein names are presented with their individual protein scores generated by the Mascot software package. The number of peptides identified by MS for each protein is presented with the % amino acid coverage this represents compared to the full length protein sequence (# peptides, % coverage).

#	Gene	Protein score	% coverage	# peptides
1	KRT10	747.3	41.27	20
2	KRT1	497.01	26.55	15
3	MB21D2	326.57	23.93	13
4	KRT2	301.86	26.13	16
5	HSPA9	221.42	15.19	8
6	KRT14	167.07	10.81	5
7	KRT5	150.18	14.31	8
8	KRT9	123.27	16.45	6
9	HIST1H2AJ	122.66	28.91	3
10	ACTG1	113.76	18.32	5
11	HSPA8	110.96	16.59	7
12	DPYSL2	70.2	7.32	4
13	PCMT1	69.22	18.52	3
14	ALB	65.32	9.83	4
15	RBM39	63.45	2.97	1
16	CDC23	57.1	7.1	4
17	H3F3A	56.25	17.65	3
18	H3F3B	56.25	17.65	3
19	H3F3AP4	56.25	17.65	3
20	H3F3AP6	56.25	17.65	3
21	HSPA5	56.07	10.02	6
22	DUT	52.45	5.67	1
23	PZP	52.29	3.94	1
24	C4A	51.17	12.55	4
25	GRN	48.87	3.2	1
26	HIST1H1D	44.8	17.19	4
27	ANAPC7	36.83	4.39	2
28	RPL35	34.93	8.13	1
29	RPL35P1	34.93	8.13	1
30	RPL35P2	34.93	8.13	1
31	SNRPD1	32.55	16.81	1
32	DDX3X	32.14	5.64	2
33	ITPR1	31.92	0.96	3

The IAP-MS experiments with the TCF4_01 and TCF4_02 antibodies resulted in a robust enrichment of TCF4 peptides, covering approximately 37% of the full length TCF4-B sequence in both cases (Table 5.6). The majority of the TCF4 peptides analysed by MS were found in bands 3 and 4 for both antibodies (Figure 5.6). Interestingly, two proteins with the highest peptide scores in the TCF4_01 IAP (YLPM1 and LUZP1, band 1) were absent in the TCF4_02 IAP, indicating these were probably cross-reactive proteins. The combined peptide data from all of the excised bands was searched and filtered to generate a list of putative TCF4-interacting proteins as outlined in Figure 5.7. Initially, proteins that were present in both TCF4_01 and TCF4_02 IAPs were exclusively selected to redress the possibility that each TCF4 antibody may have cross-reactive proteins. A threshold protein score of 20 was initially applied to the MS data to remove the lowest confidence proteins from the list. Proteins that had immunoprecipitated non-specifically with the pre-immune IgG IAP were subsequently removed. In the final step, proteins in the bead proteome were removed from the analysis (Table 5.5). The filtering process generated a list of 72 TCF4-associated proteins that spanned an array of biological functions.

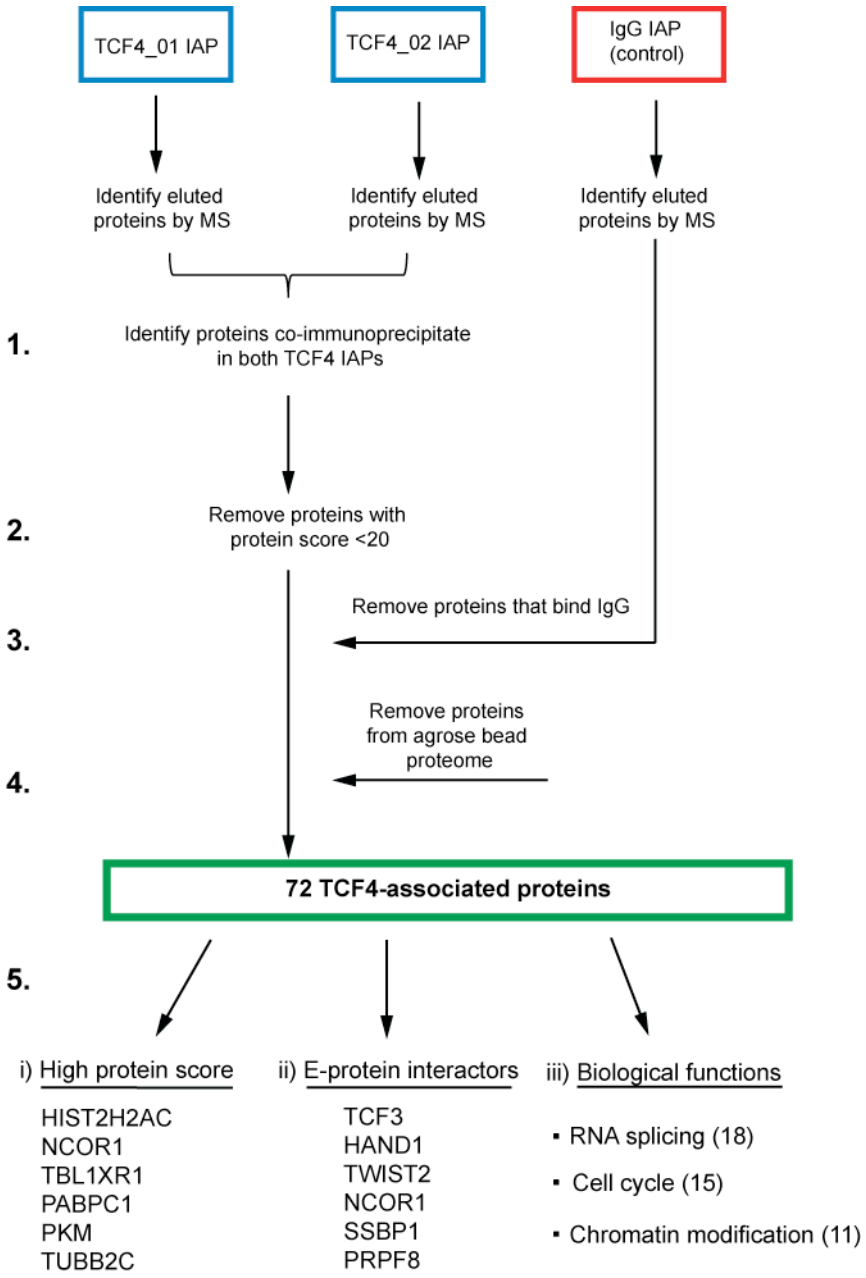


Figure 5.7 Flowchart of the quality control process used to establish potential TCF4-interacting proteins. The schematic illustrates the data processing steps used to generate the list of TCF4-associated interactions presented in Table 5.6. Bands excised from the polyacrylamide gel in Figure 5.6 were analysed by mass spectrometry (MS) and the data was processed in five steps. **1.** Proteins cross-reacting with either antibody were avoided by generating a list of co-immunoprecipitated proteins that appeared in both TCF4_01 and TCF4_02 IAP experiments. **2.** Low confidence protein predictions from MS data were eliminated by applying a threshold protein score of 20. **3.** Proteins that bound non-specifically to pre-immune IgG antibodies in the negative control IAP experiment were removed. **4.** Proteins that precipitated non-specifically onto the agarose bead matrix were removed (Table 5.5). The filtering process created a list of 72 TCF4-associated proteins (Table 5.6). **5.**

Potential TCF4-interacting proteins were annotated in three ways (Section 5.2.5). i) Proteins with high confidence protein prediction scores (score >100) in both IAPs were highlighted. ii) E-protein interaction datasets from the Human Protein Research Database (HPRD) and Biological General Repository for Interaction Datasets (BioGRID) public databases were searched to identify published E-protein interactions. iii) Top biological processes were identified using DAVID v6.7 (BP_FAT) (Table 5.7).

Table 5.6 TCF4-associated proteins determined by mass spectrometry. The proteins in this list were identified in IAP-MS experiments using the TCF4_01 and TCF4_02 antibodies. Probable false-positive interactions were removed using the filtering process outlined in Figure 5.7. The protein score (prot. score), percentage amino acid coverage (% coverage) and number of peptides (# peptides) associated to each co-precipitated protein is displayed in the table for each antibody. The proteins are ranked in descending order according to the score generated from the TCF4_02 IAP. Proteins that have a score >100 for both IAPs are highlighted in orange. These form the list of high confidence interactors. Known E-protein interactions for either TCF3 (E47) or TCF12 (HEB) are marked in bold. Official HUGO gene names of proteins have been used. The indicated protein prediction scores were generated by the Mascot software (Section 2.7.4).

#	Gene	TCF4_01 Prot. score	TCF4_01 % coverage	TCF4_01 # peptides	TCF4_02 Prot. score	TCF4_02 % coverage	TCF4_02 # peptides
1	**TCF4**	**224.65**	**37.85**	**21**	**292.88**	**36.66**	**21**
2	HIST2H2AC	160.81	58.14	5	177.09	58.14	5
3	GPHN	30.38	7.23	4	172.64	30.84	17
4	DOCK7	35.78	3.57	6	144.6	13.95	27
5	**NCOR1**	**136.98**	**9.06**	**17**	**139.56**	**13.44**	**27**
6	DHX9	98.51	15.2	15	131.97	13.94	16
7	TBL1XR1	103.82	27.63	11	128.71	32.68	11
8	PABPC1	108.92	32.46	12	127.51	21.24	7
9	PKM	228.38	34.17	11	119.63	28.57	9
10	OSBPL8	66.06	17	11	109.74	27.56	20
11	TUBB2C	172.39	32.95	11	101.55	27	9
12	HDAC3	39.74	12.4	4	99.39	17.25	5
13	SLC25A13	146.47	16.74	9	93.37	12	7
14	FASN	86.29	5.97	11	92.88	4.9	11
15	H2AFX	36.13	16.78	2	79.6	47.55	4
16	ILF2	129.84	39.2	9	77.39	20.17	5
17	FAM98B	26.68	7.33	1	77.31	8.94	2
18	**SSBP1**	**55.98**	**34.09**	**3**	**77.06**	**37.88**	**4**
19	KIAA1429	58.2	2.96	4	76.31	6.81	10
20	USP7	59.64	8.56	8	69.28	8.29	8
21	**PRPF8**	**101.98**	**5.13**	**13**	**69.24**	**4.01**	**10**
22	NUP205	61.36	2.63	5	64.62	1.64	3
23	ARHGEF2	160.8	19.1	15	62.5	19.42	17
24	**TWIST2**	**31.59**	**13.13**	**2**	**61.7**	**13.13**	**2**
25	PTCD3	31.18	3.48	2	60.28	17.79	3
26	SYNCRIP	84.81	23.02	9	59.87	11.03	4
27	LGALS1	111.07	60	6	59.82	40.3	4
28	SNRPD3	32.33	15.83	2	59.67	33.33	3
29	NAGK	89.22	21.8	7	58.38	23.84	6
30	BUB3	50.68	23.15	7	56.69	16.05	5

31	TMEM33	58.93	4.86	1	56.6	4.86	1
32	HNRNPD	77.2	22.33	4	52.22	10.23	2
33	FARSA	62.46	5.24	2	52.2	5.24	2
34	TJP1	52.91	4.38	6	50.59	8.45	13
35	SNRPD2P1	59.81	29.66	4	48.67	46.61	6
36	TIFA	56.31	29.89	2	47.96	45.98	3
37	PRDX2	33.98	21.31	5	46.71	31.15	6
38	PSMD11	49.59	10.19	4	45.66	11.15	3
39	CRMP1	341.56	30.74	13	45.26	8.91	4
40	SSBP2	50.41	10.42	1	45.26	10.42	1
41	MTPN	39.95	14.41	1	44.72	14.41	1
42	RTL1	97.53	5.89	8	44.65	4.49	6
43	**TCF3**	**33.32**	**8.33**	**1**	**44.54**	**8.33**	**1**
44	NPM1P42	35.73	10.57	3	42.11	31.32	5
45	SESTD1	30.47	4.24	1	41.98	7.61	5
46	EIF4A1	57.94	10.04	2	41.75	10.04	2
47	DHX15	54.4	5.41	2	41.74	7.27	1
48	TRA2B	47.03	21.38	6	41.51	25	4
49	HADHB	59.17	8.85	3	40.99	5.31	2
50	PPP2R1A	72.4	15.81	7	40.25	5.17	1
51	**HAND2**	**55.86**	**25.37**	**3**	**39.19**	**8.96**	1
52	SRSF7	35.44	13.86	1	39.19	20.79	2
53	SEC23B	28.95	5.26	3	37.89	5.93	3
54	PPP2R4	23.59	5.83	2	37.84	10.25	2
55	DDX1	42.93	10.13	5	35.98	8.36	4
56	CENPV	66.61	13.55	2	35.18	4.8	1
57	EXOSC9	40.42	5.92	2	34.17	2.82	1
58	FLOT1	39.4	4.35	1	34.09	4.64	1
59	CCT4	75.44	11.59	4	33.43	5.4	2
60	RPL23	52.38	20.9	2	31.8	39.55	4
61	MCCC2	69.64	12.08	6	30.37	3.01	1
62	CHTF8	32.38	5.73	3	28.89	5.73	3
63	MYEF2	72.18	6.22	3	27.44	4.45	1
64	DCAF7	109.93	21.64	6	27.32	10.23	3
65	CAD	92.89	3.21	6	26.02	1.08	2
66	MCM7	322.11	42.56	24	25.92	5.56	3
67	PSMD14	56.05	6.45	2	25.64	14.84	3
68	MCM4	541.51	44.59	32	25.33	2.92	2
69	KPRP	32.88	6.22	4	25.25	1.73	1
70	ARID1A	236.86	14.22	27	24.93	2.55	4
71	PCMTD2	59.92	10.53	4	24.84	9.57	3
72	BLVRA	37.3	6.16	2	22.66	3.42	1
73	POLDIP2	20.77	12.63	3	37.1	10.29	1

5.2.5 Functional annotation of TCF4-associated proteins

To establish whether any of the TCF4-associated proteins identified by IAP-MS were known to interact with E-proteins, the list from Table 5.6 was compared to a combined set of E-protein (TCF4, TCF3, TCF12) interactions obtained from the Human Protein Reference Database (HPRD), and the Biological General Repository for Interaction Datasets (BioGRID) database. This analysis revealed that several of the proteins identified by IAP-MS were reported to interact with TCF3 (E47). The TCF3-interacting proteins in the list included the bHLH proteins TCF3 (self association), HAND2 and TWIST2, the DNA binding protein SSBP1, the splicing factor PRPF8 and the co-repressor NCOR1 (El Ghouzzi et al., 2000; Murakami et al., 2004; Murayama et al., 2004; Teachenor et al., 2012). In addition, HAND2 was already reported to interact with TCF4 (Jogi et al., 2002; Murakami et al., 2004). Each of these proteins is highlighted in bold in Table 5.6.

TCF4-associated proteins identified by IAP-MS were functionally annotated using biological process annotations in DAVID v 6.7 (Section 2.10.1). Three distinct biological processes were identified from these annotations; RNA splicing, chromatin modification and the cell cycle (Table 5.7).

Table 5.7 Biological process annotations for TCF4-associated proteins identified by mass spectrometry. The top biological process annotations related to proteins identified Table 5.6. Biological process annotations (BP_FAT) were determined in DAVID. The percentage of genes in each annotation is relative to the total number of genes in Table 5.6 (72 genes).

Function	%	Genes
RNA splicing	18	DDX1, DHX9, DHX15, HNRNPD, PPP2R1A, SRSF7, TRA2B, SNRPD3, SYNCRIP, PRPF8, KIAA1429, SNRPD2P1, PABPC1
Cell cycle	17	CHTF8, H2AFX, ARHGEF2, CENPV, HDAC3, MCM4, MCM7, NCOR1, NPM1P42, PSMD11, PSMD14, TCF3
Chromatin organization	11	H2AFX, ARID1A, HIST2H2AC, HDAC3, NCOR1, TBL1XR1, CENPV, NPM1P42

Among the RNA splicing proteins are two members of the spliceosome; the catalytic RNA-protein complex responsible for pre-mRNA splicing. DEAH (Asp-Glu-Ala-His) box polypeptide 15 (DHX15) is an RNA helicase-like protein and is involved in

spliceosome disassembly (Arenas and Abelson, 1997). The pre-mRNA processing factor 8 homolog (PRPF8) is a protein in the catalytic core of the spliceosome that participates in assembling the catalytic RNAs for splicing events (Galej et al., 2013). Interestingly, PRPF8 was also co-immunoprecipitated with E47 (TCF3) in a similar IAP-MS experiment performed in HEK cells (Teachenor et al., 2012).

The chromatin organization annotation contains histones (H2AFX, HIST2H2AC) and chromatin modifiers such as AT-rich interactive domain-containing protein 1A (ARID1A), histone deacetylase 3 (HDAC3), nuclear receptor co-repressor 1 (NCOR1) and transducin beta-like 1 X-linked receptor 1 (TBL1XR1). ARID1A forms an essential part of the SWI/SNF (SWItch/Sucrose NonFermentable) chromatin-remodeling complex; an ATP-dependent nuclear protein complex that rearranges nucleosomes to modulate transcription. Remarkably, NCOR1, HDAC3 and TBL1XR1 were all co-purified and together form the nuclear co-repressor complex, functioning with transducin beta-like 1 (TBL1) to repress gene expression (Yoon et al., 2003). NCOR1 and TBL1XR1 are two of the few co-precipitated proteins with high scores in both IAP experiments. NCOR1 can interact with the TCF3 (E47), providing more support for this repressor as a novel TCF4-interacting protein (Murayama et al., 2004). Taken together my data suggests that NCOR1 and TBL1XR1 are promising TCF4-interacting proteins to investigate. The peptide data for these proteins is provided in Table 5.8.

The cell cycle category overlaps with the chromatin organization category for certain terms such as the centromeric protein V (CENPV), as it forms a structural part of the kinetochore and aids chromosomal alignment during mitosis (Tadeu et al., 2008). Unique terms in the cell cycle category include two members of the minichromosome maintenance complex (MCM4 and MCM7) responsible for orchestrating the precise replication of DNA and two regulatory subunits of the 26S non-ATPase proteasome (PSMD11, PSMD14) that both play a key role in maintaining the high proteosomal activity required for stem cell pluripotency (Blow and Dutta, 2005; Buckley et al., 2012; Vilchez et al., 2012).

Table 5.8 Mass spectrometry data for the TCF4-associated proteins NCOR1 and TBL1XR1. The table shows the peptide sequences of NCoR complex members NCOR1 and TBL1XR1 derived from bands 1 and 4 in Figure 5.6. High confidence peptides are presented for each IAP together with the protein scores and % coverage values generated from the MS data.

Protein	Peptide sequences	Parameters
NCOR1 2440 amino acids 270kDa	ALDPAAAAYLFQR SIVQIIYDENR SPGSISYLPSFFTK NFGLIASYLER TTITAANFIDVIITR GTAGAIQEGSITR YNTAADALAALVDAAASAPQmDVSK ADSVDVEVR GHSFADPASNLGLEDIIR QDSLLLLSQR ISVESIPSLR GMPPLEIVPENIK	**TCF4_01 IAP** Protein score: 136.98 Coverage: 9.06% Source: band 1B
	NFGLIASYLER GTAGAIQEGSITR TVLSGSIMQGTPR ALDPAAAAYLFQR SAAVSEQQQLEQK ESPVSAPLEGLIcR AQLSPGIYDDTSAR SIVQIIYDENR ADSVDVEVR QEQIDLECR EELIQSMDR	**TCF4_02 IAP** Protein score: 139.56 Coverage: 13.44% Source: band 1C
TBL1XR1 514 amino acids 55.5 kDa	IWNLSENSTSGSTQLVLR LAQQQAAAAAAAAAAASQQGSAK GTGGIFEVcWNAAGDK DKLAQQQAAAAAAAAAAASQQGSAK GNFILSAGVDK VGASASDGSVcVLDLR	**TCF4_01 IAP** Protein score: 103.82 Coverage: 27.63% Source: band 4B
	IWNLSENSTSGSTQLVLR GNFILSAGVDK VGASASDGSVcVLDLR LAQQQAAAAAAAAAAASQQGSAK GTGGIFEVcWNAAGDK SISSDEVNFLVYR WDPTGNLLAScSDDmTLK	**TCF4_02 IAP** Protein score: 128.71 Coverage: 32.68% Source: band 4C

5.3 Discussion

Multiple alternatively spliced TCF4 transcripts have been described in humans however our understanding of the proteins they encode is very limited. In addition, the absence of specific TCF4 antibodies has constrained the study of TCF4 in human cells and tissues. To further characterise the TCF4 protein, polyclonal antibodies were generated to study TCF4 at the proteomic level. TCF4 antibodies were designed to target a shared epitope present in all known TCF4 isoforms that were subsequently characterised to validate their efficacy. The antibodies were used in IAP-MS experiments that identified distinct TCF4 isoforms in SH-SY5Y cells. The peptide data generated by IAP-MS was used to determine TCF4 phosphorylation sites and potentially novel TCF4-interacting proteins that are involved in chromatin modification, splicing and cell cycle regulation.

5.3.1 TCF4 polyclonal antibodies detect short and long TCF4 isoforms in SH-SY5Y cells and can be used in multiple applications

TCF4 polyclonal antibodies were initially designed in order to immunoprecipitate endogenous TCF4 from SH-SY5Y cells. However, to ensure the quality of the antibodies, the antibodies were subjected to rigorous quality control experiments to assess their specificity and efficacy in different experiments. TCF4_01and TCF4_02 were able to detect natively folded and denatured protein in various experimental contexts. The first set of experiments, using western blots, demonstrated the presence of two distinct TCF4 protein isoforms in SH-SY5Y cells that were highly reduced after TCF4 knockdown (Figure 5.2A). E-proteins (TCF4, TCF3 and TCF12) share a high degree of sequence similarity that contributes to the difficulty in studying each protein individually. Importantly, these antibodies did not cross-react with over-expressed TCF3 on western blots illustrating the sequence-specificity of TCF4 antibodies when detecting denatured TCF4 epitopes (Figure 5.2B). The utility of these antibodies for detecting native epitopes was also demonstrated in the immunocytochemistry experiments presented in Chapter 4 (Figure 4.2). Finally the antibodies proved to be efficacious in immunoprecipitation experiments. Both TCF4_01 and TCF4_02 antibodies were capable of depleting TCF4 from cellular extracts of SH-SY5Y, enabling TCF4 to be immunoaffinity purified for downstream MS (Figure 5.3)

5.3.2 Specific TCF4 isoforms are detected in SH-SY5Y cells

To precisely identify the TCF4 isoforms in SH-SY5Y cells, the TCF4 protein was immunoaffinity purified and analyzed by MS. Peptides generated from this analysis revealed a strong enrichment of TCF4, identifying a total of 497 TCF4 peptides across the 7 bands that were sent for MS. The band with the highest enrichment of TCF4 was band 4 (Figure 5.4, 80kDa), with peptides covering approximately 70% full length TCF4-B amino acid sequence. The peptides in this data set were examined for the presence of isoform-specific amino acid sequences. The diversity of TCF4 isoforms was investigated by exploring the peptide data for unique isoform-specific sequences.

Unique peptides corresponding to TCF4-A (NCBI isoform j/k) were readily identified (Table 5.2). In addition, a single peptide for isoform l was identified among the 497 TCF4 peptides generated by the IAP-MS experiment (Table 5.2). The peptide data suggested that although isoform l and TCF4-A are present in SH-SY5Y cells, TCF4-A seemed to be the predominant short isoform. TCF4-A and TCF4 isoform l are both shorter TCF4 isoforms lacking the N-terminal sequences containing the AD1 domain and the NLS and have approximate molecular weight of 54kDa. Numerous peptides corresponding to TCF4-B (approximately 71kDa) were also identified (Figure 5.2). Form these experiments it can be concluded that SH-SY5Y cells synthesise both TCF4-A and TCF4-B isoforms.

Additionally, TCF4 peptides were explored for the presence of the RSRS insertion sequence. This four amino acid insertion is located in the core region shared by all TCF4 isoforms, positioned N-terminal to bHLH domain. Although very little is known about the function of this insertion sequence, it has been proposed to affect the DNA binding properties of the bHLH domain, allowing TCF4 to bind imperfect E-boxes (Liu et al., 1998). The peptide data contained evidence that certain isoforms contained the insertion whilst others did not. Both + and - isoforms were relatively abundant in SH-SY5Y cells as determined from the unique peptide counts relating to each isoform (Table 5.2). Analysis of TCF4 mRNA has previously shown that both of these isoforms are abundant in human tissues (Sepp et al., 2011). Interestingly, the - isoform was accompanied by a potential PKA/PKC phosphorylation site (T545)

located one amino acid proximally to the RSRS insertion site. This site was not apparently phosphorylated in the + isoform, indicating that it may be isoform-specific.

5.3.3 TCF4 phosphorylation

MS data also provides potentially important information about post-translational modifications in proteins. Accordingly, the peptide data was searched for TCF4 phosphopeptides (Table 5.3). My data suggest that many of the N-terminal serines in the vicinity of AD1 (S27, S66, S92, S140) and those surrounding AD2 (S351, S372) can be phosphorylated by large range of potential kinases including glycogen synthase kinase 3 (GSK3B), members of the mitogen-activated protein kinase signaling pathway (ERK1, ERK2, MAPKAPK2) and several of the cyclin-dependent kinases (CDK1, CDK2, CDK4, CDK5, CDK6). GSK3B is an important regulator of neural development and is required for coordinated neurogenesis, neuronal polarization and axon growth (Hur and Zhou, 2010). MAPK signaling can be activated by a variety of signals including growth factor stimulation and is essential for controlling proliferation, survival, and differentiation (Kim and Choi, 2010). Cyclin-dependent kinases work in concert with cyclins and cyclin-dependent kinase inhibitors (CDKIs) to harmonise progression through the cell cycle. Emerging evidence also suggests CDKs may have neuronal-specific functions such as neuronal migration, axon guidance, and synaptic transmission (Lim and Kaldis, 2013).

In addition, some of the phosphorylated residues seem to be near more distinctive motifs with perhaps more specialised functions. S318 is a motif recognised and phosphorylated by the catalytic subunit of the DNA-dependent protein kinase (PRKDC) and ataxia telangiectasia mutated (ATM) that both have roles in double-strand break repair, an essential process for DNA damage response and normal development (O'Driscoll and Jeggo, 2006). Interestingly, T545, specific to the - isoform, is potentially a target for cAMP-dependent protein kinase alpha (PKA) and the calcium-sensitive protein kinase C (PKC). Both of PKA and PKC have been shown to regulate signaling events during long-term potentiation (LTP) and long-term depression (LTD), two key mechanism underlying synaptic plasticity (Rebola et al., 2010). Although empirical validation will be required to support these *in silico*

predictions, this analysis may be useful to create hypotheses about upstream regulators of TCF4 function in neuronal cells.

5.3.4 Co-immunopecipitation of bHLH proteins

Three bHLH proteins were identified in the IAP-MS experiments providing support for the validity the data. The E-protein TCF3 (E47) has high degree of sequence similarity to TCF4 raising the possibility that it may be able to form heterodimers with TCF4. This interaction has also been proposed from a large-scale yeast two-hybrid screen of transcription factors (Ravasi et al., 2010). HAND2 (dHAND) and TWIST2 (Dermo-1) are two related bHLH proteins that have both been shown to interact with E-proteins. HAND2 can specifically interact with TCF4 in mammalian two-hybrid assays and in GST pull down experiments whilst TCF3 interacts with TWIST2 in yeast two-hybrid screens (Li et al., 1995; Dai and Cserjesi, 2002; Jogi et al., 2002; Murakami et al., 2004). TWIST2 is regulator of epithelial-to-mesenchymal transition and causes Setleis syndrome, a rare inherited form of ectodermal dysplasia (Tukel et al., 2010; Fang et al., 2011). HAND2 is involved in terminal differential sympathetic neurons but its interaction with TWIST1 contributes to pathogenic mechanisms in Saethre–Chotzen syndrome, another developmental disorder associated with craniosynostosis, craniofacial and limb dysplasias (Howard et al., 1997; Firulli et al., 2005).

5.3.5 Co-immunoprecipitation suggests a role for TCF4 in chromatin organization, splicing and cell cycle regulation

The repertoire of previously validated TCF4 interactions is essentially limited to a range of bHLH transcription factors that modulate TCF4 function (Sections 1.5 and 1.7.2). In an attempt to identify new protein interactions and further characterise the function of TCF4 in neuronal cells, IAP-MS experiments were used to identify proteins that co-immunoprecipitated with TCF4 in SH-SY5Y cells (Figure 5.7). Despite the filtering process, the interactions identified remain putative until they have been validated using independent methods (Table 5.6). Nevertheless, several potential interactions that were discovered in these experiments merit further attention.

A subset of the co-immunoprecipitated proteins had a functional role in chromatin organization; a core cellular process required for dynamic control of transcription, cell division and development. This category included chromatin modifiers such as ARID1A and members of the NCoR complex. Chromatin modifiers are crucial during neural development as they can activate or repress gene expression by rearranging chromatin structure (Hirabayashi and Gotoh, 2010). The importance of this role has been highlighted in human sequencing studies that have associated several neurological and psychiatric diseases to chromatin regulators (Ronan et al., 2013).

ARID1A is part of the SWI/SNF family of chromatin remodelers that causes Coffin-Sirius syndrome (Tsurusaki et al., 2012). Interestingly, multiple members of the SWI/SNF remodeling complex were co-immunoprecipitated individually by the TCF4_01 (SMARCA4, SMARCB1, SMARCC1, SMARCD1, SMARCD2, SMARCD3) and TCF4_02 (SMARCC2) antibodies, although none of these subunits were found in both IAP experiments (data not shown). bHLH proteins may require interactions with members of the SWI/SNF remodeling complex to activate gene expression programs. The proneural bHLH proteins NEUROG1 and NEUROD1 are strong inducers of the neuronal fate however they fail to promote neuronal differentiation when the function of SWI/SNF remodeling complex protein SMARCA4 (Brg1) is abolished (Seo et al., 2005). Another bHLH protein Olig2, required of oligodendrocyte differentiation, has been shown to recruit SMARCA4 at distinct sets of myelin regulatory genes to drive fate specification in oligodendrocyte precursors. Accordingly, TCF4 may also function in combination with chromatin remodeling complexes to regulate its coordinated transcriptional program.

NCOR1 is another chromatin modifier that co-immunoprecipitated with TCF4 in the IAP-MS experiments. NCOR1 participates in transcriptional repression, an important mechanism in diverse biological functions. The NCOR1 protein can be recruited by many classes of transcription factors including bHLH proteins and is often associated with large protein complexes (Jepsen and Rosenfeld, 2002). The majority of the transcriptional repression mediated by NCoR (nuclear co-repressor) complexes is conferred through histone deacetylase activity such as that of HDAC3, a co-precipitated protein in this experiment (Wen et al., 2000). The NCoR complex can exist in several forms although some of the core interactions include NCOR1, TBL1,

TBL1XR1 and HDAC3 acting together to enforce transcriptional repression (Yoon et al., 2003). Strikingly, NCOR1, TBL1XR1 and HDAC3 were all co-immunoprecipitated with TCF4 in these experiments (Table 5.6, Table 5.8). NCOR1 provides a functional scaffold for recruitment of the complex and acts as a co-factor promoting histone deactylase activity (Guenther et al., 2001). NCOR1 and TBLXLR1 were two of the more prominent proteins identified by the proteomic experiments as a high score was associated to each of the TCF4 IAP experiments (Table 5.6, Table 5.8). In addition to TBL1, TBL1XR1 and HDAC3, some of the characterised NCoR complexes contain multiple components of the mammalian SWI/SNF complex (SMARCC1, SMARCC2, SMARCA4) (Underhill et al., 2000).

NCOR1 is an interesting and plausible candidate TCF4-interacting protein because it is already implicated in bHLH transcription factor function. The NCOR1 protein can associate with the E-protein TCF3 as part of a vitamin D-dependent nuclear receptor co-repressor complex in kidney cells (Murayama et al., 2004). The presence of the vitamin D ligand causes the dissociation of the p300 co-activator from TCF3, and the association of a co-repressor complex including HDAC2 and NCOR1. Co-factor switching may be a crucial element controlling of E-protein activity in neuronal cells.

TCF4-associated proteins in the IAP-MS experiments also suggest a role for TCF4 in the cell cycle. Here, a selection of proteins with multiple functions was identified that can each contribute to aspects cell cycle regulation (Table 5.7). Control of the cell cycle has a direct effect on the balance between proliferation and neurogenesis in the brain (Ohnuma and Harris, 2003). Interestingly, the NCOR1/TBL1/HDAC3 complex may also have a role independent of transcriptional regulation but directly related to the cell cycle (Ishii et al., 2008). Co-localization experiments have demonstrated that the NCoR complex localises to the mitotic spindle during mitosis. Follow up experiments illustrated the necessity of the complex for cytokinesis, as the mitotic spindle collapsed upon knockdown of either NCOR1 or HDAC3. The centromeric protein CENPV is also associated to the kinetochore during spindle formation and is co-immunoprecipitated in the proteomic experiments suggesting that TCF4 may also play a structural role during cell division (Tadeu et al., 2008).

Finally, RNA splicing is another core biological process that may involve TCF4 through associations with different proteins. Recent evidence suggests that splicing occurs co-transcriptionally in the nucleus and therefore the splicing machinery is in close association with chromatin (Brown et al., 2012). It may therefore be possible that TCF4 interacts with certain splicing factors although this has never been demonstrated for E-proteins. The lack of evidence for bHLH factors being involved in splicing demands caution in the interpretation of these data. Nevertheless, the TCF4 homologue TCF3 was also associated with splicing factor PRPF8 in IAP-MS experiment carried out in HEK293T cells (Teachenor et al., 2012).

In summary, the proteomic analysis of TCF4 has provided valuable information to understand the context in which this transcription factor may operate. The expression of long and short isoforms in SH-SY5Y cells suggests that both proteins may be important mediators of TCF4 function in neurons. The identification of TCF4 phosphorylation sites will be useful in predicting upstream modulators of TCF4 function and their related signaling pathways. Furthermore, the interaction data generated by IAP-MS positions TCF4 in a potential biological framework intersecting with transcriptional regulation, RNA splicing and cell cycle regulation. However, the proteomic data is preliminary and further work will be required to confirm the validity and context of these interactions.

Chapter 6

TCF4 Isoform-Specific Gene Expression Profiling

6.1 Introduction

TCF4-A and TCF4-B are the two main protein isoforms that are encoded by the *TCF4* gene. Most of the current knowledge concerning the functions of these isoforms is based on transient transfection experiments in mammalian cells. Initial experiments were aimed at determining the transcriptional activity of these isoforms on specific promoters. TCF4-B and TCF4-A isoforms were initially reported to have different transcriptional activities on specific promoters *in vitro* (Skerjanc et al., 1996; Petropoulos and Skerjanc, 2000). In these studies, TCF4-B was described as a transcriptional inhibitor whilst the shorter isoforms TCF4-A was found to be a mild transcriptional activator. The difference in isoform-specific activities is likely to be mediated by differences in their domains architecture (Figure 6.1).

However, it has also been demonstrated that TCF4-B can activate transcription in certain cell lines and on specific promoter constructs (Pscherer et al., 1996; Chen and Lim, 1997; de Pontual et al., 2009; Forrest et al., 2012). The function of TCF4-B is therefore highly context dependent as described in chapter 3. The reason for this dichotomy may reside in its interaction with transcriptional regulators through AD1 and AD2 (Section 1.4). The spatio-temporal availability of co-factors during development may be crucial for determining TCF4's transcription program.

Interestingly, although TCF4-B and TCF4-A are reported to have different transcriptional activities, their physiological effects on myogenesis and EMT are reported to be similar (Petropoulos and Skerjanc, 2000; Sobrado et al., 2009). Transfection of MyoD with either TCF4-B or TCF4-A converts fibroblasts to myoblasts with the same apparent efficiency (Petropoulos and Skerjanc, 2000).

Similarly, TCF4-B and TCF4-A isoforms induce full EMT in epithelial MDCK cells stably over-expressing either isoform (Sobrado et al., 2009).

To date, the isoform-specific effects on transcription in neuronal cells remain uncharacterised. Having established that SH-SY5Y cells express long and short TCF4 isoforms, RNAi was used to knockdown the TCF4-B and TCF4-A isoforms individually in order to identify isoform-specific gene expression programs. Using the experimental design outlined in the Chapter 4, selected gene expression changes relating to each isoform were validated by qPCR and the effects of each isoform on the cellular transcriptome were determined using gene enrichment analysis.

6.2 Results

6.2.1 Knockdown of TCF4 isoforms for microarray analysis

To design isoform specific siRNAs, careful inspection of the unique nucleotide sequence between the long and short isoforms of TCF4 was required. TCF4-B has an additional 183 amino acids in its N-terminus compared to TCF4-A (Figure 6.1). The nucleotide sequence encoding this region was used to design two siRNAs that would specifically target TCF4-B (KDB1 and KDB2) (Figure 6.1). TCF4-A shares 488 of its 511 amino acids with TCF4-B, however it contains a unique 5' coding exon that is only found in this isoform (Figure 6.1). Consequently, this unique sequence was used to design two non-overlapping TCF4-A specific siRNAs (KDA1 and KDA2) (Figure 6.1). Due to the limited length of the unique sequence in TCF4-A, the siRNA sequences tended to have lower predicted targeting efficacies compared to the TCF4-B siRNA sequences (data not shown). Each of the four duplexes was transfected separately in SH-SY5Y cells over a period of 48h alongside the same anti-GAPDH siRNA used in Chapter 4 (KDGAP) and a mock transfection control (mock). Five technical replicates of each condition were prepared; four were used for RNA extraction while protein was prepared from the fifth.

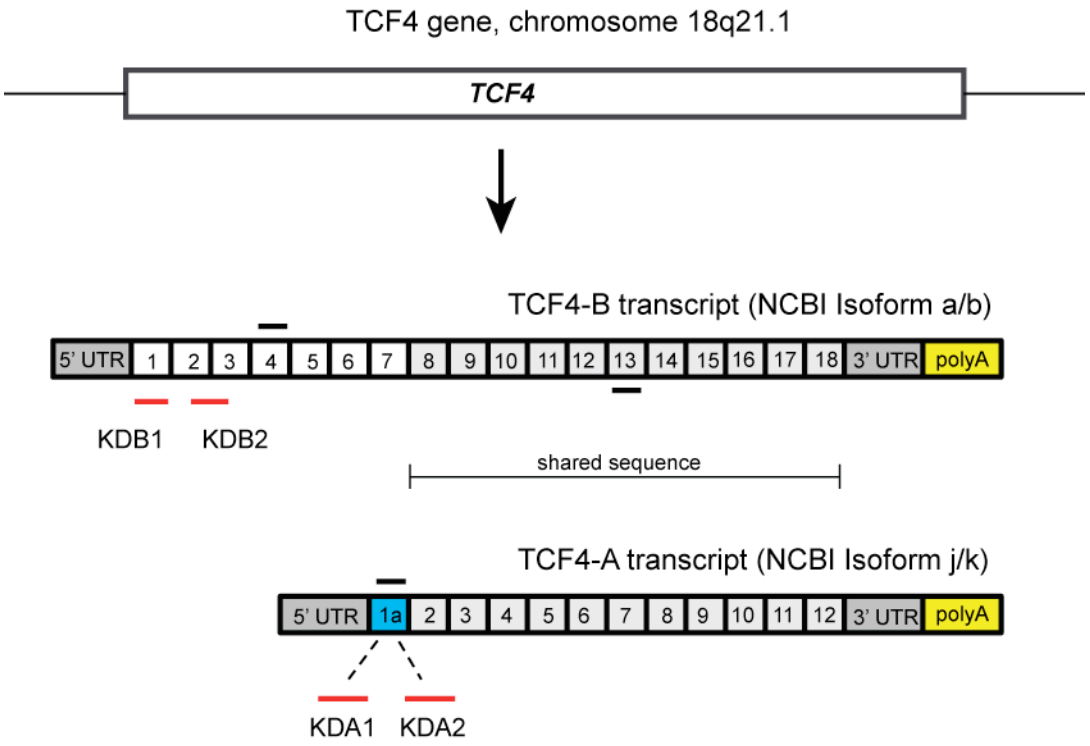


Figure 6.1 **TCF4 isoforms and siRNA recognition sites**. This illustration shows the transcript structure for the TCF4-B and TCF4-A isoforms. The approximate location of the recognition sequences for each of the four siRNA duplexes is indicated (red bars, KDB1; TCF4-B exon 1, KDB2; TCF4-B exon 2-3, KDA1; TCF4-A exon 1, KDA2; TCF4-A exon 1). Importantly, The KDA1 and KDA2 sequences do not overlap. The location of primers used for mRNA quantification of TCF4-B transcripts (exon 4), TCF4-A transcripts (unique exon 1a) and total TCF4 transcripts (shared exon 13) are indicated (black bars). The coding exons are numbered separately for each isoform (TCF4-B; 1-18, TCF4-A; 1-12). The 11 C-terminal coding exons are identical in the two isoforms (light grey, TCF4-B; 8-18, TCF4-A; 2-12). Note that TCF4-A contains a unique 5' exon (exon 1a, blue). Illustration not to scale.

The efficacy of each knockdown was evaluated using qPCR and semi-quantitative western blotting (Figure 6.2). The western blot demonstrated the isoform-specific siRNAs successfully knocked down the long and short TCF4 isoforms, validating the siRNA design (Figure 6.2A). Importantly, mock and KDGAP treatments did not have any appreciable effect on TCF4 protein levels. The knockdown efficacy for both siRNAs directed against the TCF4-B (72kDa) was similar, reducing the TCF4 isoform to approximately 30% of mock transfections (Figure 6.2B, left). However, the knockdown efficiency for each siRNA designed against TCF4-A was not the same; KDA1 reduced TCF4-A abundance to approximately 30% of mock whilst KDA2 reduced abundance to only 50% (Figure 6.2B, middle). Total TCF4 protein levels were also measured and ranged from approximately 50% (KDB2) to approximately 75% (KDA2) of mock (Figure 6.2B, right). KDA1 and KDB1 had an intermediate amount of total TCF4 protein, at approximately 65% of mock. The qPCR experiments closely reflected the knockdown efficiencies observed at the protein level (Figure 6.2C). Interestingly, both the mRNA and protein quantification detected a small isoform-specific compensatory mechanism; knockdown of the short isoform seemed to slightly increase the abundance of the long isoform (Figure 6.2B and C). In contrast, knockdown of the long isoform appeared to moderately decrease short isoform abundance (Figure 6.2B and C).

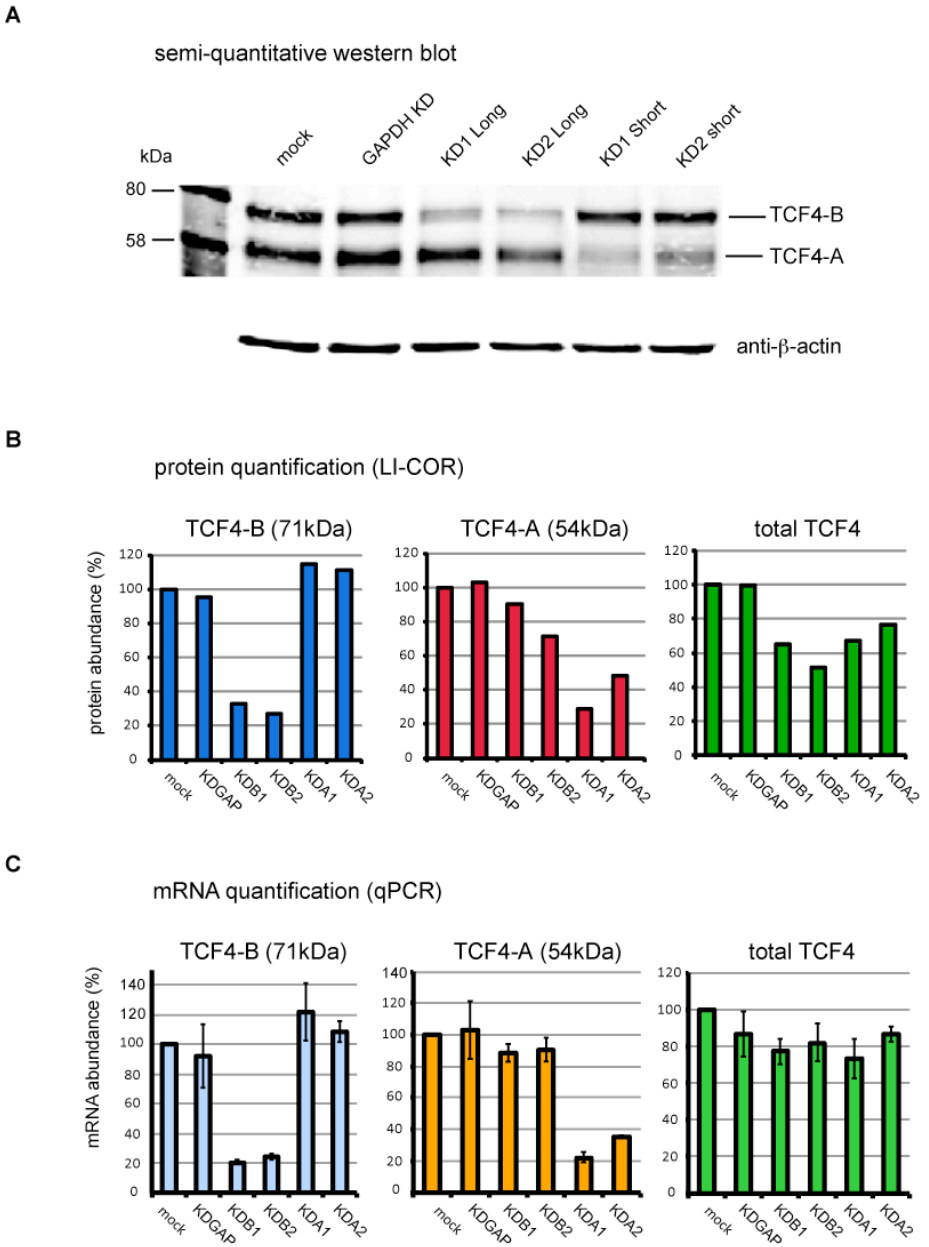


Figure 6.2 siRNA-mediated knockdown of *TCF4* isoforms in SH-SY5Y cells. SH-SY5Y cells were transfected with siRNA oligonucleotides targeting TCF4-B (KDB1 and KDB2), TCF4-A (KDA1 and KDA2) or GAPDH (KDGAP). After 48h, RNA and protein were extracted to assess the knockdown efficiency. (A) Western blots of protein lysates prepared from siRNA-treated SH-SY5Y cells demonstrated that TCF4 could be knocked down in an isoform-specific manner. α-tubulin was used as a loading control and for normalisation. (B) LI-COR quantitation of TCF4 protein levels in siRNA-treated cells. The levels of TCF4-B (71kDa), TCF4-A (54kDa) and total TCF4 (TCF4-A and TCF4-B) in each experiment were quantified and normalised to α-tubulin. (C) Primers complimentary to specific coding exons in TCF4-B (TCF4-B, exon 4), TCF4-A (TCF4-A, exon 1a) and a constitutive exon present in all TCF4 isoforms (exon 13) were used to measure transcript abundance by qPCR in four biological replicates of each group (primer binding sites are shown in Figure 6.1).

6.2.2 Microarray quality control and data processing

To examine the genome wide expression changes in an isoform-specific manner, the RNA samples described above were also used for microarray analysis, as a reproducible knockdown had been measured for each isoform. RNA quality was assessed with an Agilent Bioanalyzer and was considered of high quality for all samples (RIN 8.5-10) except one sample that was of lower quality (KDB2 replicate 1, RIN 5.2). The RNA from each condition was converted to cDNA, fragmented, amplified and labeled by CBS for hybridisation to Affymetrix Human Gene 2.0 ST arrays (Section 2.9.10). During this procedure, one of the KDGAP replicates failed quality control and had to be removed from the downstream analysis, reducing the number of replicates for this group to 3. The rest of microarray data was received in the form of .CEL files that were imported into Partek genomics suite 6.6 and RMA normalised. Principal component analysis (PCA) was subsequently performed using Partek to assess the treatment effects (Figure 6.3.1). Mapping of the three principal components of the microarray experiment explained approximately 39% of the data and showed successful grouping of the replicates in each of the six treatment groups (Figure 6.3.1). The long isoform knockdown groups (KDB1, dark blue; KDB2, light blue) and one of the short isoform (KDA2, light green) treatment groups had the least amount of variability as determined by the close clustering of individual replicates (Figure 6.3.1). The remaining groups were more variable but most still clustered distinctly in the three-dimensional space of the PCA. The KDGAP group contained one sample that appeared to be an outlier as it was neighbouring two replicates of the KDB1 group and did not cluster closely to the other KDGAP replicates. However, performing the downstream analysis with and without the sample did not have a substantial effect on the results and consequently the sample was included in the reported data (data not shown).

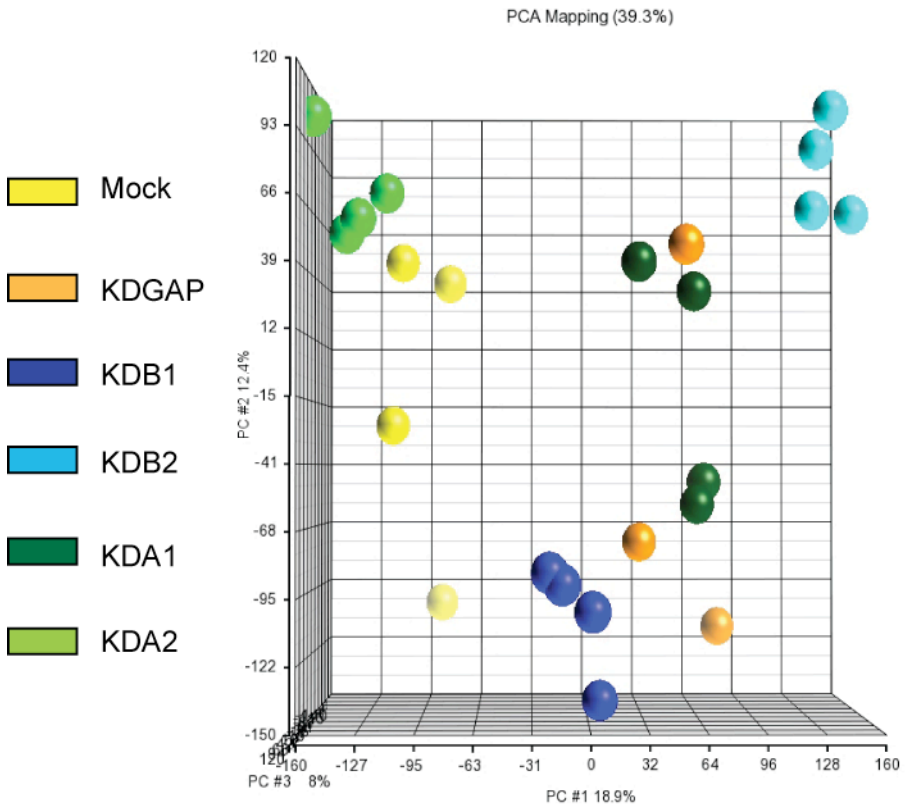


Figure 6.3.1 Principle component analysis of the isoform-specific microarray. The genome-wide expression data from each sample is reduced to three principle components that capture 39.3% of the variation in the data. The principle components (PC) are plotted against 3 axes: PC1; X-axis, PC2; Y-axis, PC3; Z-axis. The percentage variance captured by each PC is displayed next to each axis label (PC1; 18.9%, PC2; 12.4%, PC3; 8%). For clarity, the treatment groups have been colour-coded (see key to the left of the graph).

According to the PCA, the second siRNA targeting the short TCF4-A isoform (KDA2) clustered very near the mock treated samples indicating that this knockdown may not have caused many gene expression changes. Accordingly, these samples also had the highest load TCF4 protein (75% remaining) and exhibited the weakest isoform-specific knockdown, with an approximately 50% reduction in the short isoform compared to mock (Figure 6.2). The KDA2 samples were subsequently excluded from the downstream analysis due to lack of significant gene expression changes and concordance with KDA1 (see below). This observation could be a testament to the importance of dosage in TCF4 function. The poor knockdown with KDA2 duplex is likely to reflect technical limitations in the design of RNAi duplexes targeting very short sequences. Finally, the KDB2 sample of lower quality identified

by the Agilent Bioanalyzer (RIN 5.2) clustered similarly to the other samples of the group and was therefore incorporated into the ensuing analysis.

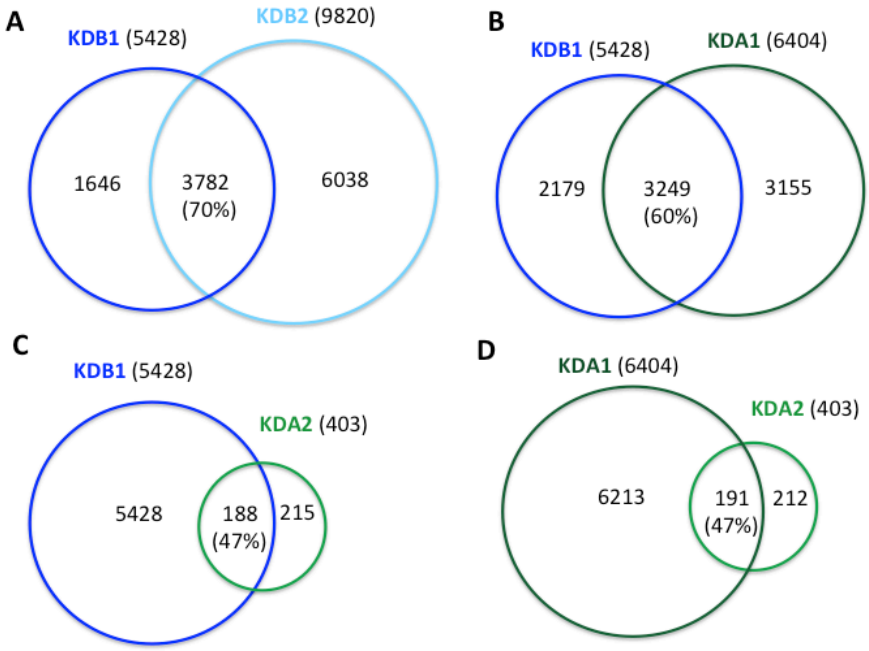


Figure 6.3.2. Comparison of differentially expressed transcripts in TCF4-knockdown treatment groups. The microarray data was used to generate a list of differentially expressed transcripts from each TCF4 knockdown treatment group. Each list was generated by performing a one-way ANOVA between the TCF4 knockdown group (KDB1, KDB2, KDA1 or KDA2) and the mock treated group. Individual lists were compared to determine the number of shared gene expression changes. The number of differentially expressed transcripts associated to each treatment and each overlap is indicated. Blue circles are used for long isoform knockdown experiments (KDB1, KDB2), green circles are used for short isoform knockdown experiments (KDA1, KDA2). The size of each circle is representative of the number of gene expression changes but is not to scale.

After PCA analysis, the transcriptional profiles of different TCF4 knockdown groups were compared to assess the degree similarity between them (Figure 6.3.2). As described in Chapter 4 (section 4.2.2), a one-way ANOVA was performed between individual TCF4 knockdown treatment groups (KDB1, KDB2, KDA1 and KDA2) and the mock treated group, generating four lists differentially expressed transcripts (corrected with an FDR of 0.01). The overlap between the two TCF4-B-specific groups (KDB1 and KDB2) was 70%, indicating a strong concordance in the effect of each treatment on differentially expressed genes (Figure 6.3.2 A). In comparison, the overlap between KDB1 and each of the TCF4-A-specific groups (KDA1 and KDA2)

were 60% and 47% respectively (Figure 6.3.2 B and C). The considerable overlap in gene expression changes between TCF4-B and TCF4-A specific groups (e.g. KDB1 and KDA1) suggests long and short isoforms may have shared effects on gene expression. However, some of the shared transcriptional changes may be non-specific in nature, caused by cells reacting to the presence of siRNA in the media and the activation of the gene silencing machinery. This analysis also underscored the lack of gene expression changes generated by the KDA2 treatment (403 differentially expressed transcripts) and the small overlap between the short isoform knockdown treatments (191 differentially expressed transcripts) (Figure 6.3.2 D).

To identify isoform-associated gene expression changes in the microarray data, a one-way ANOVA was implemented between the control and isoform-specific samples as described previous in Chapter 4. For example, to detect gene expression changes due to knockdown of TCF4-B transcripts, an ANOVA between the control group (4 mock samples and 3 KDGAP samples) and the long-isoform specific group (4 KDB1 samples and 4 KDB2 samples) was performed. To generate a high confidence list of gene expression changes a FDR correction of 0.01 was applied. The hierarchical clustering of the genes generated from the pooled KDB1 and KDB2 ANOVA (FDR0.01) generated a significant amount of shared gene expression changes (457) that clustered according to their treatment groups (Figure 6.4)

The TCF4-A associated ANOVA performed in this manner only generated limited amount of differentially expressed genes (14 genes, FDR 0.01) that were not sufficient for an informative enrichment analysis (Appendix V). Consequently, the less effective TCF4-A knockdown samples (KDA2) were removed from the downstream analysis and the following sections describes gene expression changes resulting from the KDA1 treatment (4 samples) alone. Hierarchical clustering of the top gene expression generated from the KDA1 ANOVA (1627, FDR0.01) showed appropriate clustering between control and treatment groups (Figure 6.4). Differentially expressed genes identified in this analysis were not represented in the KDA2 samples, as these appeared to be most similar to control groups (Appendix V). The hierarchical clustering of control and isoform-specific knockdown groups validated the use of the ANOVA for detecting gene expression changes associated to depletion of each isoform.

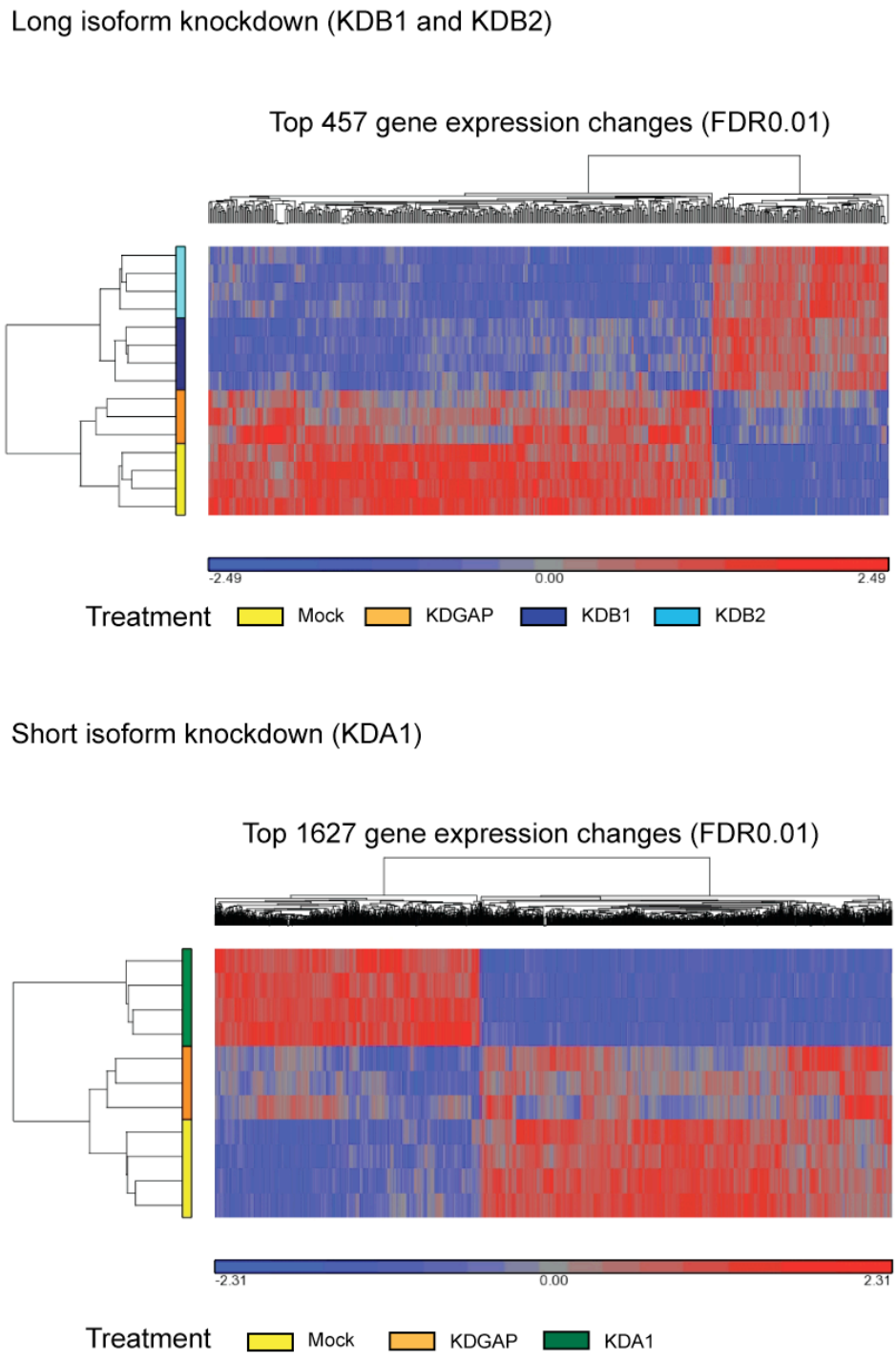


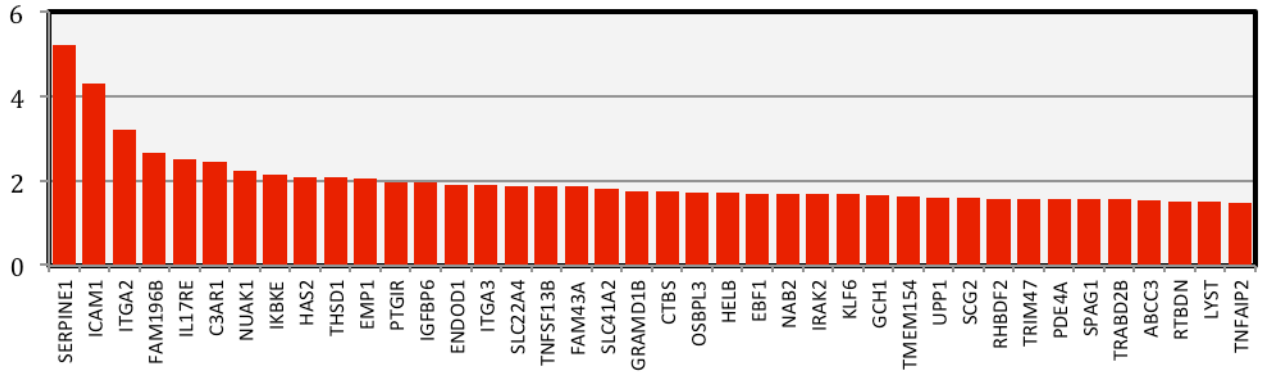
Figure 6.4 Hierarchical clustering of the top differentially expressed genes. The top differentially expressed genes (FDR 0.01) from each isoform specific ANOVA were subjected to hierarchical clustering. Each sample is demonstrated to cluster by treatment group as evident from the dendrogram to left of the heatmap. In addition, the control treatments (mock and KDGAP) and specific isoform knock down treatment (KDB1 and KDB1 or KDA1) are shown to cluster into defined groups. The top section of the heat map displays differential expression relative to controls: up-regulated genes are shown in red while down-regulated genes are coloured blue.

6.2.3 TCF4 isoform-associated genes expression changes

The gene expression changes obtained from the ANOVA described in the previous section established a list of high confidence transcriptional changes related to the TCF4-A and TCF4-B isoforms. The 40 top largest gene expression changes associated with each knockdown were apparently different and also diverged considerably from the top 40 gene expression changes identified from the global TCF4 knockdown described in Chapter 4 (Figure 4.6, Figure 6.5 and Figure 6.6). Nevertheless, downregulated genes such as dermatopontin (*DPT*), dachshund homolog 1 (*DACH1*), neurogenin 2 (*NEUROG2*), calsyntenin 2 (*CLSTN2*) and the upregulated gene, insulin-like growth factor binding protein 5 (*IGFBP5*) had amongst the highest fold changes in both global and isoform-specific TCF4 knockdown experiments (Figure 4.5, Figure 6.5 and Figure 6.6). This convergence may indicate that these genes are particularly important mediators of TCF4 function.

To experimentally validate the gene expression changes detected on the microarray, genes across discrete functional categories and with the fold change above or below 1.5 were selected for analysis by qPCR (Figure 6.7). *SNAI1*, *NEUROG2*, *RCOR2*, *TLE1* and *SKP2* were selected from the list of genes differentially expressed in the TCF4-B knockdown experiments (genes are denoted with a superscript B in Figure 6.7). In the TCF4-A knockdown group, *SNAI2, SEMA3A*, *RELN* and *DPT* were chosen for qPCR validation (genes are denoted with a superscript A in Figure 6.7). The expression of each gene was measured across all the treatment groups used on the microarrays (except KDA2) in order to assess the isoform specific effects. For each gene detected in the TCF4-B associated ANOVA (KDB), the qPCR data validated the microarray analysis showing significant differential expression between the control group and the KDB samples (Figure 6.7, superscript B). In the TCF4-A related ANOVA (KDA1), all the selected genes were validated by qPCR except for *SEMA3A*. However, a trend of downregulation was visible for *SEMA3A* (Figure 6.7, superscript A). It is also apparent from the qPCR data that knockdown of each isoform can have shared (Figure 6.7, A) and differential (Figure 6.7, B and C) effects on gene expression. Within the genes that display isoform-specific effects, *RCOR2*, *DPT*, *SEMA3A* and *RELN* were significantly different between the TCF4-B and TCF4-A isoform knockdowns (Figure 6.7, B and C).

Upregulated genes



Downregulated genes

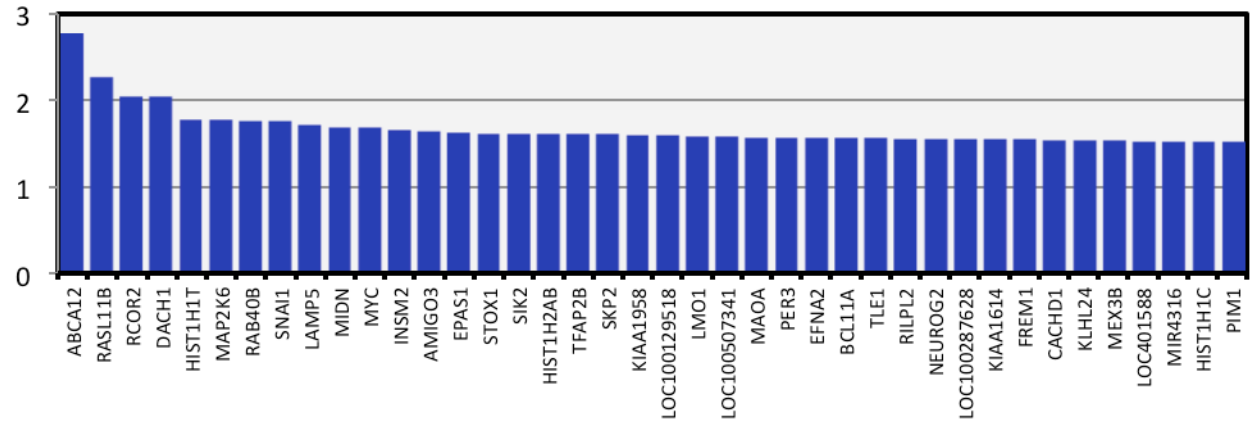


Figure 6.5 Gene expression changes in TCF4-B depleted cells. The top 40 upregulated (upper panel) and downregulated (lower panel) genes in TCF4-B knockdown cells after FDR correction (0.01) ranked by fold change. The y-axis represents fold up-regulation (upper panel) and fold down-regulation (lower panel).

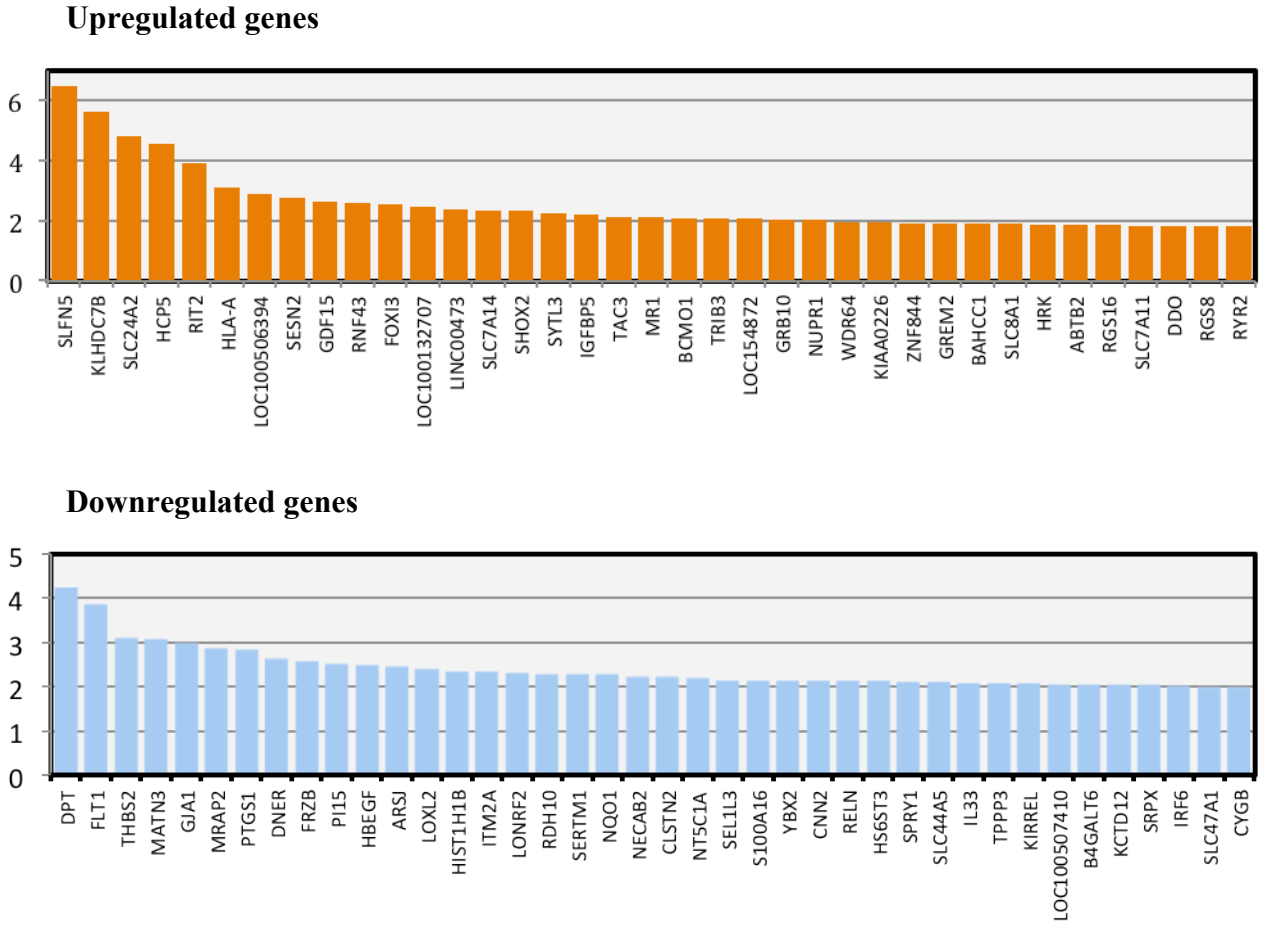


Figure 6.6 Gene expression changes in TCF4-A depleted cells. The top 40 upregulated (upper panel) and downregulated (lower panel) genes in TCF4-A knockdown cells after FDR correction (0.01) ranked by fold change. The y-axis represents fold up-regulation (upper panel) and fold down-regulation (lower panel).

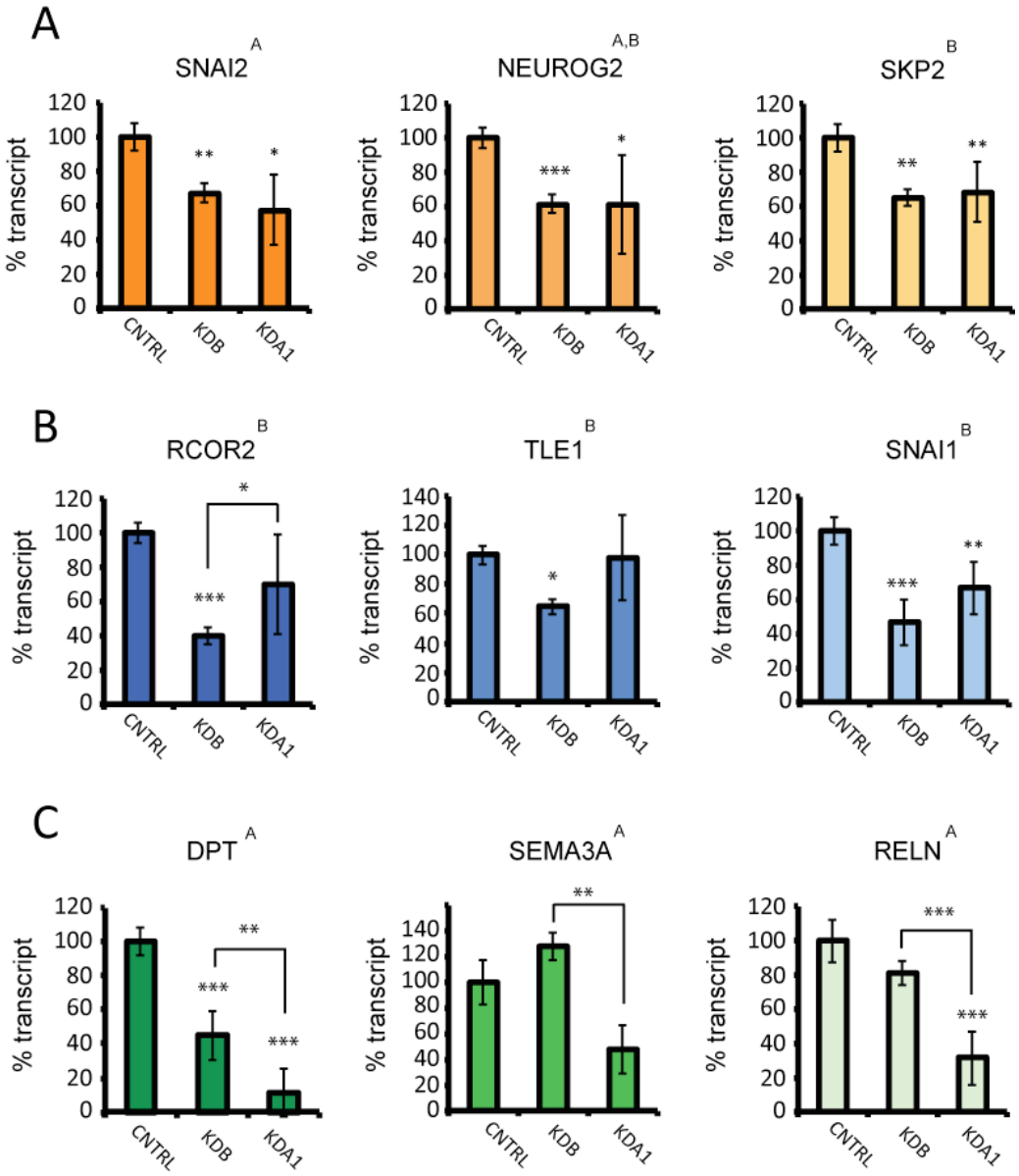


Figure 6.7 qPCR validation of differentially expressed genes. Nine differentially expressed genes belonging to various functional categories were chosen for qPCR validation. The genes identified as differentially expressed in the TCF4-B associated ANOVA are denoted with a superscript "B". Similarly, genes identified as differentially expressed in the TCF4-A associated ANOVA are denoted with a superscript "A". NEUROG2 was identified in both analyses and is denoted with "A" and "B". The expression of each gene was measured across all groups regardless of its significance in each ANOVA. (A) Genes similarly affected by TCF4-B and TCF4-A knockdown, (B) Genes more susceptible to TCF4-B knockdown (C) Genes more susceptible to TCF4-A knockdown. A Student's *t*-test was performed for each TCF4-associated knockdown category (KDB or KDA1) in comparison to controls (KDGAP and mock). In addition, a Student's *t*-test was also performed between the KDB and KDA1 knockdown groups to identify transcripts differentially expressed between each treatment group. Significant results are presented: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$

6.2.4 TCF4 isoforms regulate distinct biological processes

The list of differentially expressed genes in the TCF4-A and TCF4-B knockdowm grpups were subsequently converted to Entrez IDs and interrogated for process network enrichment in Metacore. The TCF4-A list was associated with 1319 unique Entrez IDs (FDR 0.01, 489 upregulated, 830 downregulated) whilst the TCF4-B list was associated with 386 unique Entrez IDs (FDR 0.01, 99 upregulated, 287 downregulated). To avoid a bias in the enrichment, each analysis was performed against a background list that contained all Affymetrix probes with Entrez IDs (23,816 unique IDs). This analysis generated distinct results for TCF4-A and TCF4-B, suggesting that each isoform may regulates different transcriptional programs in SH-SY5Y cells (Table 6.5 and Table 6.6, *all differentially expressed genes*). Specifically, knockdown of TCF4-B was associated with gene expression changes in processes associated to the cell cycle, DNA damage regulation and chromatin modification. By contrast, knockdown of TCF4-A resulted in differential expression of genes involved in cell adhesion, extracellular matrix interactions and cytoskeletal rearrangements. For each isoform the gene expression changes were significantly associated to several process network annotations that were distinct but functionally interconnected. The results of the enrichment analysis demonstrate a high degree of functional specificity for each isoform at the genome-wide level in SH-SY5Y cells.

The genes in each list were also examined for network processes restricted to up- and downregulated genes. Each list was deconstructed into the upregulated and downregulated components and re-analysed using the same parameters as above. Remarkably, all of the enrichment categories associated with each isoform-specific knockdown were only found in the downregulated genes. Accordingly no enrichment of any category was obtained from the upregulated genes (Table 6.5 and Table 6.6, *downregulated genes* and data not shown). In general, categories identified in the analysis of downregulated genes had more significant *P*-values, indicating a stronger enrichment of specific processes in downregulated genes compared to all the differentially expressed genes (Table 6.5 and Table 6.6 *downregulated genes*). This observation may suggest that TCF4 is predominantly a transcriptional activator in these cells, as gene expression changes are positively correlated to the abundance of TCF4.

Table 6.5 Enrichment analysis of differentially expressed genes after TCF4-B knockdown. The Entrez gene IDs derived from all the differentially expressed genes (386 genes, upper table) or just the downregulated genes (287 genes, lower table) in TCF4-B knockdown cells were analysed for enrichment using MetaCore. Each term is presented with its functional category, process network and corresponding FDR-corrected *P*-value (FDR). Categories highlighted in yellow pass multiple test correction (FDR 0.05).

All differentially expressed genes

Category	Process network	FDR	Genes
Cell cycle	S phase	5.89E-03	RFC5, TOP3A, POLD3, HIST1H1T, CDC6, CCNB1, CCNB2, HIST1H4D, PTTG1, CHAF1A, PRIM1, E2F1, CDC45
Cell cycle	Core	2.32E-02	CDC25A, CDC18L (CDC6), CAP-G, Cyclin B1, BUB3, CCNB2, BIRC5, Securin, E2F1, CDC45L
DNA damage	Checkpoint	2.73E-02	RFC5, RAD17, CDC25A, Histone H2AX, Cyclin B2, MAP2K6, RAD1, Securin, c-Myc, E2F1
Transcription	Chromatin modification	3.96E-02	H2AFZ, H2AFX, SMARCB1, HIST1H4D, RBBP4, TAF6L, HDAC5, Histone H3.3, SUDS3, HDAC2
Development	Hedgehog signaling	2.27E-01	MMP-9, SNAIL1, Cyclin B1, SREBP1 precursor, SKP2, BMPR1B, DLX5, ID3, NAB2, DYRK2, ZIC1, c-Myc, ZNF143
Development	Regulation of angiogenesis	2.55E-01	SREBP2 (nuclear), MMP-9, PAI1, SREBP1 precursor, DBH, N-cadherin, SREBP1 (nuclear), SREBP2 precursor, Secretogranin II, c-Myc, Ephrin-B2
Cell cycle	G0-G1	3.59E-01	CDC25A, RBBP8 (CtIP), RBBP4 (RbAp48), E2F1, HDAC2
Cell cycle	G2-M	3.59E-01	E2N(UBC13), Histone H1 testis, CDC25A, CAP-G, Cyclin B1, BUB3, SKP2, Cyclin B2, Securin, c-Myc
Signal Transduction	BMP/GDF signaling	4.14E-01	TLE1, BMPR1B, MEK6(MAP2K6), ActRIIB, c-Myc, ALK-1
DNA damage	Core	4.14E-01	RAD17, Histone H2AX, RAD1

Downregulated genes

Category	Process network	FDR	Genes
Cell cycle	S phase	2.86E-04	RFC5, TOP3A, POLD reg (p68), Histone H1 testis, CDC18L (CDC6), Cyclin B1, Cyclin B2, Histone H4, Securin, ChAF1 subunit A, PRIM1, E2F1, CDC45L
Cell cycle	Core	2.21E-03	CDC25A, CDC18L (CDC6), CAP-G, Cyclin B1, BUB3, Cyclin B2, Survivin, Securin, E2F1, CDC45L
DNA damage	Checkpoint	2.72E-03	RFC5, RAD17, CDC25A, Histone H2AX, Cyclin B2, MEK6(MAP2K6), RAD1, Securin, c-Myc, E2F1
Transcription	Chromatin modification	4.15E-03	H2AFZ, H2AFX, BAF47, Histone H4, RBBP4 (RbAp48), PAF65A, HDAC5, Histone H3.3, SDS3, HDAC2
Cell cycle	G2-M	8.33E-02	E2N(UBC13), Histone H1 testis, CDC25A, CAP-G, Cyclin B1, BUB3, SKP2, Cyclin B2, Securin, c-Myc
Cell cycle	G0-G1	1.26E-01	CDC25A, RBBP8 (CtIP), RBBP4 (RbAp48), E2F1, HDAC2
Development	Hedgehog signaling	1.26E-01	SNAIL1, Cyclin B1, SREBP1 precursor, SKP2, BMPR1B, DLX5, ID3, DYRK2, ZIC1, c-Myc, ZNF143
DNA damage	DBS repair	1.97E-01	E2N(UBC13), HMG2, Histone H2AX, Histone H4, RBBP4 (RbAp48), ChAF1 subunit A
DNA damage	Core	2.12E-01	RAD17, Histone H2AX, RAD1
Transcription	mRNA processing	2.12E-01	CPSF6, SFRS4, SAD1 (USP39), hnRNP A1, U2AF35, hnRNP K, snRNP-G, U1-70K

Table 6.6 Enrichment analysis of differentially expressed genes after TCF4-A knockdown. The Entrez gene IDs derived from all the differentially expressed genes (1319 genes, this page) or just the downregulated genes (830 genes, next page) in TCF4-A knockdown cells were analysed for enrichment using MetaCore. Each term is presented with its functional category, process network and corresponding FDR-corrected *P*-value (FDR). Categories highlighted in yellow pass multiple test correction (FDR 0.05). Abbreviations: ECM, extracellular matrix.

All differentially expressed genes

Category	Process	FDR	Genes
Cell adhesion	Cell-matrix interactions	3.95E-05	Biglycan, Fibrillin 1, Vitronectin, Nidogen-2, ECM1, LAMG1, ADAM23, ITGA6, ADAM-TS3, Fibulin-1, ADAM-TS9, CD44 (ICD), Endostatin, TIMP3, ADAM9, ITGAV, Tenascin-R, Aggrecanase-1, COL5A1, ITGA4, ADAM12, CD44, MAGP1, Tenascin-C, LAMA5, Nidogen, ITGB3, Layilin, MMP-24, EGFR, CD44 soluble, TIMP1, Connexin 43, CD44 (EXT), Collagen III
Cell adhesion	Integrin-mediated cell-matrix adhesion	4.05E-03	TGM2, CD151, RhoA, Fibrillin 1, Vitronectin, Profilin II, Tensin 3, CRK, ITGA6, Filamin A, Actin cytoplasmic 2, ITGAV, CD9, ITGA4, Tubulin alpha 1A, RhoGDI gamma, Tenascin-C, RhoB, LAMA5, Migfilin, ITGB3, Filamin B (TABP), NAG-2, ILK, Paxillin, Filamin C, Tetraspanin-8, Collagen III
Cytoskeleton	Regulation of cytoskeleton rearrangement	3.93E-02	Ankyrin 1, Ankyrin-G, RhoA, Socius, G-protein alpha-i1, RhoGAP6, LyGDI, SVIL, Profilin II, CRK, PDZ-RhoGEF, Filamin A, SPA1, G-protein alpha-s, mDIA2(DIAPH3), Tubulin alpha 1A, CD44, Vimentin, Filamin B (TABP), PLK1, Destrin, Paxillin, Filamin C, ELMO1
Cell adhesion	Platelet-endothelium-leucocyte interactions	3.93E-02	HGF, DDR1, TGF-beta 2, Vitronectin, TGF-beta 1, Protein C inhibitor, Tetraspanin-2, CD9, Thrombospondin 4, ITGA4, CD44, TGF-beta 3, Protein S, LDLR, TGF-beta receptor type II, EGFR, Endoplasmin, ZO-1, PDGF-D, Semaphorin 4D, ILK, Thrombospondin 3, Collagen III, Thrombospondin 2
Proteolysis	Connective tissue degradation	5.41E-02	Prolylcarboxypeptidase, Cystatin C, MTCBP-1, Vitronectin, Protein C inhibitor, ADAM23, ADAM-TS3, ADAM-TS9, TIMP3, ADAM9, Aggrecanase-1, ADAM12, SERPINB6, Tenascin-C, Nidogen, MMP-24, ADAM19, TIMP1, Collagen III
Proteolysis	ECM remodeling	5.41E-02	Cystatin C, MTCBP-1, Vitronectin, Protein C inhibitor, ADAM-TS3, Endostatin, TIMP3, Aggrecanase-1, SERPINB6, Tenascin-C, Nidogen, MMP-24, TIMP1, Clusterin, Collagen III
Immune response	Phagosome in antigen presentation	6.840E-02	Cathepsin L, RhoA, HLADPA1, Pyk2(FAK2), PSMD3, Vitronectin, Profilin II, PSMF1, CRK, PSMD14, GRP78, iC3b, Actin cytoplasmic 2, CD74, TRAM1, C3dg, JMJD6, ENDO180, CD63, C3, Endoplasmin, ERp72, PSMD13, PLC-beta1, Paxillin, ELMO1
Cytoskeleton	Actin filaments	8.05E-02	MYH9, Ankyrin 1, Ankyrin-G, Twinfilin, RhoA, SVIL, Profilin II, CRK, ABLIM1, Filamin A, mDIA2(DIAPH3), Piccolo, CD44, WASF3 (WAVE3), DAAM1, MYLK1, Migfilin, Filamin B (TABP), Destrin, Utrophin, Paxillin, Filamin C
Cell adhesion	Cadherins	1.37E-01	CD151, RhoA, PTP-2, PTPRF (LAR), Cadherin 24, PTPR-mu, SSX2IP, IQGAP1, CLR3, Actin cytoplasmic 2, PEZ, FHL2, Protocadherin 15, RET, Cadherin 23, DKK2, EGFR, Reelin, ZO-1, ILK, WNT6
Development	Neurogenesis Axonal guidance	1.76E-01	Semaphorin 3C, MYH9, NEK1, RhoA, Semaphorin 5A, Plexin A1, AHNAK, ADAM23, Plexin D1, IQGAP1, ABLIM1, PDZ-RhoGEF, DCAMKL1, Ephrin-A1, SPOCK2, G-protein alpha-s, Tenascin-R, MICAL, cAMP-GEFII, Semaphorin 3A, Ryanodine receptor 2, Reelin, Semaphorin 4D, Neuropilin-1, Semaphorin 3D

Downregulated genes

Category	Process	FDR	Genes
Cell adhesion	Cell-matrix interactions	5.80E-09	Biglycan, Fibrillin 1, Vitronectin, Nidogen-2, ECM1, LAMG1, ADAM23, ITGA6, ADAM-TS3, Fibulin-1, ADAM-TS9, CD44 (ICD), Endostatin, TIMP3, ADAM9, ITGAV, Aggrecanase-1, COL5A1, ITGA4, ADAM12, CD44, MAGP1, Tenascin-C, LAMA5, Nidogen, ITGB3, Layilin, MMP-24, EGFR, CD44 soluble, TIMP1, Connexin 43, CD44 (EXT), Collagen III
Cell adhesion	Integrin-mediated cell-matrix adhesion	1.75E-06	TGM2, CD151, RhoA, Fibrillin 1, Vitronectin, Profilin II, Tensin 3, CRK, ITGA6, Filamin A, Actin cytoplasmic 2, ITGAV, CD9, ITGA4, Tubulin alpha 1A, RhoGDI gamma, Tenascin-C, RhoB, LAMA5, Migfilin, ITGB3, Filamin B (TABP), NAG-2, ILK, Paxillin, Filamin C, Tetraspanin-8, Collagen III
Proteolysis	Connective tissue degradation	2.15E-03	Prolylcarboxypeptidase, MTCBP-1, Vitronectin, Protein C inhibitor, ADAM23, ADAM-TS3, ADAM-TS9, TIMP3, ADAM9, Aggrecanase-1, ADAM12, SERPINB6, Tenascin-C, Nidogen, MMP-24, ADAM19, TIMP1, Collagen III
Cell adhesion	Platelet-endothelium-leucocyte interactions	2.75E-03	HGF, DDR1, TGF-beta 2, Vitronectin, TGF-beta 1, Protein C inhibitor, Tetraspanin-2, CD9, ITGA4, CD44, TGF-beta 3, Protein S, LDLR, TGF-beta receptor type II, EGFR, Endoplasmin, PDGF-D, ILK, Thrombospondin 3, Collagen III, Thrombospondin 2
Proteolysis	ECM remodeling	3.80E-03	MTCBP-1, Vitronectin, Protein C inhibitor, ADAM-TS3, Endostatin, TIMP3, Aggrecanase-1, SERPINB6, Tenascin-C, Nidogen, MMP-24, TIMP1, Clusterin, Collagen III
Cytoskeleton	Actin filaments	2.82E-02	MYH9, Ankyrin-G, Twinfilin, RhoA, Profilin II, CRK, ABLIM1, Filamin A, mDIA2(DIAPH3), CD44, WASF3 (WAVE3), DAAM1, MYLK1, Migfilin, Filamin B (TABP), Destrin, Paxillin, Filamin C
Cytoskeleton	Regulation of cytoskeleton rearrangement	2.82E-02	Ankyrin-G, RhoA, G-protein alpha-i1, RhoGAP6, LyGDI, Profilin II, CRK, Filamin A, SPA1, mDIA2(DIAPH3), Tubulin alpha 1A, CD44, Vimentin, Filamin B (TABP), PLK1, Destrin, Paxillin, Filamin C
Cell adhesion	Cadherins	5.72E-02	CD151, RhoA, PTP-2, PTPRF (LAR), Cadherin 24, PTPR-mu, SSX2IP, IQGAP1, Actin cytoplasmic 2, PEZ, FHL2, RET, Cadherin 23, DKK2, EGFR, Reelin, ILK
Development	Regulation of EMT	5.81E-02	HGF, RhoA, TGF-beta 2, TGF-beta 1, CRK, BMP7, Lysyl oxidase, NOX4, TGF-beta 3, Vimentin, ITGB3, ETS1, TGF-beta receptor type II, SLUG, EGFR, PDGF-D, ILK, EDNRA, LOXL2, Collagen III
Cell adhesion	Integrin priming	9.59E-02	Vitronectin, Actin cytoplasmic 2, ITGAV, ITGB3, Calpain 2(m), CXCR4, PLC-beta1, Paxillin, Collagen III

6.3 Discussion

The *TCF4* gene is associated with two major protein isoforms: TCF4-B and TCF4-A. To advance our understanding of *TCF4* function in the nervous system, experiments were designed to explore the system-wide effects of TCF4 isoforms on the neuronal transcriptome. RNAi experiments revealed that each isoform affects a unique set of cellular processes. TCF4-B affected genes involved in cell cycle regulation and chromatin structure whilst TCF4-A was associated to gene expression changes in processes such as cell adhesion and cytoskeletal dynamics. The experiments in this chapter uncover novel cellular processes associated to TCF4 function in neuronal cells.

6.3.1 Knockdown of either TCF4-B or TCF4-A affects gene expression in SH-SY5Y cells

Having shown that two major isoforms are expressed in SH-SY5Y cells, siRNAs were used to knockdown TCF4-A or TCF4-B independently. The data derived from these experiments supports the MS identification of TCF4-A (54kDa) and TCF4-B (71kDa) as the two major TCF4 isoforms in SH-SY5Y cells (Table 5.2). Interestingly, knockdown of TCF4-A caused a compensatory increase in the expression of TCF4-B, indicating that homeostatic regulation may occur at the *TCF4* locus. This data suggests that a precise balance of each isoform may be required in cells. Despite compensatory mechanisms, a specific depletion of each isoform was possible using siRNA treatment. The gene expression profiles of TCF4-B and TCF4-A knockdown cells were analysed by microarray and compared individually to control groups (KDGAP and mock) to identify gene expression changes related to each isoform. Hierarchical clustering demonstrated that each isoform was associated with gene expression changes distinct from control groups (Figure 6.4). Depletion of each isoform resulted in both positive and negative effects on gene expression, however each isoform had a predominance of downregulated genes in this study (75% for TCF4-B and 62% for TCF4-A), indicating that TCF4-B and TCF4-A have similar effects on transcriptional activity in these cells. qPCR was used to validate gene expression changes from the array and demonstrated that knockdown of each isoform could have both shared and differential effects on gene expression (Figure 6.7).

6.3.2 TCF4-B and TCF4-A isoforms have shared effects on specific genes

Experimental validation of gene expression changes by qPCR demonstrated that in some cases TCF4-A and TCF4-B knockdown affected the expression of the same genes (Figure 6.7). This effect was particularly noticeable for *SNAI1*, *SNAI2*, *NEUROG2*, *SKP2* and *DPT*. In addition *SNAI2*, *DPT* and *NEUROG2* were identified in Chapter 4 as genes with high fold changes after global knockdown of TCF4 (*SNAI2*; -2.5, *DPT*; -6, *NEUROG2*; -3, Figure 4.5). .

SNAI2 is a zinc finger transcription factor that participates in developmental processes and EMT as described previously (Section 4.3.3). By contrast, *NEUROG2* is proneural protein essential for early fate commitment of neural precursors and has a role in the early steps of neural delamination, a process that may be related to EMT (Famulski and Solecki, 2013). Both of these transcription factors bind E-box motifs, suggesting that *TCF4*, *NEUROG2* and *SNAI2* may operate in a regulatory E-box network. Dermatopontin is a widely expressed extracellular matrix (ECM) protein that interacts with a range of ECM components to promote cell adhesion and migration (Kato et al., 2011; Liu et al., 2013). *DPT* is consistently the most differentially expressed gene across all data sets (TCF4 KD; -6, KDB; -2.2, KDA; -9). These findings may imply that *DPT* is a direct downstream target of TCF4.

DACH1, *CLSTN2* and *IGFBP5* were also genes with high fold changes (> 2) in the global TCF4 known and isoform-specific knockdown experiments (Figure 4.5, Figure 6.5, Figure 6.6). *IGFBP5* is an essential regulator of IGF signaling, that sequesters IGF ligands reducing their biological availability (Fernandez and Torres-Aleman, 2012). The *Drosophila* paralogue of *DACH1* (*dac*) is a master regulator of eye development and is also involved in developing mushroom bodies, a region of the *Drosophila* brain required for elementary cognitive functions (Chen et al., 1997; Kurusu et al., 2000). Abrogation of *dac* in flies results in disrupted and aberrant axonal projections in mushroom bodies (Kurusu et al., 2000). Calsyntenin 2 is a neuronally expressed transmembrane protein with N-terminal cadherin-like repeats and may therefore have a role in cell adhesion (Hintsch et al., 2002). However, very little is known about the functions of *DACH1*and *CLSTN2* in mammalian neurons.

6.3.3 Knockdown of TCF4-B affects genes involved in cell cycle regulation and chromatin modification

TCF4-B has been classified as both an activator and repressor of gene transcription depending on the cellular contexts and the promoter being studied (Skerjanc et al., 1996; Forrest et al., 2012). Enrichment analysis of gene expression changes in TCF4-B depleted cells identified a role for this isoform in cell cycle regulation and chromatin modification. The MetaCore database of process networks highlighted four terms that were statistically enriched for TCF4-B associated gene expression changes including, "Cell cycle_S phase", "Cell cycle_Core", "DNA damage_Checkpoint", "Transcription_Chromatin modification" (Table 6.5). These terms all became more significant when only the downregulated genes were included in the analysis, whilst no significant categories were found in the dataset composed of upregulated genes (Table 6.5 and data not shown). This observation, allied with the fact that a greater proportion of genes in the analysis (75%) were downregulated after the knockdown treatment suggest that TCF4-B may function predominantly as a transcriptional activator in these cells. However, the small number of upregulated genes (99) may have prevented the identification of enriched processes in this analysis.

Interestingly, the proteomic analysis of phosphorylation sites revealed that TCF4 contains amino acid motifs potentially recognised by cyclin-dependent kinases (CDKs) as well as the PRKDC and ATM proteins (Table 5.4). CDKs are directly involved in cell cycle regulation whereas the PRKDC and ATM kinases that are activated in response to DNA strand breaks; these two mechanisms are represented in TCF4-B related gene expression changes (O'Driscoll and Jeggo, 2006; Lim and Kaldis, 2013). Thus several lines of evidence suggest that TCF4-B may have a function in cell cycle control and DNA repair.

In the nervous system, neurogenesis is highly dependent on the different states of the cells cycle (Dehay and Kennedy, 2007). Controlling the timing and length of the cycle is decisive in regulating the balance of proliferative and differentiating neural precursors. In this model, decreasing the rate of the cell cycle progression increases the neuron generation. Hence, determining the effect of *TCF4* expression in proliferating neuronal cells could provide valuable information for its role in neuronal

development. There is already evidence that E-protein expression can influence the cell cycle and proliferation. In microarray experiments, cells overexpressing the TCF4 paralogue TCF3 were also associated to gene expression changes related to the cell cycle, indicative of a conserved function in E-proteins (Schwartz et al., 2006). Furthermore, expression of TCF4 has been associated to reduced proliferation and cell cycle arrest in a number of non-neuronal cancer cell lines (Pagliuca et al., 2000; Herbst et al., 2009). In neuroblastoma, overexpression of TCF4 can also cause a reduction of proliferation although the mechanism is undetermined (Rothschild et al., 2006). Similar evidence is available for TCF3 as it can induce cell cycle arrest in a number of different cell lines (Peverali et al., 1994; Engel and Murre, 2004). However, the precise role of E-proteins in cell cycle control is controversial as studies have also shown they can promote cell cycle progression (Zhao et al., 2001). Further cellular assays in neuronal models will be required to determine the exact function of TCF4-B on the cell cycle.

6.3.4 Knockdown of TCF4-A affects genes involved in cell adhesion and cytoskeletal structure

TCF4-A was identified as the second predominant isoform in SH-SY5Y cells. This isoform shares much of its sequence with TCF4-B although it is devoid of the N-terminal activation domain (AD1) and the NLS (Figure 6.1). Although only one siRNA resulted in robust knockdown of TCF4-A, process enrichment analysis in MetaCore identified a selection of terms associated to cell adhesion and cytoskeletal remodeling. These data differed considerably from the gene expression changes in TCF4-B-depleted cells (Table 6.6 and Table 6.5). Performing the enrichment analysis with all the differentially expressed genes or with exclusively the downregulated genes did not change the functional categories identified (Table 6.6). However, additional terms associated to proteolysis and cytoskeletal rearrangement became significant when the upregulated were excluded form the analysis. These results suggest the upregulated genes did not have a substantial role in functional processes identified.

The distinct enrichment categories identified in TCF4-A knockdown cells may suggest a form of cellular migration is being affected, as this process involves

restructuring of cell adhesion molecules and the cytoskeleton (Parsons et al., 2010). During cell migration, cells extending protrusions of the cell membrane that polarise the morphology of the cell. These protrusions are driven by the polymerisation of actin filaments that restructure the cytoskeleton. Cells can stabilise protrusions through cell adhesions molecules linking the actin cytoskeleton to ECM proteins. The depletion TCF4-A may therefore be associated to cellular migration or a related cellular process. Other terms that have been identified in the enrichment analysis could provide insight into the cellular processes that are being disrupted following TCF4-A knockdown (Table 6.6). In particular, the terms of "Development_Regulation of EMT" and "Neurogenesis_axon guidance" are intriguing, as EMT and axon guidance share some basic features of cellular migration.

Experiments presented in Chapter 4 in addition to the published literature would support a role for TCF4-A in EMT. Although EMT is not a significant term in the present MetaCore analysis, the significant terms are highly reminiscent of cellular processes in EMT. Notably, EMT involves the loss of adhesion and a morphogenic transition to a migratory phenotype that requires cytoskeletal rearrangements (Lim and Thiery, 2012). The qPCR validation experiments also demonstrate that the EMT-regulators *SNAI1* and *SNAI2* are differentially expressed in TCF4-A knockdown cells (Figure 6.7). This evidence may suggest that TCF4-A is the predominant isoform involved in regulating EMT.

The TCF4-A knockdown may also affect mechanisms associated to axon guidance, as this process shares many general features associated to neuronal migration, including changes in adhesive properties and cellular cytoarchitecture (Araujo and Tear, 2003). In addition, genes relating to this mechanism have high fold changes after TCF4-A is knocked down in SH-SY5Y cells (Table 6.6). For example, *RELN* and *SEMA3A* are two proteins related to this process that were selected for qPCR validation experiments (Figure 6.7). The semaphorins act as axonal growth cone guidance molecules and are crucial for managing precise neuronal connections in the brain (Pasterkamp, 2012). In addition to *SEMA3A*, many other semaphorin family members are also differentially expressed in this experiment indicating they may be associated to TCF4-A function. Interestingly, *SEMA3A* is also reduced in the cerebellum of schizophrenia patients suggesting it may contribute to the pathology of

neurodevelopmental disorders (Eastwood et al., 2003). The differential expression of *RELN* is also notable in this experiment because mutations in this gene cause cerebellar malformations and several post-mortem studies have shown reduced *RELN* expression in the brain of schizophrenia patients (Impagnatiello et al., 1998; Fatemi et al., 2000; Guidotti et al., 2000; Hong et al., 2000). Through affecting the expression of important regulators of axon guidance genes, TCF4-A could therefore disrupt aspects of neurodevelopment that confer susceptibility to schizophrenia.

In conclusion, experiments using RNAi have allowed for isoform-specific knockdown to be performed on TCF4 transcripts. In combination with the MS data presented in Chapter 5, this technique has established that TCF4-B and TCF4-A are the major TCF4 isoforms in SH-SY5Y cells. Knockdown experiments in these cells demonstrate that the TCF4-B and TCF4-A cause distinctly different gene expression changes. TCF4-B depletion is associated with gene expression changes associated to the cell cycle and chromatin modification whilst the TCF4-A depletion affects genes involved in cell adhesion, remodeling of the cytoskeleton and proteolysis of extracellular matrix. This study provides the first evidence that TCF4 isoforms can have unique effects in neuronal cells. The cellular processes identified have important implications for understanding the role of *TCF4* in neurodevelopment and disease.

Chapter 7

General Discussion

7.1 The function of *TCF4* missense mutations in PTHS

PTHS is caused by mutations in *TCF4* that result in haploinsufficiency of the encoded protein (Amiel et al., 2007). At the beginning of this study, a number of *TCF4* missense mutations had been identified in PTHS patients although their impact on the function of the protein had only been studied in the context of transactivation on synthetic promoters (de Pontual et al., 2009). Accordingly, one of the aims of this thesis was to further examine the effects of missense mutations on TCF4 function. Missense mutations are especially informative to study as they allow changes in single amino acids to be related to particular functional deficits. In this thesis I showed that most *TCF4* missense mutations affected either the subcellular localization, dimerisation or transcriptional activity as assayed on one of two potentially physiological promoters: *NRXN1β* and *CNTNAP2* (Forrest et al., 2012). Mutations in each of these loci have been shown to cause a PTHS-like disorder suggesting that *TCF4*, *NXNR1β* and *CNTNAP2* may form a regulatory network (Zweier et al., 2009; Blake et al., 2010).

One of the unexpected findings of this study was the idiosyncratic behavior of each mutant in the different functional tests (3.2.4 and 3.2.6). These observations were surprising because each of the missense mutations modeled cause PTHS, and no apparent correlation between genotype and phenotype has been reported (Marangi et al., 2011). However, the data in chapter 3 emphasises the importance of context in measuring functional deficits. In particular, distinct functional deficits were observed at the level of transcriptional activity when the same mutant was assayed on the *NXNR1β* and *CNTNAP2* promoters (Figures 3.11 and 3.12). The same mutations also displayed context dependent deficits when different dimer configurations were used to measure protein:protein interactions (Figures 3.6 and 3.7). These results suggest

that measuring the functional impact of a mutation *in vivo* requires an appreciation of the relevant physiological context in which TCF4 operates.

One of the considerations in the functional analysis of *TCF4* missense mutations is the choice of DNA sequence used to assay transcriptional activity. The *NXNR1β* and *CNTNAP2* promoters were selected on the basis that TCF4 may regulate the expression of these genes, since common neurodevelopmental phenotypes can be caused by *TCF4*, *NRXN1* and *CNTNAP2* mutations (Zweier et al., 2009). Although I was able to show that these promoters are responsive to TCF4 expression and contain conserved E-boxes, it is still uncertain whether TCF4 directly regulates these genes physiologically (Figure 3.10). Consequently, it will be imperative to identify TCF4's genomic binding sites, in order to select promoter or enhancer sequences that are relevant to its transcriptional activity. The use of chromatin immunoprecipitation combined and next generation sequencing (ChIP-seq) would directly facilitate the discovery of TCF4 binding sites *in vivo* (Furey, 2012). The gene expression changes from global *TCF4* knockdown cells (Chapter 4) or the isoform-specific knockdown cells (Chapter 6) could be particularly useful to interpret such data. Nevertheless, the use of native promoter sequences, such as those used in Chapter 3, clearly demonstrate that the effects of *TCF4* mutations can be more subtle than those observed on synthetic promoters containing multimerised E-boxes (Forrest et al., 2012; Sepp et al., 2012b). Although useful, synthetic promoters do not model the inherent complexity of transcriptional regulation *in vivo* that results from the additive effects of multiple *cis*- and *trans*- acting factors (Coulon et al., 2013). Indeed E-box binding sites have been shown to cooperate with neighbouring sequences to mediate their functional effects (Castro et al., 2006; McDonald et al., 2013). In light of this information, it is crucial to identify physiologically regulated TCF4 targets to better our understanding of the molecular pathophysiology of PTHS, learning disability and schizophrenia.

Differences in the transcriptional activities of the PTHS-associated TCF4 missense mutants could be mediated by impaired homo- and heterodimerisation (Figures 3.6 and 3.7). Mutations in the bHLH domain (R578P, R580W and A610V) particularly affected protein association to partner bHLH proteins. Defective interactions with bHLH proteins may have important consequences for neurodevelopmental

phenotypes in PTHS, as proneural proteins are important for multiple aspects regulating neurogenesis and neuronal differentiation (Bertrand et al., 2002). The interaction of TCF4 with ATOH1 may represent such an example. In mice, Atoh1 is required for the migration and differentiation of hindbrain neurons critical for perinatal breathing (Rose et al., 2009). One of the salient features of PTHS is an abnormal breathing pattern characterised by intermittent hyperventilation and apnea (Peippo and Ignatius, 2012). In addition, studies on *Tcf4* knockout mice have shown that a critical amount of the Tcf4/Atoh1 heterodimer is necessary to develop specific populations of cells in the hindbrain (Flora et al., 2007) (Section 1.9.2). These studies raise the intriguing possibility that mutations affecting the TCF4/ATOH1 interaction may subtly disrupt the development of hindbrain structures regulating autonomic function and cause the abnormal breathing patterns observed in PTHS.

Although TCF4 missense mutations impaired dimerization with a selection of known TCF4-interacting proteins, each heterodimer was affected differently by individual mutations (Figure 3.7). Therefore, another consideration for the function analysis of PTHS missense mutations is the choice of binding partner used to assess the effects on protein:protein interactions. The proteomic analysis of proteins that co-purify with TCF4 (Chapter 5) generated an interesting resource of potential TCF4-interacting proteins. Interestingly, this analysis did not identify any of the known heterodimer partners used to characterise TCF4 missense mutants (ASCL1, ATOH1, NEUROD1 or ID2). However, the bHLH proteins TCF3, HAND2 (dHAND) and TWIST2 co-immunoprecipitated with TCF4 in SH-SY5Y cells (Table 5.6). HAND2 and TWIST2 have been reported to interact with E-proteins (Howard et al., 1997; Firulli et al., 2005). TWIST2 is also a mediator of EMT that was identified as an important TCF4-associated cellular process (Section 4.2.5)(Fang et al., 2011). It would therefore be interesting to assess the effects of TCF4 missense mutations on interactions with each of these proteins.

Whilst mutations in the bHLH domain impaired homo- and heterodimer formation, the more N-terminal TCF4 mutations G358V and D535G had a lesser effect on protein:protein interactions (Figures 3.6 and 3.7). By contrast, N-terminal mutations had context-dependent deficits in transcriptional activation measured on the NRXN1 and CNTNAP2 promoters, indicating that their ability to regulate transcription was

affected (Figures 3.11 and 3.12). G358V is located in AD2 a region of TCF4 proposed to interact with the transcriptional co-repressor ETO (Figure 3.1)(Zhang et al., 2004). D535G is located in the Rep domain that can functionally regulate the activity of AD1 and AD2 (Figure 3.1)(Markus et al., 2002b). Therefore, the mutations G358V and D535G might impair the recruitment of regulator proteins important for TCF4 function. In addition to ETO, interesting candidates to investigate in this context would be interactions with the NCoR complex (NCOR1, TBL1XR1 and HDAC3) that co-purified with TCF4 (Section 5.2.4). A recent study found that *Mecp2* missense mutations found in Rett syndrome patients abrogate the interaction between Mecp2 and NCoR in transgenic mice (Lyst et al., 2013). As mentioned previously, Rett syndrome has substantial phenotypic overlap with PTHS (Armani et al., 2012). Similar to PTHS, *MECP2* missense mutations that cause Rett syndrome cluster in two distinct regions of the protein (Lyst et al., 2013). The majority of *MECP2* mutations are located in the methyl-binding domain (MBD) of the protein however other mutations are located in the transcriptional repression domain (TRD). Whereas mutations in the MBD impair DNA binding to methylated DNA, the mutation R306C, located in the TRD, eliminates the protein interactions with NCOR1, TBL1XR1 and HDAC3 (Yusufzai and Wolffe, 2000; Lyst et al., 2013). Providing that the interactions described in Chapter 5 can be confirmed using different methodology, it would be interesting to determine whether the G358V and D535G mutations impair the interaction between TCF4 and NCoR.

In addition to modeling mutations on TCF4-B, it may be useful to interpret the effects of PTHS-associated missense mutations on TCF4-A. TCF4-A appears to be equally abundant as TCF4-B in some human cell lines of neuronal origin (Chapter 5, Figure 5.2A). TCF4-A does not contain AD1 therefore the effects of N-terminal mutations affecting regulatory domains may reveal stronger functional impairments in transcriptional assays. TCF4-A may also have differential effects on protein:protein interactions that could be investigated in the light of the findings described above.

Finally, additional experiments could be designed to demonstrate the physiological effects of missense mutations in PTHS. One possibility would be to generate knock-in mice with alleles containing PTHS missense mutations, as has been done to understand the pathophysiology of certain autism spectrum disorders (Tabuchi et al.,

2007; Lyst et al., 2013). Finally, induced pluripotent stem cells (iPS) could be created from PTHS fibroblasts allowing disease-specific cell models to be formed. Re-programming iPS cells from PTHS patients into neurons could be an effective strategy to identify novel molecular and cellular mechanisms relevant to PTHS pathology (Brennand et al., 2011).

7.2 TCF4-associated cellular processes and neurodevelopmental disease.

As mentioned above, genetic studies have uncovered an important role for *TCF4* in human neurodevelopmental disorders. However, the precise way in which TCF4 contributes to disease in humans is unknown. Consequently, another aim of this thesis was to delineate the gene expression program that TCF4 may regulate in neuronal cells. I found that knockdown of all TCF4 transcripts in SH-SY5Y cells results in the differential expression of genes involved in cell signaling, cell survival, EMT and neuronal differentiation (Forrest et al., 2013).

Although these microarray experiments uncovered potentially important cellular processes related to TCF4 function, this method did not allow for direct and indirect effects on gene expression to be distinguished. As previously mentioned in Section 7.1, one of the aims of future experiments would be to examine the genome-wide binding sites for TCF4 in neuronal cells in order to identify direct genomic targets. Such experiments would provide additional understanding of TCF4 function and the genes it regulates. However, the RNAi-mediated knockdown approach was useful as it created a cellular model of TCF4 deficiency that may be relevant in PTHS and schizophrenia. Haploinsufficiency of TCF4 is widely assumed to be the cause of PTHS however the effects of common schizophrenia risk variants on *TCF4* expression are so far unknown. Preliminary evidence from mouse models and from the blood of psychosis patients may suggest that *TCF4* up-regulation is associated with schizophrenia risk, however rigorous evidence of this is lacking (Brzozka et al., 2010; Wirgenes et al., 2012). Nevertheless, the cellular processes revealed by knockdown of *TCF4* are likely to be important in understanding the role of *TCF4* in schizophrenia risk and PTHS (Figure 7.1).

One the of the most intriguing findings in *TCF4* knockdown cells was the differential expression of certain neurodevelopmental genes involved in dominantly inherited

forms of intellectual disability (Chapter 4, Table 4.6). In addition, several E-box binding transcription factors (*ASCL1*, *NEUROG2*, *SNAI1*, *SNAI2*, *BHLHE40*, *ZEB2*, *TWIST1*, *MEF2C*) were differentially expressed in TCF4-depleted cells. *TCF4*, *SNAI1*, *SNAI2*, *ZEB1* and *TWIST1* may operate in a transcriptional hierarchy in epithelial cells undergoing full EMT (Cano and Portillo, 2010). Interestingly, MEF2C is involved in the activity dependent transcription of *TCF4* and synergises with neurogenic bHLH transcription factors at E-box sites (Black et al., 1996; Flavell et al., 2008). Since heterozygous *ZEB2* and *MEF2C* mutations cause severe forms mental retardation, these data may suggest a functional network of E-box regulators is also operating in neuronal cells to regulate crucial neurodevelopmental events important in human disorders.

The knockdown experiments in neuroblastoma cells confirmed that *TCF4* has a role in regulating EMT (Chapter 4). The process of EMT has been extensively studied in epithelial cells and cancer however, the role of an EMT-like process in the brain is less clear (Acloque et al., 2009). Nevertheless, there are considerable parallels between EMT and the events regulating neuronal development (Famulski and Solecki, 2013). During early steps of corticogenesis, neuroepithelial cells (neural stem cells) must delaminate from the basal surface of the ventricular zone to become multipolar cells (mesenchymal-like), in order to initiate migration into the cortical layers. Switching between polar (epithelial-like) and multipolar (mesenchymal-like) states is also required during migration through cortical layers and is an essential step of cortical development. The delamination of progenitor cells may be an important process to investigate in *Tcf4* knockout mice, as Tcf4 expression is particularly high in the ventricular zones of the developing and adult brain (Soosaar et al., 1994). The change in the morphological and migratory properties of cells associated with EMT may also be of relevance to neurodevelopmental disorders because mutations in the EMT-regulator *ZEB2*, cause Mowat-Wilson syndrome (MWS) (Vandewalle et al., 2005). MWS is associated with severe intellectual disability and is a differential diagnosis for PTHS suggesting that a common disease mechanisms may exist (Peippo and Ignatius, 2012) . However, ZEB2 also plays a crucial role in interneuron subtype specification indicating that TCF4 and ZEB2 may have pleiotropic functions in the nervous system (McKinsey et al., 2013).

The proteomic analysis of TCF4-interacting proteins suggests that TCF4 may have novel molecular functions in RNA splicing, cell cycle regulation and chromatin organisation (Table 5.6). Although these interactions require further validation, they potentially reveal a different aspect of TCF4 function that has not yet been addressed. The identification of members of the NCoR complex in IAP-MS experiment provides evidence that TCF4 may act as a repressor in particular contexts. This may explain the presence of up-regulated genes in the *TCF4* knockdown experiment (Chapters 4 and 6). In addition, the phosphorylation sites identified by MS may provide useful insight to predict upstream regulators of TCF4 function (Table 5.4). Understanding how TCF4 is regulated in neurons will be essential in determining its biological role in the integrated network of the cell. Proteomics also confirmed the present of the TCF4-B and TCF4-A isoforms in SH-SY5Y cells (Table 5.2). These data enabled a more refined analysis of TCF4 function to be performed using isoform-specific knockdowns (Chapter 6). These experiments revealed strikingly different functions for each isoform that were unexpected due to the apparent redundant effects of TCF4-B and TCF4-A in epithelial (MDCK) cells (Sobrado et al., 2009). The results suggest that TCF4 isoforms may have unique functions in neuronal cells.

Multiple terms in the enrichment analysis of TCF4-B knockdown cells suggested that this isoform has a role in cell cycle regulation (Table 6.5). In addition, cell cycle proteins co-purified with TCF4, indicating TCF4 may have an integrated role in the cell cycle. Tight regulation of cell cycle is essential during brain development as it has a direct impact on the balance between proliferation and differentiation of cells, and affects cortical architecture (Dehay and Kennedy, 2007). Further experiments assessing the effects of TCF4 on neuronal cell proliferation and differentiation would be very useful to understand the functional implications of these findings. Evidence from magnetic resonance imaging (MRI) studies of the brain in PTHS patients may provide support for TCF4 having a role in cell cycle control and proliferation. One of the most frequent brain abnormalities observed in PTHS patients are small hippocampi (34%) (Whalen et al., 2012b). In addition, dilated ventricles and frontal lobe hypoplasia have also been reported (Amiel et al., 2007; Zweier et al., 2008a). These structural defects could be the result of defective proliferation of developing neurons in the brain. Volumetric reduction of different brain regions is also a prominent finding in schizophrenia patients (Goldman et al., 2009; Balu and Coyle,

2011). However, the structural abnormalities in the brain can be caused by various mechanisms such as cell loss or cell packing, therefore further studies would be required to test these predictions (Balu and Coyle, 2011).

Gene expression changes associated with TCF4-A knockdown suggested a role for this isoform in cell adhesion, cytoskeletal remodeling and proteolysis of the extracellular matrix (Table 6.6). These gene expression changes are reminiscent of processes related to EMT or forms of neuronal migration (Ayala et al., 2007). Functional characterization of these processes will be essential to determine whether TCF4-A knockdown and global knockdown of *TCF4* can affect EMT and cell migration. Importantly, several schizophrenia candidate genes such as *RELN* (reelin) and *DISC1* (disrupted in schizophrenia 1) are involved in neuronal migration suggesting this may be an important cellular process governing neuropsychiatric phenotypes (Anton et al., 2004; Franco et al., 2011; Tomita et al., 2011). Gene enrichment analyses of schizophrenia susceptibility genes also support a role for biological process such as cell adhesion and migration in disease pathophysiology (O'Dushlaine et al., 2011; Gilman et al., 2012). To further our understanding of TCF4 in schizophrenia it will be important to determine whether TCF4 regulates validated schizophrenia susceptibility genes. Identification of new schizophrenia susceptibility loci by GWAS, CNV analysis and exome sequencing will be a useful resource to examine this possibility (Section 1.11.2).

In summary, the data presented in this thesis describes novel molecular and cellular functions for TCF4 in SH-SY5Y cells that have important consequences for nervous system development. Global knockdown of *TCF4* affects the expression of genes that are associated with intellectual disability and EMT. Furthermore, TCF4 isoforms are shown to have distinct effects in neuronal cells that together affect diverse cellular processes such as the cell cycle, cell adhesion and cytoskeletal dynamics. Accordingly, the transcriptional dysregulation of these processes may cause neurodevelopmental abnormalities that are associated to PTHS and schizophrenia (Figure 7.1).

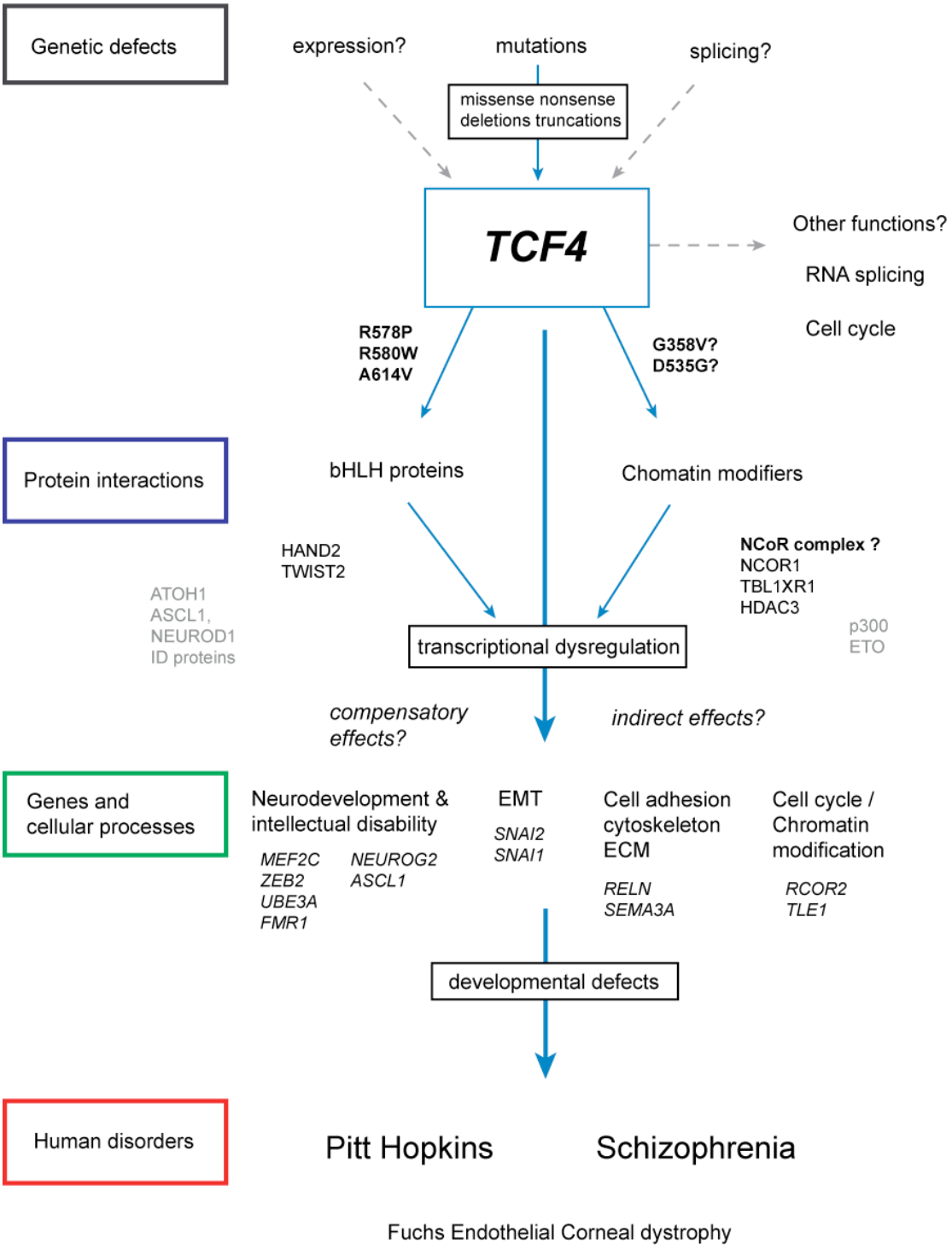


Figure 7.1 Overview of key findings from this study and their relation to potential disease mechanisms. Genetic defects cause a change in the activity or expression levels of TCF4. These mechanisms disrupt TCF4-associated protein interactions and transcription, predisposing to developmental phenotypes in human disorders. Genes and proteins identified in this study are in black. Known TCF4-interacting proteins previously found in other studies are in grey.

Bibliography

Aberg KA et al. (2013) A comprehensive family-based replication study of schizophrenia genes. JAMA Psychiatry 70:573-581.

Acloque H, Adams MS, Fishwick K, Bronner-Fraser M, Nieto MA (2009) Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. The Journal of Clinical Investigation 119:1438-1449.

Akazawa C, Sasai Y, Nakanishi S, Kageyama R (1992) Molecular characterization of a rat negative regulator with a basic helix-loop-helix structure predominantly expressed in the developing nervous system. Journal of Biological Chemistry 267:21879-21885.

Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, Sebat J, Wigler M, Martin CL, Ledbetter DH, Nelson SF, Cantor RM, Geschwind DH (2008) Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. American Journal of Human Genetics 82:150-159.

Amiel J, Rio M, de Pontual L, Redon R, Malan V, Boddaert N, Plouin P, Carter NP, Lyonnet S, Munnich A, Colleaux L (2007) Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. American Journal of Human Genetics 80:988-993.

Anton ES, Ghashghaei HT, Weber JL, McCann C, Fischer TM, Cheung ID, Gassmann M, Messing A, Klein R, Schwab MH, Lloyd KC, Lai C (2004) Receptor tyrosine kinase ErbB4 modulates neuroblast migration and placement in the adult forebrain. Nature Neuroscience 7:1319-1328.

Araujo SJ, Tear G (2003) Axon guidance mechanisms and molecules: lessons from invertebrates. Nature Reviews: Neuroscience 4:910-922.

Arenas JE, Abelson JN (1997) Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proceedings of the National Academy of Sciences of the United States of America 94:11798-11802.

Argenton F, Arava Y, Aronheim A, Walker MD (1996) An activation domain of the helix-loop-helix transcription factor E2A shows cell type preference in vivo in microinjected zebra fish embryos. Molecular and Cellular Biology 16:1714-1721.

Armani R, Archer H, Clarke A, Vasudevan P, Zweier C, Ho G, Williamson S, Cloosterman D, Yang N, Christodoulou J (2012) Transcription factor 4 and myocyte enhancer factor 2C mutations are not common causes of Rett syndrome. American journal of medical genetics Part A 158A:713-719.

Aronheim A, Shiran R, Rosen A, Walker MD (1993) The E2A gene product contains two separable and functionally distinct transcription activation domains. Proceedings of the National Academy of Sciences of the United States of America 90:8063-8067.

Ayala R, Shu T, Tsai LH (2007) Trekking across the brain: the journey of neuronal migration. Cell 128:29-43.

Bacanu SA, Chen J, Sun J, Richardson K, Lai CQ, Zhao Z, O'Donovan MC, Kendler KS, Chen X (2013) Functional SNPs are enriched for schizophrenia association signals. Molecular Psychiatry.

Bain G, Murre C (1998) The role of E-proteins in B- and T-lymphocyte development. Seminars in Immunology 10:143-153.

Balu DT, Coyle JT (2011) Neuroplasticity signaling pathways linked to the pathophysiology of schizophrenia. Neuroscience & Biobehavioral Reviews 35:848-870.
Baratz KH, Tosakulwong N, Ryu E, Brown WL, Branham K, Chen W, Tran KD, Schmid-Kubista KE, Heckenlively JR, Swaroop A, Abecasis G, Bailey KR, Edwards AO (2010) E2-2 protein and Fuchs's corneal dystrophy New England Journal Medicine. 363:1016-1024.
Bayly R, Chuen L, Currie RA, Hyndman BD, Casselman R, Blobel GA, LeBrun DP (2004) E2A-PBX1 interacts directly with the KIX domain of CBP/p300 in the induction of proliferation in primary hematopoietic cells. Journal of Biological Chemistry 279:55362-55371.
Bayly R, Murase T, Hyndman BD, Savage R, Nurmohamed S, Munro K, Casselman R, Smith SP, LeBrun DP (2006) Critical role for a single leucine residue in leukemia induction by E2A-PBX1. Molecular and Cellular Biology 26:6442-6452.
Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell 61:49-59.
Bergqvist I, Eriksson M, Saarikettu J, Eriksson B, Corneliussen B, Grundstrom T, Holmberg D (2000) The basic helix-loop-helix transcription factor E2-2 is involved in T lymphocyte development. European Journal of Immunology 30:2857-2863.
Bertrand N, Castro DS, Guillemot F (2002) Proneural genes and the specification of neural cell types. Nature Reviews: Neuroscience 3:517-530.
Bhakar AL, Dolen G, Bear MF (2012) The pathophysiology of fragile X (and what it teaches us about synapses). Annual Review of Neuroscience 35:417-443.
Bhalla S, Spaulding C, Brumbaugh RL, Zagort DE, Massari ME, Murre C, Kee BL (2008) differential roles for the E2A activation domains in B lymphocytes and macrophages. Journal of Immunology 180:1694-1703.
Bhattacharya A, Kaphzan H, Alvarez-Dieppa AC, Murphy JP, Pierre P, Klann E (2012) Genetic removal of p70 S6 kinase 1 corrects molecular, synaptic, and behavioral phenotypes in fragile X syndrome mice. Neuron 76:325-337.
Black BL, Ligon KL, Zhang Y, Olson EN (1996) Cooperative transcriptional activation by the neurogenic basic helix-loop-helix protein MASH1 and members of the myocyte enhancer factor-2 (MEF2) family. Journal of Biological Chemistry 271:26659-26663.
Blake DJ, Forrest M, Chapman RM, Tinsley CL, O'Donovan MC, Owen MJ (2010) TCF4, schizophrenia, and Pitt-Hopkins Syndrome. Schizophrenia Bulletin 36:443-447.
Blow JJ, Dutta A (2005) Preventing re-replication of chromosomal DNA. Nature Reviews: Molecular Cell Biology 6:476-486.
Bolos V, Peinado H, Perez-Moreno MA, Fraga MF, Esteller M, Cano A (2003) The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. Journal of Cell Science 116:499-511.
Bond AM, Bhalala OG, Kessler JA (2012) The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation. Developmental Neurobiology 72:1068-1084.
Boutin C, Hardt O, de Chevigny A, Core N, Goebbels S, Seidenfaden R, Bosio A, Cremer H (2010) NeuroD1 induces terminal neuronal differentiation in

olfactory neurogenesis. Proceedings of the National Academy of Sciences of the United States of America 107:1201-1206.
Bracko O, Singer T, Aigner S, Knobloch M, Winner B, Ray J, Clemenson GD, Jr., Suh H, Couillard-Despres S, Aigner L, Gage FH, Jessberger S (2012) Gene expression profiling of neural stem cells and their neuronal progeny reveals IGF2 as a regulator of adult hippocampal neurogenesis. The Journal of Neuroscience 32:3376-3387.
Bradney C, Hjelmeland M, Komatsu Y, Yoshida M, Yao TP, Zhuang Y (2003) Regulation of E2A activities by histone acetyltransferases in B lymphocyte development. Journal of Biological Chemistry 278:2370-2376.
Braff DL, Geyer MA (1990) Sensorimotor gating and schizophrenia. Human and animal model studies. Archives of General Psychiatry 47:181-188.
Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, Li Y, Mu Y, Chen G, Yu D, McCarthy S, Sebat J, Gage FH (2011) Modelling schizophrenia using human induced pluripotent stem cells. Nature 473:221-225.
Brockschmidt A, Todt U, Ryu S, Hoischen A, Landwehr C, Birnbaum S, Frenck W, Radlwimmer B, Lichter P, Engels H, Driever W, Kubisch C, Weber RG (2007) Severe mental retardation with breathing abnormalities (Pitt-Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. Human Molecular Genetics 16:1488-1494.
Brown SJ, Stoilov P, Xing Y (2012) Chromatin and epigenetic regulation of pre-mRNA processing. Human Molecular Genetics 21:R90-96.
Brzozka MM, Rossner MJ (2013) Deficits in trace fear memory in a mouse model of the schizophrenia risk gene TCF4. Behavioural Brain Research 237:348-356.
Brzozka MM, Radyushkin K, Wichert SP, Ehrenreich H, Rossner MJ (2010) Cognitive and sensorimotor gating impairments in transgenic mice overexpressing the schizophrenia susceptibility gene Tcf4 in the brain. Biological Psychiatry 68:33-40.
Buckley SM, Aranda-Orgilles B, Strikoudis A, Apostolou E, Loizou E, Moran-Crusio K, Farnsworth CL, Koller AA, Dasgupta R, Silva JC, Stadtfeld M, Hochedlinger K, Chen EI, Aifantis I (2012) Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. Cell Stem Cell 11:783-798.
Cano A, Portillo F (2010) An emerging role for class I bHLH E2-2 proteins in EMT regulation and tumor progression. Cell Adhesion & Migration 4:56-60.
Cano A, Perez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, Portillo F, Nieto MA (2000) The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nature Cell Biology 2:76-83.
Casarosa S, Fode C, Guillemot F (1999) Mash1 regulates neurogenesis in the ventral telencephalon. Development 126:525-534.
Castro DS, Skowronska-Krawczyk D, Armant O, Donaldson IJ, Parras C, Hunt C, Critchley JA, Nguyen L, Gossler A, Gottgens B, Matter JM, Guillemot F (2006) Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. Developmental Cell 11:831-844.
Castro DS, Martynoga B, Parras C, Ramesh V, Pacary E, Johnston C, Drechsel D, Lebel-Potter M, Garcia LG, Hunt C, Dolle D, Bithell A, Ettwiller L, Buckley N, Guillemot F (2011) A novel function of the proneural factor Ascl1 in

progenitor proliferation identified by genome-wide characterization of its targets. Genes & Development 25:930-945.
Chen B, Lim RW (1997) Physical and functional interactions between the transcriptional inhibitors Id3 and ITF-2b. Evidence toward a novel mechanism regulating muscle-specific gene expression. Journal of Biological Chemistry 272:2459-2463.
Chen DY, Stern SA, Garcia-Osta A, Saunier-Rebori B, Pollonini G, Bambah-Mukku D, Blitzer RD, Alberini CM (2011) A critical role for IGF-II in memory consolidation and enhancement. Nature 469:491-497.
Chen R, Amoui M, Zhang Z, Mardon G (1997) Dachshund and eyes absent proteins form a complex and function synergistically to induce ectopic eye development in Drosophila. Cell 91:893-903.
Chiaramello A, Soosaar A, Neuman T, Zuber MX (1995) Differential expression and distinct DNA-binding specificity of ME1a and ME2 suggest a unique role during differentiation and neuronal plasticity. Molecular Brain Research 29:107-118.
Cisse B, Caton ML, Lehner M, Maeda T, Scheu S, Locksley R, Holmberg D, Zweier C, den Hollander NS, Kant SG, Holter W, Rauch A, Zhuang Y, Reizis B (2008) Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. Cell 135:37-48.
Cobaleda C, Perez-Caro M, Vicente-Duenas C, Sanchez-Garcia I (2007) Function of the zinc-finger transcription factor SNAI2 in cancer and development. Annual Review of Genetics 41:41-61.
Corneliussen B, Thornell A, Hallberg B, Grundstrom T (1991) Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. Journal of Virology 65:6084-6093.
Corneliussen B, Holm M, Waltersson Y, Onions J, Hallberg B, Thornell A, Grundstrom T (1994) Calcium/calmodulin inhibition of basic-helix-loop-helix transcription factor domains. Nature 368:760-764.
Coulon A, Chow CC, Singer RH, Larson DR (2013) Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nature Reviews: Genetics 14:572-584.
Cromwell HC, Mears RP, Wan L, Boutros NN (2008) Sensory gating: a translational effort from basic to clinical science. Clin EEG Neurosci 39:69-72.
Cross-Disorder Group of the Psychiatric Genomics C et al. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature Genetics 45:984-994.
Dai YS, Cserjesi P (2002) The basic helix-loop-helix factor, HAND2, functions as a transcriptional activator by binding to E-boxes as a heterodimer. Journal of Biological Chemistry 277:12604-12612.
Davis RL, Cheng PF, Lassar AB, Weintraub H (1990) The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell 60:733-746.
de Pontual L et al. (2009) Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. Human Mutation 30:669-676.
de Pooter RF, Kee BL (2010) E proteins and the regulation of early lymphocyte development. Immunological Reviews 238:93-109.
Degorce F, Card A, Soh S, Trinquet E, Knapik GP, Xie B (2009) HTRF: A technology tailored for drug discovery - a review of theoretical aspects and recent applications. Current Chemical Genomics 3:22-32.

Dehay C, Kennedy H (2007) Cell-cycle control and cortical development. Nature Reviews: Neuroscience 8:438-450.
Denis CM, Chitayat S, Plevin MJ, Wang F, Thompson P, Liu S, Spencer HL, Ikura M, LeBrun DP, Smith SP (2012) Structural basis of CBP/p300 recruitment in leukemia induction by E2A-PBX1. Blood 120:3968-3977.
Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, Geczy CL (2013) Functions of S100 proteins. Current Molecular Medicine 13:24-57.
Dorflinger U, Pscherer A, Moser M, Rummele P, Schule R, Buettner R (1999) Activation of somatostatin receptor II expression by transcription factors MIBP1 and SEF-2 in the murine brain. Molecular and Cellular Biology 19:3736-3747.
Eastwood SL, Law AJ, Everall IP, Harrison PJ (2003) The axonal chemorepellant semaphorin 3A is increased in the cerebellum in schizophrenia and may contribute to its synaptic pathology. Molecular Psychiatry 8:148-155.
El Ghouzzi V, Legeai-Mallet L, Aresta S, Benoist C, Munnich A, de Gunzburg J, Bonaventure J (2000) Saethre-Chotzen mutations cause TWIST protein degradation or impaired nuclear location. Human Molecular Genetics 9:813-819.
Elia J et al. (2010) Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Molecular Psychiatry 15:637-646.
Ellenberger T, Fass D, Arnaud M, Harrison SC (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes & Development 8:970-980.
Engel I, Murre C (2004) E2A proteins enforce a proliferation checkpoint in developing thymocytes. EMBO Journal 23:202-211.
Famulski JK, Solecki DJ (2013) New spin on an old transition: epithelial parallels in neuronal adhesion control. Trends in Neurosciences 36:163-173.
Fang X, Cai Y, Liu J, Wang Z, Wu Q, Zhang Z, Yang CJ, Yuan L, Ouyang G (2011) Twist2 contributes to breast cancer progression by promoting an epithelial-mesenchymal transition and cancer stem-like cell self-renewal. Oncogene 30:4707-4720.
Farah MH, Olson JM, Sucic HB, Hume RI, Tapscott SJ, Turner DL (2000) Generation of neurons by transient expression of neural bHLH proteins in mammalian cells. Development 127:693-702.
Farkas LM, Huttner WB (2008) The cell biology of neural stem and progenitor cells and its significance for their proliferation versus differentiation during mammalian brain development. Current Opinion in Cell Biology 20:707-715.
Fatemi SH, Earle JA, McMenomy T (2000) Reduction in Reelin immunoreactivity in hippocampus of subjects with schizophrenia, bipolar disorder and major depression. Molecular Psychiatry 5:654-663, 571.
Fernandez AM, Torres-Aleman I (2012) The many faces of insulin-like peptide signalling in the brain. Nature Reviews Neuroscience 13:225-239.
Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391:806-811.
Firulli BA, Krawchuk D, Centonze VE, Vargesson N, Virshup DM, Conway SJ, Cserjesi P, Laufer E, Firulli AB (2005) Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. Nature Genetics 37:373-381.

Flavell SW, Kim TK, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, Greenberg ME (2008) Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron 60:1022-1038.

Flora A, Garcia JJ, Thaller C, Zoghbi HY (2007) The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. Proceedings of the National Academy of Sciences of the United States of America 104:15382-15387.

Fode C, Ma Q, Casarosa S, Ang SL, Anderson DJ, Guillemot F (2000) A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. Genes & Development 14:67-80.

Forrest M, Chapman RM, Doyle AM, Tinsley CL, Waite A, Blake DJ (2012) Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome. Human Mutation 33:1676-1686.

Forrest MP, Waite AJ, Martin-Rendon E, Blake DJ (2013) Knockdown of Human TCF4 Affects Multiple Signaling Pathways Involved in Cell Survival, Epithelial to Mesenchymal Transition and Neuronal Differentiation. PLoS One 8:e73169.

Franco SJ, Martinez-Garay I, Gil-Sanz C, Harkins-Perry SR, Muller U (2011) Reelin regulates cadherin function via Dab1/Rap1 to control neuronal migration and lamination in the neocortex. Neuron 69:482-497.

Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, Knoers NV, Cahn W, Kahn RS, Edelmann L, Davis KL, Silverman JM, Brunner HG, van Kessel AG, Wijmenga C, Ophoff RA, Veltman JA (2008) CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Molecular psychiatry 13:261-266.

Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nature Reviews: Genetics 13:840-852.

Galej WP, Oubridge C, Newman AJ, Nagai K (2013) Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 493:638-643.

Gao Z, Ure K, Ables JL, Lagace DC, Nave KA, Goebbels S, Eisch AJ, Hsieh J (2009) Neurod1 is essential for the survival and maturation of adult-born neurons. Nature Neuroscience 12:1090-1092.

Garavelli L, Mainardi PC (2007) Mowat-Wilson syndrome. Orphanet Journal of Rare Diseases 2:42.

Ghosh HS, Cisse B, Bunin A, Lewis KL, Reizis B (2010) Continuous expression of the transcription factor e2-2 maintains the cell fate of mature plasmacytoid dendritic cells. Immunity 33:905-916.

Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M, Vitkup D (2012) Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. Nature Neuroscience 15:1723-1728.

Girard SL et al. (2011) Increased exonic de novo mutation rate in individuals with schizophrenia. Nature Genetics 43:860-863.

Giurgea I, Missirian C, Cacciagli P, Whalen S, Fredriksen T, Gaillon T, Rankin J, Mathieu-Dramard M, Morin G, Martin-Coignard D, Dubourg C, Chabrol B, Arfi J, Giuliano F, Claude Lambert J, Philip N, Sarda P, Villard L, Goossens M, Moncla A (2008) TCF4 deletions in Pitt-Hopkins Syndrome. Human Mutation 29:E242-251.

Goldfarb AN, Lewandowska K, Pennell CA (1998) Identification of a highly conserved module in E proteins required for in vivo helix-loop-helix dimerization. Journal of Biological Chemistry 273:2866-2873.
Goldman AL, Pezawas L, Mattay VS, Fischl B, Verchinski BA, Chen Q, Weinberger DR, Meyer-Lindenberg A (2009) Widespread reductions of cortical thickness in schizophrenia and spectrum disorders and evidence of heritability. Archives of General Psychiatry 66:467-477.
Guenther MG, Barak O, Lazar MA (2001) The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. Molecular and Cellular Biology 21:6091-6101.
Guidotti A, Auta J, Davis JM, Di-Giorgi-Gerevini V, Dwivedi Y, Grayson DR, Impagnatiello F, Pandey G, Pesold C, Sharma R, Uzunov D, Costa E (2000) Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study. Archives of General Psychiatry 57:1061-1069.
Guillemot F, Joyner AL (1993) Dynamic expression of the murine Achaete-Scute homologue Mash-1 in the developing nervous system. Mechanisms of Development 42:171-185.
Guo C, Hu Q, Yan C, Zhang J (2009) Multivalent binding of the ETO corepressor to E proteins facilitates dual repression controls targeting chromatin and the basal transcription machinery. Molecular and Cellular Biology 29:2644-2657.
Gutierrez H, Davies AM (2011) Regulation of neural process growth, elaboration and structural plasticity by NF-kappaB. Trends in Neurosciences 34:316-325.
Hamdan FF, Daoud H, Patry L, Dionne-Laporte A, Spiegelman D, Dobrzeniecka S, Rouleau GA, Michaud JL (2013) Parent-child exome sequencing identifies a de novo truncating mutation in TCF4 in non-syndromic intellectual disability. Clinical Genetics 83:198-200.
Hauser J, Saarikettu J, Grundstrom T (2008a) Calcium regulation of myogenesis by differential calmodulin inhibition of basic helix-loop-helix transcription factors. Molecular Biology of the Cell 19:2509-2519.
Hauser J, Sveshnikova N, Wallenius A, Baradaran S, Saarikettu J, Grundstrom T (2008b) B-cell receptor activation inhibits AID expression through calmodulin inhibition of E-proteins. Proceedings of the National Academy of Sciences of the United States of America 105:1267-1272.
Heng JI, Nguyen L, Castro DS, Zimmer C, Wildner H, Armant O, Skowronska-Krawczyk D, Bedogni F, Matter JM, Hevner R, Guillemot F (2008) Neurogenin 2 controls cortical neuron migration through regulation of Rnd2. Nature 455:114-118.
Henthorn P, Kiledjian M, Kadesch T (1990) Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. Science 247:467-470.
Herbst A, Kolligs FT (2008) A conserved domain in the transcription factor ITF-2B attenuates its activity. Biochemical and Biophysical Research Communications 370:327-331.
Herbst A, Bommer GT, Kriegl L, Jung A, Behrens A, Csanadi E, Gerhard M, Bolz C, Riesenberg R, Zimmermann W, Dietmaier W, Wolf I, Brabletz T, Goke B, Kolligs FT (2009) ITF-2 is disrupted via allelic loss of chromosome 18q21, and ITF-2B expression is lost at the adenoma-carcinoma transition. Gastroenterology 137:639-648, 648 e631-639.

Hermann S, Saarikettu J, Onions J, Hughes K, Grundstrom T (1998) Calcium regulation of basic helix-loop-helix transcription factors. Cell Calcium 23:135-142.
Hintsch G, Zurlinden A, Meskenaite V, Steuble M, Fink-Widmer K, Kinter J, Sonderegger P (2002) The calsyntenins--a family of postsynaptic membrane proteins with distinct neuronal expression patterns. Molecular and Cellular Neuroscience 21:393-409.
Hirabayashi Y, Gotoh Y (2010) Epigenetic control of neural precursor cell fate during development. Nature Reviews: Neuroscience 11:377-388.
Hong SE, Shugart YY, Huang DT, Shahwan SA, Grant PE, Hourihane JO, Martin ND, Walsh CA (2000) Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN mutations. Nature Genetics 26:93-96.
Howard TD, Paznekas WA, Green ED, Chiang LC, Ma N, Ortiz de Luna RI, Garcia Delgado C, Gonzalez-Ramos M, Kline AD, Jabs EW (1997) Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. Nature Genetics 15:36-41.
Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4:44-57.
Hug BA, Lazar MA (2004) ETO interacting proteins. Oncogene 23:4270-4274.
Hur EM, Zhou FQ (2010) GSK3 signalling in neural development. Nature Reviews: Neuroscience 11:539-551.
Huttner WB, Brand M (1997) Asymmetric division and polarity of neuroepithelial cells. Current Opinion in Neurobiology 7:29-39.
Hyman BT, Yuan J (2012) Apoptotic and non-apoptotic roles of caspases in neuronal physiology and pathophysiology. Nature Reviews Neuroscience 13:395-406.
Igo RP, Jr., Kopplin LJ, Joseph P, Truitt B, Fondran J, Bardenstein D, Aldave AJ, Croasdale CR, Price MO, Rosenwasser M, Lass JH, Iyengar SK, Group FGM-cS (2012) Differing roles for TCF4 and COL8A2 in central corneal thickness and fuchs endothelial corneal dystrophy. PLoS One 7:e46742.
Ik Tsen Heng J, Tan SS (2003) The role of class I HLH genes in neural development--have they been overlooked? Bioessays 25:709-716.
Impagnatiello F, Guidotti AR, Pesold C, Dwivedi Y, Caruncho H, Pisu MG, Uzunov DP, Smalheiser NR, Davis JM, Pandey GN, Pappas GD, Tueting P, Sharma RP, Costa E (1998) A decrease of reelin expression as a putative vulnerability factor in schizophrenia. Proceedings of the National Academy of Sciences of the United States of America 95:15718-15723.
Insel TR (2010) Rethinking schizophrenia. Nature 468:187-193.
International Schizophrenia C (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455:237-241.
Ishii S, Kurasawa Y, Wong J, Yu-Lee LY (2008) Histone deacetylase 3 localizes to the mitotic spindle and is required for kinetochore-microtubule attachment. Proceedings of the National Academy of Sciences of the United States of America 105:4179-4184.
Itoh Y, Moriyama Y, Hasegawa T, Endo TA, Toyoda T, Gotoh Y (2013) Scratch regulates neuronal migration onset via an epithelial-mesenchymal transition-like mechanism. Nature Neuroscience 16:416-425.
Jepsen K, Rosenfeld MG (2002) Biological roles and mechanistic actions of co-repressor complexes. Journal of Cell Science 115:689-698.

Jogi A, Persson P, Grynfeld A, Pahlman S, Axelson H (2002) Modulation of basic helix-loop-helix transcription complex formation by Id proteins during neuronal differentiation. Journal of Biological Chemistry 277:9118-9126.
Johnson SE, Wang X, Hardy S, Taparowsky EJ, Konieczny SF (1996) Casein kinase II increases the transcriptional activities of MRF4 and MyoD independently of their direct phosphorylation. Molecular and Cellular Biology 16:1604-1613.
Jossin Y, Cooper JA (2011) Reelin, Rap1 and N-cadherin orient the migration of multipolar neurons in the developing neocortex. Nature Neuroscience 14:697-703.
Kageyama R, Ohtsuka T, Kobayashi T (2007) The Hes gene family: repressors and oscillators that orchestrate embryogenesis. Development 134:1243-1251.
Kalluri R, Weinberg RA (2009) The basics of epithelial-mesenchymal transition. Journal of Clinical Investigation 119:1420-1428.
Kalscheuer VM, Feenstra I, Van Ravenswaaij-Arts CM, Smeets DF, Menzel C, Ullmann R, Musante L, Ropers HH (2008) Disruption of the TCF4 gene in a girl with mental retardation but without the classical Pitt-Hopkins syndrome. American Journal of Medical Genetics Part A 146A:2053-2059.
Kato A, Okamoto O, Ishikawa K, Sumiyoshi H, Matsuo N, Yoshioka H, Nomizu M, Shimada T, Fujiwara S (2011) Dermatopontin interacts with fibronectin, promotes fibronectin fibril formation, and enhances cell adhesion. Journal of Biological Chemistry 286:14861-14869.
Kavanagh DH, Dwyer S, O'Donovan MC, Owen MJ (2013) The ENCODE project: implications for psychiatric genetics. Molecular Psychiatry 18:540-542.
Kee BL (2009) E and ID proteins branch out. Nature Reviews: Immunology 9:175-184.
Keshavan MS, Tandon R, Boutros NN, Nasrallah HA (2008) Schizophrenia, "just the facts": what we know in 2008 Part 3: neurobiology. Schizophr Research 106:89-107.
Kim DH, Behlke MA, Rose SD, Chang MS, Choi S, Rossi JJ (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. Nature Biotechnology 23:222-226.
Kim EK, Choi EJ (2010) Pathological roles of MAPK signaling pathways in human diseases. Biochimica et Biophysica Acta (BBA) - Reviews in Biomembranes 1802:396-405.
Kim S, Cho H, Lee D, Webster MJ (2012) Association between SNPs and gene expression in multiple regions of the human brain. Translational Psychiatry 2:e113.
Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ (2009) Neurexin 1 (NRXN1) deletions in schizophrenia. Schizophrenia Bulletin 35:851-854.
Kirov G et al. (2012) De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Molecular Psychiatry 17:142-153.
Knapp M, Mangalore R, Simon J (2004) The global costs of schizophrenia. Schizophrenia Bulletin 30:279-293.
Knippschild U, Gocht A, Wolff S, Huber N, Lohler J, Stoter M (2005) The casein kinase 1 family: participation in multiple cellular processes in eukaryotes. Cellular Signalling 17:675-689.
Kolligs FT, Nieman MT, Winer I, Hu G, Van Mater D, Feng Y, Smith IM, Wu R, Zhai Y, Cho KR, Fearon ER (2002) ITF-2, a downstream target of the

Wnt/TCF pathway, is activated in human cancers with beta-catenin defects and promotes neoplastic transformation. Cancer Cell 1:145-155.
Kousoulidou L, Tanteles G, Moutafi M, Sismani C, Patsalis PC, Anastasiadou V (2013) 263.4 kb deletion within the TCF4 gene consistent with Pitt-Hopkins syndrome, inherited from a mosaic parent with normal phenotype. European Journal of Medical Genetics 56:314-318.
Kurusu M, Nagao T, Walldorf U, Flister S, Gehring WJ, Furukubo-Tokunaga K (2000) Genetic control of development of the mushroom bodies, the associative learning centers in the Drosophila brain, by the eyeless, twin of eyeless, and Dachshund genes. Proceedings of the National Academy of Sciences of the United States of America 97:2140-2144.
Larsen MR, Sorensen GL, Fey SJ, Larsen PM, Roepstorff P (2001) Phospho-proteomics: evaluation of the use of enzymatic de-phosphorylation and differential mass spectrometric peptide mass mapping for site specific phosphorylation assignment in proteins separated by gel electrophoresis. Proteomics 1:223-238.
Larsson G, Schleucher J, Onions J, Hermann S, Grundstrom T, Wijmenga SS (2001) A novel target recognition revealed by calmodulin in complex with the basic helix--loop--helix transcription factor SEF2-1/E2-2. Protein Science 10:169-186.
Larsson G, Schleucher J, Onions J, Hermann S, Grundstrom T, Wijmenga SS (2005) Backbone dynamics of a symmetric calmodulin dimer in complex with the calmodulin-binding domain of the basic-helix-loop-helix transcription factor SEF2-1/E2-2: a highly dynamic complex. Biophysical Journal 89:1214-1226.
Le Guen T, Fichou Y, Nectoux J, Bahi-Buisson N, Rivier F, Boddaert N, Diebold B, Heron D, Chelly J, Bienvenu T (2011) A missense mutation within the fork-head domain of the forkhead box G1 Gene (FOXG1) affects its nuclear localization. Human Mutation 32:E2026-2035.
Le Meur N, Holder-Espinasse M, Jaillard S, Goldenberg A, Joriot S, Amati-Bonneau P, Guichet A, Barth M, Charollais A, Journel H, Auvin S, Boucher C, Kerckaert JP, David V, Manouvrier-Hanu S, Saugier-Veber P, Frebourg T, Dubourg C, Andrieux J, Bonneau D (2010) MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. Journal of Medical Genetics 47:22-29.
Ledent V, Paquet O, Vervoort M (2002) Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biology 3:RESEARCH0030.
Lee JK, Cho JH, Hwang WS, Lee YD, Reu DS, Suh-Kim H (2000) Expression of neuroD/BETA2 in mitotic and postmitotic neuronal cells during the development of nervous system. Developmental Dynamics 217:361-367.
Lee SH, DeCandia TR, Ripke S, Yang J, Schizophrenia Psychiatric Genome-Wide Association Study C, International Schizophrenia C, Molecular Genetics of Schizophrenia C, Sullivan PF, Goddard ME, Keller MC, Visscher PM, Wray NR (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nature Genetics 44:247-250.
Lehalle D, Williams C, Siu VM, Clayton-Smith J (2011) Fetal pads as a clue to the diagnosis of Pitt-Hopkins syndrome. American Journal of Medical Genetics Part A 155A:1685-1689.
Lennertz L, Rujescu D, Wagner M, Frommann I, Schulze-Rauschenbach S, Schuhmacher A, Landsberg MW, Franke P, Moller HJ, Wolwer W, Gaebel W,

Hafner H, Maier W, Mossner R (2011) Novel schizophrenia risk gene TCF4 influences verbal learning and memory functioning in schizophrenia patients. Neuropsychobiology 63:131-136.
Levinson DF et al. (2011) Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. American Journal of Psychiatry 168:302-316.
Lewis DA, Levitt P (2002) Schizophrenia as a disorder of neurodevelopment. Annual Review of Neuroscience 25:409-432.
Li HS, Yang CY, Nallaparaju KC, Zhang H, Liu YJ, Goldrath AW, Watowich SS (2012a) The signal transducers STAT5 and STAT3 control expression of Id2 and E2-2 during dendritic cell development. Blood 120:4363-4373.
Li L, Cserjesi P, Olson EN (1995) Dermo-1: a novel twist-related bHLH protein expressed in the developing dermis. Developmental Biology 172:280-292.
Li L, Du Y, Li N, Wu X, Wu Y (2009) Top-down modulation of prepulse inhibition of the startle reflex in humans and rats. Neuroscience & Biobehavioral Reviews 33:1157-1167.
Li S, Mattar P, Zinyk D, Singh K, Chaturvedi CP, Kovach C, Dixit R, Kurrasch DM, Ma YC, Chan JA, Wallace V, Dilworth FJ, Brand M, Schuurmans C (2012b) GSK3 temporally regulates neurogenin 2 proneural activity in the neocortex. The Journal of Neuroscience 32:7791-7805.
Lijffijt M, Lane SD, Meier SL, Boutros NN, Burroughs S, Steinberg JL, Moeller FG, Swann AC (2009) P50, N100, and P200 sensory gating: relationships with behavioral inhibition, attention, and working memory. Psychophysiology 46:1059-1068.
Lim J, Thiery JP (2012) Epithelial-mesenchymal transitions: insights from development. Development 139:3471-3486.
Lim S, Kaldis P (2013) Cdks, cyclins and CKIs: roles beyond cell cycle regulation. Development 140:3079-3093.
Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, International Schizophrenia C, Holmans PA, O'Donovan MC, Purcell SM, Smit AB, Verhage M, Sullivan PF, Visscher PM, Posthuma D (2012) Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Molecular Psychiatry 17:996-1006.
Liu X, Meng L, Shi Q, Liu S, Cui C, Hu S, Wei Y (2013) Dermatopontin promotes adhesion, spreading and migration of cardiac fibroblasts in vitro. Matrix Biology 32:23-31.
Liu Y, Ray SK, Yang XQ, Luntz-Leybman V, Chiu IM (1998) A splice variant of E2-2 basic helix-loop-helix protein represses the brain-specific fibroblast growth factor 1 promoter through the binding to an imperfect E-box. The Journal of Biological Chemistry 273:19269-19276.
Liu YP, Burleigh D, Durning M, Hudson L, Chiu IM, Golos TG (2004) Id2 is a primary partner for the E2-2 basic helix-loop-helix transcription factor in the human placenta. Molecular and Cellular Endocrinology 222:83-91.
Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature Genetics 33:177-182.
Longo A, Guanga GP, Rose RB (2008) Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. Biochemistry 47:218-229.

Lyden D, Young AZ, Zagzag D, Yan W, Gerald W, O'Reilly R, Bader BL, Hynes RO, Zhuang Y, Manova K, Benezra R (1999) Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. Nature 401:670-677.
Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, Rappsilber J, Greenberg ME, Bird A (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. Nature Neuroscience 16:898-902.
Mabb AM, Judson MC, Zylka MJ, Philpot BD (2011) Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. Trends in Neurosciences 34:293-303.
Mackey DA, Warrington NM, Hewitt AW, Oates SK, Yazar S, Soloshenko A, Crawford GJ, Mountain JA, Pennell CE (2012) Role of the TCF4 gene intronic variant in normal variation of corneal endothelium. Cornea 31:162-166.
Mao YS, Zhang B, Spector DL (2011) Biogenesis and function of nuclear bodies. Trends in Genetics 27:295-306.
Marangi G et al. (2011) The Pitt-Hopkins syndrome: report of 16 new patients and clinical diagnostic criteria. American Journal of Medical Genetics Part A 155A:1536-1545.
Markus M, Du Z, Benezra R (2002a) Enhancer-specific modulation of E protein activity. Journal of Biological Chemistry 277:6469-6477.
Markus M, Du Z, Benezra R (2002b) Enhancer-specific modulation of E protein activity. The Journal of biological chemistry 277:6469-6477.
Massague J (2012) TGFbeta signalling in context. Nature Reviews: Molecular Cell Biology 13:616-630.
Massari ME, Murre C (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Molecular and Cellular Biology 20:429-440.
Massari ME, Jennings PA, Murre C (1996) The AD1 transactivation domain of E2A contains a highly conserved helix which is required for its activity in both Saccharomyces cerevisiae and mammalian cells. Molecular and Cellular Biology 16:121-129.
Massari ME, Grant PA, Pray-Grant MG, Berger SL, Workman JL, Murre C (1999) A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. Molecular Cell 4:63-73.
McCarthy FM, Bridges SM, Wang N, Magee GB, Williams WP, Luthe DS, Burgess SC (2007) AgBase: a unified resource for functional analysis in agriculture. Nucleic Acids Research 35:D599-603.
McClellan JM, Susser E, King MC (2007) Schizophrenia: a common disease caused by multiple rare alleles. British Journal of Psychiatry 190:194-199.
McDonald JJ, Alinikula J, Buerstedde JM, Schatz DG (2013) A critical context-dependent role for e boxes in the targeting of somatic hypermutation. Journal of Immunology 191:1556-1566.
McKinsey GL, Lindtner S, Trzcinski B, Visel A, Pennacchio LA, Huylebroeck D, Higashi Y, Rubenstein JL (2013) Dlx1&2-Dependent Expression of Zfhx1b (Sip1, Zeb2) Regulates the Fate Switch between Cortical and Striatal Interneurons. Neuron 77:83-98.
Mitchell KJ, Porteous DJ (2011) Rethinking the genetic architecture of schizophrenia. Psychological Medicine 41:19-32.

Miyata T, Maeda T, Lee JE (1999) NeuroD is required for differentiation of the granule cells in the cerebellum and hippocampus. Genes & Development 13:1647-1652.
Molkentin JD, Olson EN (1996) Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. Proceedings of the National Academy of Sciences of the United States of America 93:9366-9373.
Moreno-Bueno G, Cubillo E, Sarrio D, Peinado H, Rodriguez-Pinilla SM, Villa S, Bolos V, Jorda M, Fabra A, Portillo F, Palacios J, Cano A (2006) Genetic profiling of epithelial cells expressing E-cadherin repressors reveals a distinct role for Snail, Slug, and E47 factors in epithelial-mesenchymal transition. Cancer Research 66:9543-9556.
Murakami M, Kataoka K, Tominaga J, Nakagawa O, Kurihara H (2004) Differential cooperation between dHAND and three different E-proteins. Biochemical and Biophysical Research Communications 323:168-174.
Murayama A, Kim MS, Yanagisawa J, Takeyama K, Kato S (2004) Transrepression by a liganded nuclear receptor via a bHLH activator through co-regulator switching. EMBO Journal 23:1598-1608.
Murre C, McCaw PS, Baltimore D (1989a) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell 56:777-783.
Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, et al. (1989b) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 58:537-544.
Navarrete K, Pedroso I, De Jong S, Stefansson H, Steinberg S, Stefansson K, Ophoff RA, Schalkwyk LC, Collier DA (2013) TCF4 (e2-2; ITF2): A schizophrenia-associated gene with pleiotropic effects on human disease. American Journal of Medical Genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 162:1-16.
Need AC et al. (2012) Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. American Journal of Human Genetics 91:303-312.
Neuman T, Keen A, Knapik E, Shain D, Ross M, Nornes HO, Zuber MX (1993) ME1 and GE1: basic helix-loop-helix transcription factors expressed at high levels in the developing nervous system and in morphogenetically active regions. European Journal of Neuroscience 5:311-318.
Nieto MA (2011) The ins and outs of the epithelial to mesenchymal transition in health and disease. Annual Review of Cell and Developmental Biology 27:347-376.
O'Driscoll M, Jeggo PA (2006) The role of double-strand break repair - insights from human genetics. Nature Reviews: Genetics 7:45-54.
O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, International Schizophrenia C, Corvin A (2011) Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Molecular Psychiatry 16:286-292.
Ohnuma S, Harris WA (2003) Neurogenesis and the cell cycle. Neuron 40:199-208.
Ohsako S, Hyer J, Panganiban G, Oliver I, Caudy M (1994) Hairy function as a DNA-binding helix-loop-helix repressor of Drosophila sensory organ formation. Genes & Development 8:2743-2755.

Okoniewski MJ, Miller CJ (2008) Comprehensive analysis of affymetrix exon arrays using BioConductor. PLoS Computational Biology 4:e6.
Onions J, Hermann S, Grundstrom T (1997) Basic helix-loop-helix protein sequences determining differential inhibition by calmodulin and S-100 proteins. Journal of Biological Chemistry 272:23930-23937.
Pacary E, Martynoga B, Guillemot F (2012) Crucial first steps: the transcriptional control of neuron delamination. Neuron 74:209-211.
Pagliuca A, Gallo P, De Luca P, Lania L (2000) Class A helix-loop-helix proteins are positive regulators of several cyclin-dependent kinase inhibitors' promoter activity and negatively affect cell growth. Cancer Research 60:1376-1382.
Panchision DM, Pickel JM, Studer L, Lee SH, Turner PA, Hazel TG, McKay RD (2001) Sequential actions of BMP receptors control neural precursor cell production and fate. Genes & Development 15:2094-2110.
Parsons JT, Horwitz AR, Schwartz MA (2010) Cell adhesion: integrating cytoskeletal dynamics and cellular tension. Nature Reviews: Molecular Cell Biology 11:633-643.
Pasterkamp RJ (2012) Getting neural circuits into shape with semaphorins. Nature Reviews: Neuroscience 13:605-618.
Peippo M, Ignatius J (2012) Pitt-Hopkins Syndrome. Molecular syndromology 2:171-180.
Penagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, Sonnenblick LI, Gruver R, Almajano J, Bragin A, Golshani P, Trachtenberg JT, Peles E, Geschwind DH (2011) Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell 147:235-246.
Perez-Moreno MA, Locascio A, Rodrigo I, Dhondt G, Portillo F, Nieto MA, Cano A (2001) A new role for E12/E47 in the repression of E-cadherin expression and epithelial-mesenchymal transitions. Journal of Biological Chemistry 276:27424-27431.
Persson P, Jogi A, Grynfeld A, Pahlman S, Axelson H (2000) HASH-1 and E2-2 are expressed in human neuroblastoma cells and form a functional complex. Biochemical and Biophysical Research Communications 274:22-31.
Petropoulos H, Skerjanc IS (2000) Analysis of the inhibition of MyoD activity by ITF-2B and full-length E12/E47. Journal of Biological Chemistry 275:25095-25101.
Peverali FA, Ramqvist T, Saffrich R, Pepperkok R, Barone MV, Philipson L (1994) Regulation of G1 progression by E2A and Id helix-loop-helix proteins. EMBO Journal 13:4291-4301.
Phillips SE (1994) Built by association: structure and function of helix-loop-helix DNA-binding proteins. Structure 2:1-4.
Piton A, Redin C, Mandel JL (2013) XLID-Causing Mutations and Associated Genes Challenged in Light of Data From Large-Scale Human Exome Sequencing. American Journal of Human Genetics.
Pitt D, Hopkins I (1978) A syndrome of mental retardation, wide mouth and intermittent overbreathing. Australian Paediatric Journal 14:182-184.
Powell LM, Jarman AP (2008) Context dependence of proneural bHLH proteins. Current Opinion in Genetics & Development 18:411-417.
Pscherer A, Dorflinger U, Kirfel J, Gawlas K, Ruschoff J, Buettner R, Schule R (1996) The helix-loop-helix transcription factor SEF-2 regulates the activity of

a novel initiator element in the promoter of the human somatostatin receptor II gene. EMBO Journal 15:6680-6690.
Qin ZH, Tao LY, Chen X (2007) Dual roles of NF-kappaB in cell survival and implications of NF-kappaB inhibitors in neuroprotective therapy. Acta Pharmacologica Sinica 28:1859-1872.
Quednow BB, Ettinger U, Mossner R, Rujescu D, Giegling I, Collier DA, Schmechtig A, Kuhn KU, Moller HJ, Maier W, Wagner M, Kumari V (2011) The schizophrenia risk allele C of the TCF4 rs9960767 polymorphism disrupts sensorimotor gating in schizophrenia spectrum and healthy volunteers. The Journal of Neuroscience 31:6684-6691.
Quednow BB et al. (2012) Schizophrenia risk polymorphisms in the TCF4 gene interact with smoking in the modulation of auditory sensory gating. Proceedings of the National Academy of Sciences of the United States of America 109:6271-6276.
Quong MW, Massari ME, Zwart R, Murre C (1993) A new transcriptional-activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells. Molecular and Cellular Biology 13:792-800.
Rauch A et al. (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380:1674-1682.
Ravanpay AC, Olson JM (2008) E protein dosage influences brain development more than family member identity. Journal of Neuroscience Research 86:1472-1481.
Ravasi T et al. (2010) An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140:744-752.
Rebola N, Srikumar BN, Mulle C (2010) Activity-dependent synaptic plasticity of NMDA receptors. Journal of Physiology 588:93-99.
Reiner A, Yekutieli D, Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 19:368-375.
Reizis B (2010) Regulation of plasmacytoid dendritic cell development. Current Opinion in Immunology 22:206-211.
Remacle JE, Kraft H, Lerchner W, Wuytens G, Collart C, Verschueren K, Smith JC, Huylebroeck D (1999) New mode of DNA binding of multi-zinc finger transcription factors: deltaEF1 family members bind with two hands to two target sites. The EMBO journal 18:5073-5084.
Ripke S et al. (2013) Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nature Genetics.
Ripke S et al. (2011) Genome-wide association study identifies five new schizophrenia loci. Nature Genetics 43:969-976.
Rodriguez-Murillo L, Gogos JA, Karayiorgou M (2012) The genetic architecture of schizophrenia: new mutations and emerging paradigms. Annual Review of Medicine 63:63-80.
Ronan JL, Wu W, Crabtree GR (2013) From neural development to cognition: unexpected roles for chromatin. Nature Reviews: Genetics 14:347-359.
Rose MF, Ren J, Ahmad KA, Chao HT, Klisch TJ, Flora A, Greer JJ, Zoghbi HY (2009) Math1 is essential for the development of hindbrain neurons critical for perinatal breathing. Neuron 64:341-354.
Rosenfeld JA, Leppig K, Ballif BC, Thiese H, Erdie-Lalena C, Bawle E, Sastry S, Spence JE, Bandholz A, Surti U, Zonana J, Keller K, Meschino W, Bejjani

BA, Torchia BS, Shaffer LG (2009) Genotype-phenotype analysis of TCF4 mutations causing Pitt-Hopkins syndrome shows increased seizure activity with missense mutations. Genetics in Medicine 11:797-805.
Ross SE, Greenberg ME, Stiles CD (2003) Basic helix-loop-helix factors in cortical development. Neuron 39:13-25.
Rossi M, Labalme A, Cordier MP, Till M, Blanchard G, Dubois R, Guibaud L, Heissat S, Javouhey E, Lachaux A, Mure PY, Ville D, Edery P, Sanlaville D (2012) Mosaic 18q21.2 deletions including the TCF4 gene: a clinical report. American Journal of Medical Genetics Part A 158A:3174-3181.
Rothschild G, Zhao X, Iavarone A, Lasorella A (2006) E Proteins and Id2 converge on p57Kip2 to regulate cell cycle in neural cells. Molecular and Cellular Biology 26:4351-4361.
Rousso DL, Pearson CA, Gaber ZB, Miquelajauregui A, Li S, Portera-Cailliau C, Morrisey EE, Novitch BG (2012) Foxp-mediated suppression of N-cadherin regulates neuroepithelial character and progenitor maintenance in the CNS. Neuron 74:314-330.
Saha S, Chant D, McGrath J (2007) A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time? Archives of General Psychiatry 64:1123-1131.
Sasai Y, Kageyama R, Tagawa Y, Shigemoto R, Nakanishi S (1992) Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. Genes & Development 6:2620-2634.
Schmedt T, Silva MM, Ziaei A, Jurkunas U (2012) Molecular bases of corneal endothelial dystrophies. Experimental Eye Research 95:24-34.
Schmeisser MJ, Baumann B, Johannsen S, Vindedal GF, Jensen V, Hvalby OC, Sprengel R, Seither J, Maqbool A, Magnutzki A, Lattke M, Oswald F, Boeckers TM, Wirth T (2012) IkappaB kinase/nuclear factor kappaB-dependent insulin-like growth factor 2 (Igf2) expression regulates synapse formation and spine maturation via Igf2 receptor signaling. The Journal of Neuroscience 32:5688-5703.
Schwab MH, Bartholomae A, Heimrich B, Feldmeyer D, Druffel-Augustin S, Goebbels S, Naya FJ, Zhao S, Frotscher M, Tsai MJ, Nave KA (2000) Neuronal basic helix-loop-helix proteins (NEX and BETA2/Neuro D) regulate terminal granule cell differentiation in the hippocampus. Journal of Neuroscience 20:3714-3724.
Schwartz R, Engel I, Fallahi-Sichani M, Petrie HT, Murre C (2006) Gene expression patterns define novel roles for E47 in cell cycle progression, cytokine-mediated signaling, and T lineage development. Proceedings of the National Academy of Sciences of the United States of America 103:9976-9981.
Sebat J, Levy DL, McCarthy SE (2009) Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. Trends in Genetics 25:528-535.
Segklia A, Seuntjens E, Elkouris M, Tsalavos S, Stappers E, Mitsiadis TA, Huylebroeck D, Remboutsika E, Graf D (2012) Bmp7 regulates the survival, proliferation, and neurogenic properties of neural progenitor cells during corticogenesis in the mouse. PloS One 7:e34088.
Seo S, Richardson GA, Kroll KL (2005) The SWI/SNF chromatin remodeling protein Brg1 is required for vertebrate neurogenesis and mediates transactivation of Ngn and NeuroD. Development 132:105-115.

Sepp M, Pruunsild P, Timmusk T (2012a) Pitt-Hopkins syndrome-associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant-negative effects. Human Molecular Genetics.

Sepp M, Pruunsild P, Timmusk T (2012b) Pitt-Hopkins syndrome-associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant-negative effects. Human Molecular Genetics 21:2873-2888.

Sepp M, Kannike K, Eesmaa A, Urb M, Timmusk T (2011) Functional diversity of human basic helix-loop-helix transcription factor TCF4 isoforms generated by alternative 5' exon usage and splicing. PloS One 6:e22138.

Shain DH, Neuman T, Zuber MX (1997) Embryonic expression and evolution of duplicated E-protein genes in Xenopus laevis: parallels with ancestral E-protein genes. Genetics 146:345-353.

Skerjanc IS, Truong J, Filion P, McBurney MW (1996) A splice variant of the ITF-2 transcript encodes a transcription factor that inhibits MyoD activity. Journal of Biological Chemistry 271:3555-3561.

Sobrado VR, Moreno-Bueno G, Cubillo E, Holt LJ, Nieto MA, Portillo F, Cano A (2009) The class I bHLH factors E2-2A and E2-2B regulate EMT. Journal of Cell Science 122:1014-1024.

Soosaar A, Chiaramello A, Zuber MX, Neuman T (1994) Expression of basic-helix-loop-helix transcription factor ME2 during brain development and in the regions of neuronal plasticity in the adult brain. Molecular Brain Research 25:176-180.

Spicer DB, Rhee J, Cheung WL, Lassar AB (1996) Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. Science 272:1476-1480.

Stefansson H et al. (2009) Common variants conferring risk of schizophrenia. Nature 460:744-747.

Steinberg S et al. (2011) Common variants at VRK2 and TCF4 conferring risk of schizophrenia. Human Molecular Genetics 20:4076-4081.

Steinbusch CV, van Roozendaal KE, Tserpelis D, Smeets EE, Kranenburg-de Koning TJ, de Waal KH, Zweier C, Rauch A, Hennekam RC, Blok MJ, Schrander-Stumpel CT (2013) Somatic mosaicism in a mother of two children with Pitt-Hopkins syndrome. Clinical Genetics 83:73-77.

Stephan KE, Friston KJ, Frith CD (2009) Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. Schizophr Bulletin 35:509-527.

Stevens JD, Roalson EH, Skinner MK (2008) Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: genomic approach to cellular differentiation. Differentiation 76:1006-1022.

Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of General Psychiatry 60:1187-1192.

Sullivan PF, Daly MJ, O'Donovan M (2012) Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nature Reviews: Genetics 13:537-551.

Sun XH, Baltimore D (1991) An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell 64:459-470.

Sun XH, Copeland NG, Jenkins NA, Baltimore D (1991) Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. Molecular and Cellular Biology 11:5603-5611.
Tabuchi K, Blundell J, Etherton MR, Hammer RE, Liu X, Powell CM, Sudhof TC (2007) A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. Science 318:71-76.
Tadeu AM, Ribeiro S, Johnston J, Goldberg I, Gerloff D, Earnshaw WC (2008) CENP-V is required for centromere organization, chromosome alignment and cytokinesis. EMBO Journal 27:2510-2522.
Takano K, Lyons M, Moyes C, Jones J, Schwartz CE (2010) Two percent of patients suspected of having Angelman syndrome have TCF4 mutations. Clinical Genetics 78:282-288.
Takano K, Tan WH, Irons MB, Jones JR, Schwartz CE (2011) Pitt-Hopkins syndrome should be in the differential diagnosis for males presenting with an ATR-X phenotype. Clinical Genetics 80:600-601.
Talkowski ME et al. (2012) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell 149:525-537.
Tanaka A, Itoh F, Nishiyama K, Takezawa T, Kurihara H, Itoh S, Kato M (2010) Inhibition of endothelial cell activation by bHLH protein E2-2 and its impairment of angiogenesis. Blood 115:4138-4147.
Tandon R, Keshavan MS, Nasrallah HA (2008) Schizophrenia, "just the facts" what we know in 2008. 2. Epidemiology and etiology. Schizophr Res 102:1-18.
Tapscott SJ, Davis RL, Thayer MJ, Cheng PF, Weintraub H, Lassar AB (1988) MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science 242:405-411.
Taus T, Kocher T, Pichler P, Paschke C, Schmidt A, Henrich C, Mechtler K (2011) Universal and confident phosphorylation site localization using phosphoRS. Journal of Proteome Research 10:5354-5362.
Teachenor R, Beck K, Wright LY, Shen Z, Briggs SP, Murre C (2012) Biochemical and phosphoproteomic analysis of the helix-loop-helix protein E47. Molecular and Cellular Biology 32:1671-1682.
Tomita K, Kubo K, Ishii K, Nakajima K (2011) Disrupted-in-Schizophrenia-1 (Disc1) is necessary for migration of the pyramidal neurons during mouse hippocampal development. Human Molecular Genetics 20:2834-2845.
Toulopoulouand T, Murray RM (2004) Verbal memory deficit in patients with schizophrenia: an important future target for treatment. Expert Review of Neurotherapeutics 4:43-52.
Tropea D, Giacometti E, Wilson NR, Beard C, McCurry C, Fu DD, Flannery R, Jaenisch R, Sur M (2009) Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. Proceedings of the National Academy of Sciences of the United States of America 106:2029-2034.
Tsuang M (2000) Schizophrenia: genes and environment. Biological Psychiatry 47:210-220.
Tsurusaki Y et al. (2012) Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. Nature Genetics 44:376-378.
Tukel T, Sosic D, Al-Gazali LI, Erazo M, Casasnovas J, Franco HL, Richardson JA, Olson EN, Cadilla CL, Desnick RJ (2010) Homozygous nonsense mutations in TWIST2 cause Setleis syndrome. American Journal of Human Genetics 87:289-296.

Turner DL, Weintraub H (1994) Expression of achaete-scute homolog 3 in Xenopus embryos converts ectodermal cells to a neural fate. Genes & Development 8:1434-1447.
Uittenbogaard M, Chiaramello A (1999) Expression of the basic Helix-Loop-Helix ME1 E-protein during development and aging of the murine cerebellum. Neuroscience Letters 274:191-194.
Underhill C, Qutob MS, Yee SP, Torchia J (2000) A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. Journal of Biological Chemistry 275:40463-40470.
Unger GM, Davis AT, Slaton JW, Ahmed K (2004) Protein kinase CK2 as regulator of cell survival: implications for cancer therapy. Current Cancer Drug Targets 4:77-84.
Van Balkom ID, Vuijk PJ, Franssens M, Hoek HW, Hennekam RC (2012) Development, cognition, and behaviour in Pitt-Hopkins syndrome. Developmental Medicine and Child Neurology 54:925-931.
Vandewalle C, Comijn J, De Craene B, Vermassen P, Bruyneel E, Andersen H, Tulchinsky E, Van Roy F, Berx G (2005) SIP1/ZEB2 induces EMT by repressing genes of different epithelial cell-cell junctions. Nucleic Acids Research 33:6566-6578.
Veltman JA, Brunner HG (2012) De novo mutations in human genetic disease. Nature Reviews: Genetics 13:565-575.
Vicent S, Luis-Ravelo D, Anton I, Garcia-Tunon I, Borras-Cuesta F, Dotor J, De Las Rivas J, Lecanda F (2008) A novel lung cancer signature mediates metastatic bone colonization by a dual mechanism. Cancer Research 68:2275-2285.
Vilchez D, Boyer L, Morantte I, Lutz M, Merkwirth C, Joyce D, Spencer B, Page L, Masliah E, Berggren WT, Gage FH, Dillin A (2012) Increased proteasome activity in human embryonic stem cells is regulated by PSMD11. Nature 489:304-308.
Voronova A, Baltimore D (1990) Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proceedings of the National Academy of Sciences of the United States of America 87:4722-4726.
Walsh T et al. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320:539-543.
Wang C, Bian W, Xia C, Zhang T, Guillemot F, Jing N (2006a) Visualization of bHLH transcription factor interactions in living mammalian cell nuclei and developing chicken neural tube by FRET. Cell Research 16:585-598.
Wang D, Claus CL, Vaccarelli G, Braunstein M, Schmitt TM, Zuniga-Pflucker JC, Rothenberg EV, Anderson MK (2006b) The basic helix-loop-helix transcription factor HEBAlt is expressed in pro-T cells and enhances the generation of T cell precursors. Journal of Immunology 177:109-119.
Wen YD, Perissi V, Staszewski LM, Yang WM, Krones A, Glass CK, Rosenfeld MG, Seto E (2000) The histone deacetylase-3 complex contains nuclear receptor corepressors. Proceedings of the National Academy of Sciences of the United States of America 97:7202-7207.
Whalen S et al. (2012a) Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: Clinical score and further delineation of the TCF4 mutational spectrum. Human Mutation 33:64-72.

Whalen S et al. (2012b) Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: clinical score and further delineation of the TCF4 mutational spectrum. Human Mutation 33:64-72.
Wieben ED, Aleff RA, Tosakulwong N, Butz ML, Highsmith WE, Edwards AO, Baratz KH (2012) A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. PLoS One 7:e49083.
Wikstrom I, Forssell J, Goncalves M, Colucci F, Holmberg D (2006) E2-2 regulates the expansion of pro-B cells and follicular versus marginal zone decisions. Journal of Immunology 177:6723-6729.
Wilson SW, Rubenstein JL (2000) Induction and dorsoventral patterning of the telencephalon. Neuron 28:641-651.
Wirgenes KV, Sonderby IE, Haukvik UK, Mattingsdal M, Tesli M, Athanasiu L, Sundet K, Rossberg JI, Dale AM, Brown AA, Agartz I, Melle I, Djurovic S, Andreassen OA (2012) TCF4 sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders. Translational Psychiatry 2:e112.
Wong MC, Castanon I, Baylies MK (2008) Daughterless dictates Twist activity in a context-dependent manner during somatic myogenesis. Developmental Biology 317:417-429.
Wu PY, Ruhlmann C, Winston F, Schultz P (2004) Molecular architecture of the S. cerevisiae SAGA complex. Molecular Cell 15:199-208.
Wu Y, Sato F, Yamada T, Bhawal UK, Kawamoto T, Fujimoto K, Noshiro M, Seino H, Morohashi S, Hakamada K, Abiko Y, Kato Y, Kijima H (2012) The BHLH transcription factor DEC1 plays an important role in the epithelial-mesenchymal transition of pancreatic cancer. International Journal of Oncology 41:1337-1346.
Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M (2011) Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nature Genetics 43:864-868.
Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, Levy S, Gogos JA, Karayiorgou M (2012) De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature Genetics 44:1365-1369.
Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C, Savagner P, Gitelman I, Richardson A, Weinberg RA (2004) Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell 117:927-939.
Yoon HG, Chan DW, Huang ZQ, Li J, Fondell JD, Qin J, Wong J (2003) Purification and functional characterization of the human N-CoR complex: the roles of HDAC3, TBL1 and TBLR1. EMBO Journal 22:1336-1346.
Yoon SO, Chikaraishi DM (1994) Isolation of two E-box binding factors that interact with the rat tyrosine hydroxylase enhancer. Journal of Biological Chemistry 269:18453-18462.
Yoshida S, Ohbo K, Takakura A, Takebayashi H, Okada T, Abe K, Nabeshima Y (2001) Sgn1, a basic helix-loop-helix transcription factor delineates the salivary gland duct cell lineage in mice. Developmental Biology 240:517-530.
Yusufzai TM, Wolffe AP (2000) Functional consequences of Rett syndrome mutations on human MeCP2. Nucleic Acids Research 28:4172-4179.
Zebedee Z, Hara E (2001) Id proteins in cell cycle control and cellular senescence. Oncogene 20:8317-8325.

Zhai Y, Wu R, Schwartz DR, Darrah D, Reed H, Kolligs FT, Nieman MT, Fearon ER, Cho KR (2002) Role of beta-catenin/T-cell factor-regulated genes in ovarian endometrioid adenocarcinomas. American Journal of Pathology 160:1229-1238.

Zhang J, Kalkum M, Yamamura S, Chait BT, Roeder RG (2004) E protein silencing by the leukemogenic AML1-ETO fusion protein. Science 305:1286-1289.

Zhao F, Vilardi A, Neely RJ, Choi JK (2001) Promotion of cell cycle progression by basic helix-loop-helix E2A. Molecular and Cellular Biology 21:6346-6357.

Zhuang Y, Cheng P, Weintraub H (1996) B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. Molecular and Cellular Biology 16:2898-2905.

Zhuang Y, Barndt RJ, Pan L, Kelley R, Dai M (1998) Functional replacement of the mouse E2A gene with a human HEB cDNA. Molecular and Cellular Biology 18:3340-3349.

Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, Bijlsma EK, Oortveld MA, Ekici AB, Reis A, Schenck A, Rauch A (2009) CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. American Journal of Human Genetics 85:655-666.

Zweier C, Peippo MM, Hoyer J, Sousa S, Bottani A, Clayton-Smith J, Reardon W, Saraiva J, Cabral A, Gohring I, Devriendt K, de Ravel T, Bijlsma EK, Hennekam RC, Orrico A, Cohen M, Dreweke A, Reis A, Nurnberg P, Rauch A (2007) Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). American Journal of Human Genetics 80:994-1001.

Zweier C et al. (2008a) Further delineation of Pitt-Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. Journal of Medical Genetics 45:738-744.

Zweier C et al. (2008b) Further delineation of Pitt-Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. Journal of Medical Genetics 45:738-744.

Zweier M, Rauch A (2012) The MEF2C-Related and 5q14.3q15 Microdeletion Syndrome. Molecular syndromology 2:164-170.

Appendix I

Classification of the basic helix-loop-helix transcription factor family.
The bHLH family is sub-divided into seven functional classes according to tissue expression, dimerisation capability and DNA binding specificity (Classes I-VII) (Massari and Murre, 2000). The official HUGO gene names of transcription factors in each class is provided with the basic characteristics of each group. Letters in brackets under the class number represent the categorisation of the bHLH family members by two other classification systems based on phylogenetic analysis (Ledent et al., 2002; Stevens et al., 2008).

Class	Family	Human genes	Defining features
I (A, B)	E-proteins	*TCF3 (E2A, E47/E12)* *TCF12 (HEB)* *TCF4 (E2-2, ITF2, SEF2)*	*Homodimerisation; Heterodimerisation with class II, class V, and class VI transcription factors; each gene contains two main isoforms; widely expressed*
II (A, A)	NEUROD	*NEUROD1, NEUROD2, NEUROD4, NEUROD6*	*Heterodimerisation with E-proteins; tissue specific expression; cell differentiation*
II (A, A)	NEUROG	*NEUROG1, NEUROG2, NEUROG3*	
II (A, A)	ATOH	*ATOH1, ATOH7, ATOH8*	
II (A, A)	ASCL	*ASCL1, ASCL2, ASCL3, ASCL4, ASCL5*	
II (A, E)	OLIG	*OLIG1, OLIG2, OLIG3*	
I (A, C)	MYOD	*MYOD1, MYOG, MYF5, MYF6*	
II (A, A)	dHAND	*HAND1, HAND2*	*Homo and heterodimerisation*
II (A, A)	TWIST	*TWIST1, TWIST2*	*Homo and heterodimerisation; activates and represses transcription*
III (B^1)	TFEB		*bHLH/Z domain; Activates and represses transcription*
III (B^1)	TFE3		
III (B^1)	SREBF	*SREBF1, SREBF2*	
III (B, E)	MYC	*MYC, MYCN, MYCL1, MYCL2*	
IV (B, C)	MXD1 (Mad)	*MXD1, MXD3, MXD4, MXI1*	
IV (B, D)	MAX	*MAX*	
V (D, B)	ID proteins	*ID1, ID2, ID3, ID4*	*Inhibitors of differentiation; lack basic domain responsible for*

			DNA binding properties; heterodimerise with E-proteins to inhibit activity
VI (E, B)	HES	*HES1, HES2, HES3, HES4, HES5, HES6, HES7*	*bHLH-O domain; Proline rich region; homodimers bind N-boxes; Inhibit the activity of E-proteins*
VI (E, B)	HEY	*HEY1, HEY2, HEYL*	
VII (C, E)	AHR	*AHR, AHRR*	*bHLH-PAS domain*
VII (C, E)	SIM2	*SIM1, SIM2*	
VII (C, E)	TRH	*NPAS3*	
VII (C, E)	CLOCK	*CLOCK, NPAS2*	
VII (C, E)	BMAL	*ARNTL, ARNTL2*	
VII (C, E)	ARNT	*ARNT, ARNT2*	

Appendix II

List of primers and probes used in this thesis. All DNA sequences are written in the 5' to 3' direction. For the mutagenesis primers, only the forward sequence is provided.

Name	DNA sequence/probe name
Plasmid primers (Section 2.3.2)	
HA/Myc-TCF4 F	CTGCGAATTCGGATGCATCACCAACAGCGAATG
HA/Myc-TCF4 R	CTGGCGGCCGCATTACATCTGTCCCATGTGATTC
TRX/GST-TCF4 F	CTGGAATTCCTGTTTGGTCTAGAAATGGAG
TRX/GST-TCF4 R	CTGGTCGACATCGTCATTATTGCTAGTAATTG
GFP-TCF4 F	CTGTCGAATTCTATGCATCACCAACAGCGAATG
GFP-TCF4 R	CTGGTCGACTTACATCTGTCCCATGTGATTC
HA-ATOH1 F	GTCCGAATTCGGATGTCCCGCCTGCTGCATG
HA-ATOH1 R	CTGCGGTCGACCCTAACTTGCCTCATCCGAGTC
myc-E47 F	CTGCGAATTCGGATGAACCAGCCGCAGAGG
myc-E47 R	CTGCGGTCGACCTCACATGTGCCCGGCGG
pLuc-CNTNAP2 F	CTGAAGCTTCTGCTTGAGATACCCATGGAA
pLuc-CNTNAP2 R	CTGAAGCTTAAGAGCAGAGGCTGCAGAAG
pLuc-NRXN1β F	CTGAAGCTTAGAGTGCGCCCTTCTGAAG
pLuc-NRXN1β R	CTGAAGCTTAATGGGAAGAAGAATCCTTGT
TCF4 600F	CTATCCTTCCTCCAAACCAG
TCF4 1200F	CCGGAACCATGCAGTGGG
TCF4 568R	GGCAAACCTGGAGGAACTTT
CNTNAP2 630F	TAC ATA TCG AAA ACA TGT GCA TT
E47 285F	TTCATCCACATTCCTGGGAC
E47 1385R	TGGTTGTGCATGAGGCTGGT
G358V	CATCTCTCTCAGCAGTCACAGCTGTTTGGTC
D535G	GAGGACAAGAAATTAGGTGACGACAAGAAGAT
R578P	AACAATGCCCGAGAGCCTCTGCGGGTCCGTGAC
R580W	GCCCGAGAGCGTCTGTGGGTCCGTGACATCAAC
A614V	CTGATCCTCCACCAGGTGGTGGCCGTCATCCTC
BHLHE40 F	TACCTGATCCCACCTTCAGC
BHLHE40 R	TTGAGGCCTGGGTATAGCAC
MEF2C F	TCGAGATACCCACAACACAC
MEF2C R	TCGTACGAACTGCTACAGCT
TCF3/E47 F	AAGCCACTGCACACAGACAG
TCF3/E47 R	CGCATCACTTTCCACATGAC

SNAI1 F	ACAGCTGCTTTGAGCTACAG
SNAI1 R	GCATAGTTAGTCACACCTCG
SNAI2 F	CAGTGCAAAAACTGCTCCAA
SNAI2 R	GCTTCGGAGTGAAGAAATGC
ZEB1 F	GTGCACAAGAAGAGCCACAA
ZEB1 R	TTGCGCAAGACAAGTTCAAG
ZEB2 F	TACGGATCCCGAAACGATAC
ZEB2 R	CCTCGTGGTCTGATTTGGTT
TCF4 EX13 F	ATGGCAAATAGAGGAAGCGG
TCF4 EX13 R	TGGAGAATAGATCGAAGCAAG
TCF4A F	GAATCCGAGACCATGTACTG
TCF4A R	AGCATAGACCGCTTTCCCAT
TCF4 EX4 F	CTGGATTTCAGTGCGATGTT
TCF4 EX4 R	GTCTTCTACATTTGAGCCAG
GAPDH F	TGCACCACCAACTGCTTAGC
GAPDH R	GGCATGGACTGTGGTCATGAG
18S F	CCATCCAATCGGTAGTAGCG
18S R	GTAACCCGTTGAACCCCATT
NDUFA4 F	AGCTGAAGAAGGAACGTCCA
NDUFA4 R	GGCTTCTGGAAGACCTTCAT
COX6A1 F	AGAGAATCTGGACCACTACC
COX6A1 R	GTAATGGTCCAAACCAGTGC
TaqMan probes (Section 2.9.5)	
IGF2	Hs01005964_g1
CDKN1C	Hs00908986_g1
NEUROG2	Hs00935087_g1
CDK6	Hs01026373_m1
BMP7	Hs01002399_m1
CASP1	Hs00354836_m1
CASP8	Hs01018151_m1
FAS	Hs00236330_m1
NTRK1	Hs01021006_g1
NOTCH1	Hs01062014_m1
18S	Hs03003631_g1
DPT	Hs00355056_m1
NEUROG2	Hs00935087_g1
SNAI2	Hs00950344_m1
SEMA3A	Hs00173810_m1
RELN	Hs01022646_m1

TLE1	Hs00270768_m1
RCOR2	Hs00293511_m1
SNAI1	Hs00195591_m1
SKP2	Hs01021864_m1

Appendix III

List of genes differentially expressed in GAPDH knockdown cells. Differentially expressed genes were established from microarray data of mock and GAPDH KD treated cells. A one-way ANOVA was performed between the two groups that generated a list of 26 genes (FDR 0.01).

Entrez ID	Gene Symbol	*P*-value	Fold-Change
3434	IFIT1	4.95E-06	186.82
2195	FAT1	1.31E-07	2.52
80314	EPC1	9.33E-06	1.79
25953	PNKD	2.10E-06	1.64
5445	PON2	7.69E-06	1.33
100293266	-	3.86E-06	-1.22
6256	RXRA	1.09E-05	-1.43
199720	GGN	7.55E-06	-1.57
497661	C18orf32	6.12E-06	-1.63
122769	PPIL5	4.00E-06	-1.67
146562	C16orf71	1.54E-06	-1.79
100293485	-	8.85E-06	-1.93
339345	NANOS2	7.08E-06	-2.13
-	RCN2	5.86E-06	-2.18
8564	KMO	5.99E-08	-2.34
144110	TMEM86A	4.26E-06	-2.85
115653	KIR3DL3	6.65E-09	-3.55
51540	SCLY	7.52E-06	-3.57
91056	AP001266.4-2	2.43E-06	-5.50
649137	ZSCAN5C	3.77E-10	-5.58
4803	NGF	5.95E-08	-5.74
-	AL732579.5	5.04E-09	-5.74
2597	GAPDH	1.88E-06	-6.91
100134235	-	3.47E-07	-8.11
-	RP11-122A8.1	1.11E-10	-8.32
-	RP4-686J16.2	1.78E-06	-8.34

Appendix IV

Amino acid sequence of the TCF4-B$^+$ and TCF4-A$^+$ isoforms. Selected functional domains are highlighted: basic helix-loop-helix domain, blue; activation domains, yellow; nuclear localisation signal, red; RSRS, bold type; isoform-specific sequences, grey.

TCF4-B$^+$ (E2-2B, ITF2-B, SEF2-1B)

671 amino acids, +RSRS
Ensembl: TCF4-201, ENSP00000455984
NCBI: Transcription factor 4 isoform a, NP_001077431.1

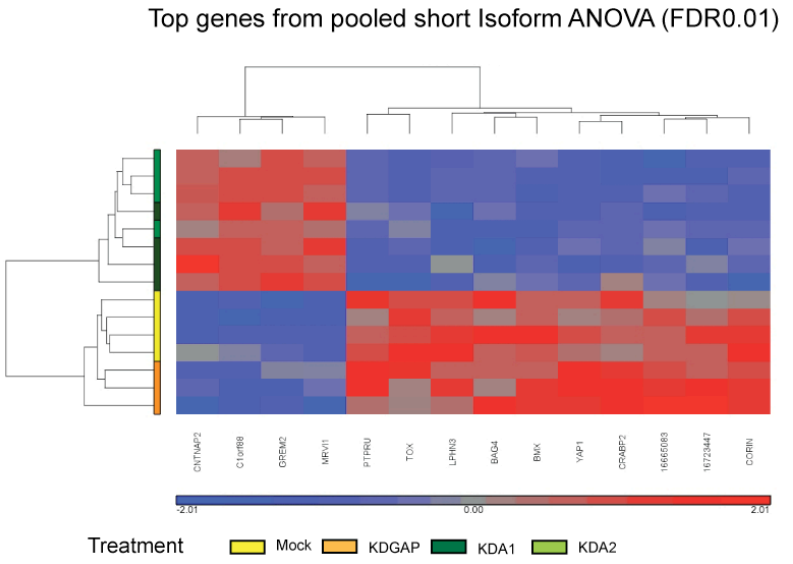
MHHQQRMAALGTDKELSDLLDFSAMFSPPVSSGKNGPTSLASGHFT
GSNVEDRSSSGSWGNGGHPSPSRNYGDGTPYDHMTSRDLGSHDNLS
PPFVNSRIQSKTERGSYSSSYGRESNLQGCHQQSLLGGDMDMGNPGT
LSPTKPGSQYYQYSSNNPRRRPLHSSAMEVQTKKVRKVPPGLPSSVY
APSASTADYNRDSPGYPSSKPATSTFPSSFFMQDGHHSSDPWSSSSG
MNQPGYAGMLGNSSHIPQSSSYCSLHPHERLSYPSHSSADINSSLPPM
STFHRSGTNHYSTSSCTPPANGTDSIMANRGSGAAGSSQTGDALGK
ALASIYSPDHTNNSFSSNPSTPVGSPPSLSAGTAVWSRNGGQASSSPN
YEGPLHSLQSRIEDRLERLDDAIHVLRNHAVGPSTAMPGGHGDMHG
IIGPSHNGAMGGLGSGYGTGLLSANRHSLMVGTHREDGVALRGSHS
LLPNQVPVPQLPVQSATSPDLNPPQDPYRGMPPGLQGQSVSSGSSEI
KSDDEGDENLQDTKSSEDKKLDDDKKDIKSITRSRSSNNDDEDLTPE
QKAEREKERRMANNARERLRVRDINEAFKELGRMVQLHLKSDKPQ
TKLLILHQAVAVILSLEQQVRERNLNPKAACLKRREEEKVSSEPPPLS
LAGPHPGMGDASNHMGQM

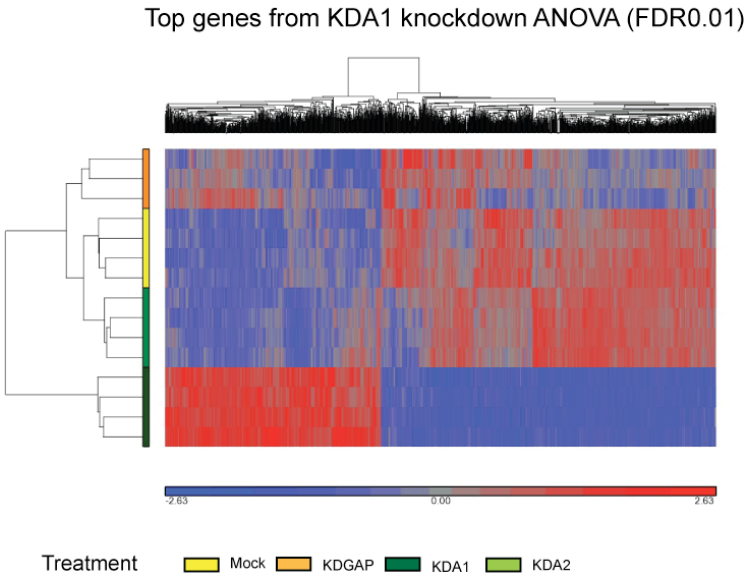
TCF4-A$^+$ (E2-2A, ITF2-A, SEF2-1A)

511 amino acids, +RSRS
Ensembl: TCF4-015, ENSP00000409447
NCBI: Transcription factor 4 isoform j, NP_001230163.1

MYCAYTIPGMGGNSLMYYYNGKAVYAPSASTADYNRDSPGYPSSK
PATSTFPSSFFMQDGHHSSDPWSSSSGMNQPGYAGMLGNSSHIPQSS
SYCSLHPHERLSYPSHSSADINSSLPPMSTFHRSGTNHYSTSSCTPPAN
GTDSIMANRGSGAAGSSQTGDALGKALASIYSPDHTNNSFSSNPSTP
VGSPPSLSAGTAVWSRNGGQASSSPNYEGPLHSLQSRIEDRLERLDD
AIHVLRNHAVGPSTAMPGGHGDMHGIIGPSHNGAMGGLGSGYGTG
LLSANRHSLMVGTHREDGVALRGSHSLLPNQVPVPQLPVQSATSPD
LNPPQDPYRGMPPGLQGQSVSSGSSEIKSDDEGDENLQDTKSSEDKK
LDDDKKDIKSITRSRSSNNDDEDLTPEQKAEREKERRMANNARERL
RVRDINEAFKELGRMVQLHLKSDKPQTKLLILHQAVAVILSLEQQVR
ERNLNPKAACLKRREEEKVSSEPPPLSLAGPHPGMGDASNHMGQM

Appendix V





Hierarchical clustering of the top differentially expressed genes with both short isoform knockdown groups. The top differentially expressed genes (FDR 0.01) from pooled short isoform ANOVA were subjected to hierarchical clustering (top). The ANOVA compared the short isoform knockdown groups (KDA1 and KDA2) to controls (KDGAP and mock). This analysis generated 14 differentially expressed genes, indicating a lack of similarity in the gene expression changes between KDA1 and KDA2. The bottom panel illustrates hierarchical clustering of the top differentially expressed genes (FDR 0.01) generated from the ANOVA comparing KDA1 samples to controls. This analysis is the same as Figure 6.4 except the KDA2 samples have been included in the clusting analysis. In this figure KDA2 clusters with control groups (Mock and KDGAP) indicating that top gene expression changes for KDA1 are not comparable to KDA2.

Appendix VI

Publications arising from this study

Blake DJ, Forrest M, Chapman RM, Tinsley CL, O'Donovan MC, Owen MJ (2010) **TCF4, schizophrenia, and Pitt-Hopkins** Syndrome. *Schizophr Bull* 36:443-447.

Forrest M, Chapman RM, Doyle AM, Tinsley CL, Waite A, Blake DJ (2012) **Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome**. *Hum Mutat* 33:1676-1686.

Forrest MP, Waite AJ, Martin-Rendon E, Blake DJ (2013) **Knockdown of Human TCF4 Affects Multiple Signaling Pathways Involved in Cell Survival, Epithelial to Mesenchymal Transition and Neuronal Differentiation**. *PLoS One* 8:e73169.