TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
26/2023

Molecular Characterization of Basic Helix-Loop-Helix Transcription Factor TCF4: From Expression to Function

ALEX SIRP

TAL TECH PRESS

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
26/2023

Aluselise heeliks-ling-heeliks transkriptsiooniteguri TCF4 ekspressiooni ja funktsiooni kirjeldamine

ALEX SIRP

TAL TECH
KIRJASTUS

Contents

List of Publications 6
Author's Contribution to the Publications 7
Introduction 8
Abbreviations 9
1 Review of literature 10
1.1 Basic helix-loop-helix transcription factors 10
1.1.1 E-proteins 10
1.2 Transcription factor 4 11
1.2.1 Functions 11
1.2.2 Animal studies 12
1.2.3 Interactome 13
1.2.4 Target genes 14
1.2.5 Regulation of activity 16
1.2.6 Gene structure and functional protein domains 17
1.2.7 Expression 19
1.3 TCF4-related diseases 20
1.3.1 Pitt-Hopkins Syndrome 20
1.3.2 Schizophrenia 21
1.3.3 Mild to moderate intellectual disability 22
1.3.4 Fuchs endothelial corneal dystrophy 22
1.3.5 Tumorigenesis 22
2 Aims of the study 24
3 Materials and Methods 25
4 Results 26
4.1 Results obtained in publication I 26
4.2 Results obtained in publication II 26
4.3 Results obtained in publication III 27
5 Discussion 28
5.1 Expression of TCF4 in rodent and human 28
5.2 The effect of FECD associated CTG trinucleotide repeat expansion on the expression of *TCF4* 29
5.3 The effect of disease-related missense variations and mutations in *TCF4* on the functionality of the protein 31
6 Conclusions 33
7 References 34
Acknowledgements 45
Abstract 46
Lühikokkuvõte 48
Appendix 51
Publication I 51
Publication II 71
Publication III 85
Curriculum vitae 103
Elulookirjeldus 106

List of Publications

The list of author's publications, on the basis of which the thesis has been prepared:

I. **Sirp, A.***, Shubina, A.*, Tuvikene, J., Tamberg, L., Kiir, C.S., Kranich, L., Timmusk, T.
Expression of alternative transcription factor 4 mRNAs and protein isoforms in the developing and adult rodent and human tissues.
Front. Mol. Neurosci., 15. 2022 Nov. DOI: 10.3389/fnmol.2022.1033224.

II. **Sirp, A**.*, Leite, K.*, Tuvikene, J.*, Nurm, K., Sepp, M., Timmusk, T.
The Fuchs corneal dystrophy-associated CTG repeat expansion in the TCF4 gene affects transcription from its alternative promoters.
Sci. Rep. 2020 Oct; 10 (1), #18424. DOI: 10.1038/s41598-020-75437-3

III. **Sirp, A**.*, Roots, K.*, Nurm, K., Tuvikene, J., Sepp, M., Timmusk, T.
Functional consequences of TCF4 missense substitutions associated with Pitt-Hopkins syndrome, mild intellectual disability, and schizophrenia.
J. Biol. Chem. 2021 Dec; 297, 101381. doi: 10.1016/j.jbc.2021.101381.

*- Equal contribution

Author's Contribution to the Publications

Contribution to the papers in this thesis are as follows:

I-III In all the publications, I participated in designing the experiments, performed the experiments and wrote most of the manuscript.

Introduction

Transcription factor 4 (TCF4) is a broadly expressed basic helix-loop-helix transcription factor that is essential in neurogenesis and functioning of the nervous system. Deficits in TCF4 function have been implicated in a number of severe neurocognitive disorders such as schizophrenia and intellectual disability. In addition, just a single mutation in the basic helix-loop-helix region can cause Pitt-Hopkins syndrome, a rare genetic autism spectrum disorder described by severe neurodevelopmental delay. Expansion of a repeat region in an intron of *TCF4* has also been tied to development of Fuchs' endothelial corneal dystrophy, a highly prevalent eye disease affecting vision.

TCF4 functions through the formation of homo- or heterodimers. Due to the many interaction partners with contrasting expression patterns, TCF4 can exert various functions, depending on tissue type and developmental stage. While homozygous deletion of *Tcf4* in rodents is lethal, *Tcf4* haploinsufficiency causes a Pitt-Hopkins syndrome-like phenotype.

In this thesis, we characterized *TCF4* mRNA and protein expression throughout rodent and human development with focus on the many distinct *TCF4* isoforms. In addition, we studied the effects of previously described disease related aberrations in *TCF4* on the expression and functionality of TCF4 protein. Results of this thesis help better understand the overall function of TCF4 and may help lay the foundation for gene therapy approaches for the many TCF4 associated diseases.

Abbreviations

ASCL1	Achaete-scute homolog 1
bHLH	Basic helix-loop-helix
ChIP	Chromatin immunoprecipitation
E-box	Ephrussi box
FECD	Fuchs endothelial corneal dystrophy
FXTAS	Fragile X-associated ataxia syndrome
ID	Inhibitor of DNA binding
iPSC	Induced pluripotent stem cell
LTP	Long term potentiation
NEUROD	Neurogenic differentiation factor
NPC	Neural progenitor cells
PDC	Plasmacytoid dendritic cell
RNA-seq	RNA sequencing
SAHA	Suberoylanilide hydroxamic acid
TCF4	Transcription factor 4
TCF7L2	Transcription factor 7-like 2

1 Review of literature

1.1 Basic helix-loop-helix transcription factors

Basic helix-loop-helix (bHLH) transcription factors are named after their highly conserved HLH protein domain (two alpha-helices connected by a loop) which is necessary for dimerization with other transcription factors. The basic region of bHLH transcription factors mediates DNA binding. Even though bHLH proteins are not present in prokaryotes, they are expressed in eukaryotic organisms including fungi, animals and plants (Murre, 2019).

The bHLH transcription factor family is divided into seven classes (I-VII). These classes are grouped based on their interactions with other bHLH transcription factors. Members of the class I bHLH transcription factors in mammals include TCF3 (also known as E2A including splice variants E12 and E47), TCF4 (E2-2) and TCF12 (HEB). The only class I protein in *Drosophila melanogaster* is daughterless and in *Caenorhabditis elegans* is helix-loop-helix protein 2 (hlh-2). To regulate transcription of target genes, class I proteins need to form either homodimers or heterodimers with class II bHLH proteins (ASCL1, NEUROD1 and 2, MyoD etc.) before binding to their target sequence CANNTG (N = any nucleotide), the sequence also known as Ephrussi box (E-box) (Massari and Murre, 2000). In addition, class I and II proteins can form heterodimers with class V proteins that function as negative regulators of transcriptional regulation as they lack the DNA binding domain (Benezra et al., 1990). In vertebrate, class V is formed by the inhibitors of DNA binding (ID) family of proteins. The interaction between class I, II and V transcription factors is very important in neural development (Massari and Murre, 2000).

The remaining classes of bHLH proteins contain additional functional protein domains which define their classes. Class III (USF1, MITF, etc.) and class IV proteins (MAX, MNT, etc.) contain a leucine zipper domain after the bHLH region, which mediates dimerization within and between these classes of bHLH proteins (Murre, 2019). Class VI proteins (HES1-7 etc.) have a proline residue in their basic region and are known for their interaction with the co-repressor Groucho. Class VII proteins (BMAL, CLOCK etc.) contain several per-ARNT-sim (PAS) domains that react to light and oxygen (Massari and Murre, 2000; Murre, 2019).

1.1.1 E-proteins

The class I bHLH transcription factors TCF3, TCF4 and TCF12 are also known as E-proteins. Homozygous null mutation for any of the E-proteins results in early postnatal lethality (Zhuang et al., 1994, 1996). The roles of E-proteins have been studied in detail with the overall function being participation in neurogenesis. However, in early studies, most of the focus was on describing the role of E-proteins in immune cell maturation. While all the E-proteins are important for the development of pro-B cells (Zhuang et al., 1996), it is important to note that replacing the mouse *Tcf3* gene with the human *TCF12* gene can compensate for the loss of *Tcf3* (Zhuang et al., 1998). Similar results have been obtained in *Caenorhabditis elegans* where substitution of *hlh-2* with human *TCF3* rescues the negative effects associated with *hlh-2* knockdown (Sallee and Greenwald, 2015). In *Drosophila melanogaster*, overexpression of human *TCF4* rescues the embryonic lethality of *daughterless* null mutation (Tamberg et al., 2015). Together, these results suggest that the functioning of E-proteins is conserved through evolution.

A compensatory effect between endogenously expressed E-proteins has been implicated in the developing rodent nervous system (Ravanpay and Olson, 2008). However, during hindbrain development loss of *Tcf4* cannot be rescued by other endogenously expressed E-proteins even though their expression patterns are comparable (Flora et al., 2007). In addition, more recent animal studies suggest that compensatory mechanisms do not exist within an E-protein as well. For example, the expression of shorter *Tcf4* isoforms cannot alleviate the negative effects arising from the loss-of-function of longer isoforms (Jung et al., 2018; Wittmann et al., 2021). This is also supported by the fact that splice variants of *TCF3* share the majority of binding sites but have differing roles in mouse embryonic neural stem cells (NSC) – E47 acts mainly as a transcriptional repressor and E12 functions as a transcriptional activator (Pfurr et al., 2017).

1.2 Transcription factor 4

TCF4 (also known as E2-2, ITF-2, SEF-2) was first described as an activator of immunoglobulin enhancers (Henthorn et al., 1990). It must be emphasized that transcription factor 4 (*TCF4*, gene ID:6925) gene should not be confused with transcription factor 7-like 2 (*TCF7L2*, gene ID:6934) gene as *TCF7L2* is historically referred to as T-cell factor 4 and abbreviated also as *TCF4*. Due to the same abbreviation, there has been much confusion and misinterpretation of data between transcription factor 4 *TCF4* and transcription factor 7-like 2 *TCF4*. When doing research on *TCF4* it is suggested to check the methods section for primers, antibodies etc. to confirm which *TCF4* is studied.

1.2.1 Functions

Early studies showed that TCF4 binds to regulate the viral glucocorticoid response element (Corneliussen et al., 1991), the rat tyrosine hydroxylase enhancer (Yoon and Chikaraishi, 1994) and the human somatostatin receptor II promotor (Pscherer et al., 1996). First animal studies on TCF4 concluded that *Tcf4* is important in the development of immune cells – B- (Zhuang et al., 1996) and T-cells (Bergqvist et al., 2000). In addition, *Tcf4* is necessary for the development of plasmacytoid dendritic cells (PDC), as deficiency of *Tcf4* reduces the number of PDCs. TCF4 regulates the expression of genes common to PDC-s and can thus regulate conversion between classical dendritic cells and PDC-s (Cisse et al., 2008; Ghosh et al., 2010). More specifically, the linage commitment of dendritic cells is coordinated by the expression levels of Tcf4 and Id2 and their upstream expression regulators Stat3 and Stat5, respectively (Li et al., 2012). Even though ID2 is a dimerization partner of TCF4, there is evidence that ID2 and ID3 interact exclusively with only TCF12 and only ID1 interacts with all of the E-proteins (Oh et al., 2021; Kantzer et al., 2022).

By now it is well known that TCF4 has a very important role in neurogenesis. TCF4 promotes differentiation and regulates proliferation of NSCs (Fischer et al., 2014; Shariq et al., 2021) and is necessary for neuronal migration, axon guidance and synapse formation (Li et al., 2019; Mesman et al., 2020; Wittmann et al., 2021). Loss of *Tcf4* causes changes in the architecture of cortical layers and cerebellum, and affects the development of corpus callosum, midline glia and hippocampus (Hellwig et al., 2019; Mesman et al., 2020). TCF4 is also involved in adult hippocampal neurogenesis, where in addition to the regulation of differentiation and proliferation it supresses the inflammatory transformation of neural progenitor cells (NPC) (Shariq et al., 2021). Interaction between MATH1 and TCF4 has been suggested to be important for development of the hindbrain (Flora et al., 2007).

TCF4 is also associated with epithelial-mesenchymal transition (EMT) and *Tcf4* overexpression causes migratory and invasive behaviour of cells *in vitro* (Sobrado et al., 2009). This is supported by newer experiments as TCF4 knockdown in SH-SY5Y cells causes differential expression of important regulators of EMT such as *DEC1* and *SNAI2*, and of genes associated with cell survival and neuronal differentiation such as *ASCL1* and *NEUROG2* (Forrest et al., 2013).

1.2.2 Animal studies

Homozygous *Tcf4* null mice were created by inserting a neo cassette into the bHLH region of *Tcf4* gene (Zhuang et al., 1996). These mice usually die around birth (Zhuang et al., 1998; Flora et al., 2007). 30% of *Tcf4* null animals which are born die by P4 (Cleary et al., 2021). Very high lethality has also been described in mice with homozygous deletion for *Tcf4* exon 4 – present only in a subset of longer TCF4 protein isoforms (Jung et al., 2018; Wittmann et al., 2021). In addition, homozygous in-frame deletion which affects six sequential amino acids (574-579) in the bHLH region of *Tcf4* is embryonically lethal (Thaxton et al., 2018). As *Tcf4* null animals die before birth there is not much information about the phenotype caused by total *Tcf4* knock-out. However, it is known that at P0, *Tcf4* null animals show clustering of neuronal precursors in the hindbrain which are supposed to migrate to the pontine nucleus (Flora et al., 2007). Total knock-out and knock-out of *Tcf4* exons present in longer isoforms results in an undeveloped forebrain commissure system (Mesman et al., 2020; Wittmann et al., 2021). RNA sequencing (RNA-seq) analyses from *Tcf4* knock-out mice indicate that TCF4 may regulate the expression of TCF4 dimerization partners such as *Ascl1*, *NeuroD1, NeuroD2* and *Id2* (Mesman et al., 2020; Wittmann et al., 2021).

Tcf4 heterozygous animals are viable and have about 30% mortality at weaning age compared to wild type littermates (Zhuang et al., 1996; Flora et al., 2007; Cleary et al., 2021). Changes in the rodent brain caused by reduced *Tcf4* expression include abnormal cortical development (Li et al., 2019; Mesman et al., 2020), neuronal migration (Flora et al., 2007; Chen et al., 2016; Wang et al., 2020), oligodendrocyte differentiation (Phan et al., 2020; Wedel et al., 2020), dendrites including changes in branching and length (Crux et al., 2018; Sarkar et al., 2021) and aberrant neuronal firing (Rannals et al., 2016; Sarkar et al., 2021). Abnormal dendrites are also present in primary hippocampal cultures where *Tcf4* expression is silenced (Rosato et al., 2021). Interestingly, while the reduction of TCF4 in embryonic development results in reduced neuronal firing of cortical neurons (Rannals et al., 2016), reduction of TCF4 expression in adults results in hippocampal neuron hyperexcitability (Sarkar et al., 2021). Mice haploinsufficient for only longer TCF4 isoforms have reduced cortical volume and agenesis of the splenium of corpus callosum (Jung et al., 2018).

Smaller body weight of *Tcf4* heterozygous mice has also been noted but the results are contradicting. According to Grubišić *et al.*, *Tcf4* heterozygous mice have no changes in body weight (Grubišić et al., 2015), whereas Thaxton *et al.* reports decreased body weight of *Tcf4* heterozygous mice (Thaxton et al., 2018). Cleary *et al*. showed that *Tcf4* haploinsufficient mice have reduced body weight in earlier stages of postnatal development, and the difference becomes insignificant at later developmental stages (Cleary et al., 2021).

Mice where *Tcf4* has only been knocked-out in a set of glial fibrillary acidic protein (GFAP) positive NPCs are viable after birth. They exhibit smaller body weight, aberrant

migration of cerebellar granule cells and overall reduced cerebellar volume (Hellwig et al., 2019).

The effects of alterations in *Tcf4* expression on rodent behaviour have been well described. *Tcf4* heterozygous mice have deficits in prepulse inhibition, learning and memory, and prefer social isolation. In addition, they display hyperactivity, anxiety and enhanced long term potentiation (LTP) of hippocampal neurons (Kennedy et al., 2016). Such results of the behaviour of *Tcf4* heterozygous mouse have been confirmed by other studies with some exceptions (Rannals et al., 2016; Thaxton et al., 2018; Sarkar et al., 2021). According to Thaxton *et al*., *Tcf4* heterozygous mice are not anxious and asocial (Thaxton et al., 2018). Differences between Kennedy *et al*. and Thaxton *et al*. may arise from the use of different mouse strains. In addition to previously mentioned phenotypes, *Tcf4* heterozygous mice exhibit reduced frequencies of action potentials (Rannals et al., 2016; Thaxton et al., 2018).

Mice with postnatal overexpression of *Tcf4* have reduced fear memory and display deficits in prepulse inhibition. However, no defects in activity, exploration, pain sensitivity or histological aberrations in the brain are present (Brzózka et al., 2010). *In utero* overexpression of Tcf4-B in rat cortex at E16 alters the distribution of pyramidal cells in the developing neocortex (Page et al., 2018). Another interesting phenomen resulting from *Tcf4* overexpression in rats is the reduction of inflammatory and neuropathic pain sensitivity. This is achieved by supressing neuronal activity of dorsal root ganglion neurons via downregulation of $Na_v1.8$ expression(Li et al., 2020). The use of different Tcf4 mice models for research is extensively reviewed in Sweatt *et al*., (Sweatt, 2013).

1.2.3 Interactome

The exact functions and potential target genes of TCF4 are dependent on developmental context and the expression pattern of the many interaction partners of TCF4 (Powell and Jarman, 2008; Quednow et al., 2014). Interaction partners of TCF4 include ASCL1 (achaete-scute complex homolog 1), ATOH1 (atonal homolog 1), NEUROD1-3 (neurogenic differentiation 1-3), MYOD1 (myogenic differentiation 1), TAL1-2 (T-cell acute lymphocytic leukaemia 1-2), MSC (musculin), LYL1 (lymphoblastic leukemia derived sequence 1) and ID1-4 (inhibitor of DNA binding). The interactome of TCF4 also consists of AR (androgen receptor), CDC73 (cell division cycle 73), JUN (jun oncogene), PARP (poly ADP-ribose polymerase 1) and RUNX1T1 (runt-related transcription factor 1) (Blake et al., 2010).

The interactome of TCF4 has been widened by a newer study which indicates that TCF4 interacts with transcription regulators such as Sox2, Twist1, Smad4, p300, Smarca4, Chd7, Zeb2, Hcfc1, Ehmt1 and Ski (Moen et al., 2017). TCF4 also interacts with the mediator multiprotein complex which activates enhancers and super enhancers to regulate the expression of neurogenic transcription factors (Quevedo et al., 2019). Mediator complex also binds possible TCF4 interaction partners Sox2, p300 and Chd7 (Moen et al., 2017; Quevedo et al., 2019). Interaction of TCF4 with numerous non bHLH transcription factors is also supported by an expression pattern and regulon activity analysis from single cell RNA sequencing (scRNA-seq) data (Wittmann et al., 2021). However, RNA-seq based analyses do not necessarily confirm physical interaction.

There is also evidence that TCF4 protein isoforms may interact with different partners based on the presence of functional protein domains. For example, *in vivo* co-immunoprecipitation experiments reveal that while the longer TCF4-B protein isoform interacts with Sox11, the shorter isoform TCF4-A does not (Wittmann et al., 2021).

1.2.4 Target genes

The first systematic analyses on TCF4 target genes were done in PDC lines using chromatin immunoprecipitation (ChIP) combined with an TCF4 antibody specific for longer TCF4 isoforms only. In total, the results included >100 high-confidence genes related to pathogen sensing, signal transduction and transcriptional regulation (Cisse et al., 2008; Ghosh et al., 2010). More recent ChIP-sequencing (ChIP-seq) experiments with an antibody specific for all TCF4 isoforms from human neuroblastoma SH-SY5Y cells revealed TCF4 binding sites close to >5000 genes involved in neurogenesis, cell signalling, cell cycle regulation and ion transport. Based on histone modifications, 77% of TCF4 binding sites were in active enhancers and only 1.7% of TCF4 binding sites were in gene promoters (Forrest et al., 2018). In addition, ChIP-seq has been done in SH-SY5Y cells using an antibody specific only for long TCF4 protein isoforms which resulted in >6500 target genes (Xia et al., 2018). ChIP-seq in mouse NSCs using tagged TCF4 shows that TCF4 binds enhancer regions in *Nrxn1* gene together with p300 and also regulates primary microcephaly genes *Mcph1* and *Wdr62* by binding enhancer regions of these genes together with microcephaly-associated transcription factors Smad4, Sox2 and Chd7 (Moen et al., 2017). ChIP based analyses of TCF4 binding are summarized in table 1.

Table 1. *Summary of ChIP based genome wide analyses on potential TCF4 target genes. ChIP, chromatin immunoprecipitation; iPSC, induced pluripotent stem cell; NSC, neural stem cell.*

Reference	Method	Cell type	Isoform
Cisse et al., 2008;	ChIP-qPCR	Human CAL-1 cell line	TCF4-B
Ghosh et al., 2010	ChIP-on-ChIP microarray	Human CAL-1 cell line	TCF4-B
Moen et al., 2017	ChIP-seq	Mouse NSCs	Tagged-TCF4-B
Hennig et al., 2017	ChIP-seq	Human iPSC-derived neurons	Tagged-TCF4-A
Forrest et al., 2018	ChIP-seq	Human SH-SY5Y cell line	All isoforms
Xia et al., 2018	ChIP-seq	Human SH-SY5Y cell line	Long isoforms

Analyses on differentially expressed genes after TCF4 silencing or complete knock-out have also revealed potential target genes and functions of TCF4. Microarray analysis in SH-SY5Y cells showed that knock-down of *TCF4* leads to >4800 differentially expressed genes associated with signal transduction and neurogenesis, out of which only around 17% contain TCF4 binding sites (Forrest et al., 2013, 2018). In another study TCF4 knock-down in human NPC line resulted in 628 differentially expressed genes mainly involved in cell cycle regulation (Hill et al., 2017). *TCF4* knockdown in human induced pluripotent stem cell (iPSC) derived NPCs changed the expression of 161 genes (60 upregulated and 101 downregulated). The majority of differentially expressed genes were involved in neuronal development and differentiation (Hennig et al., 2017). RNA interference-mediated knockdown of *Tcf4* in mouse NSCs led to dysregulation of genes

associated with intellectual disability, schizophrenia, autism spectrum disorders and mental disorders (Moen et al., 2017).

Possible target genes of TCF4 have also arisen from RNA-seq data from mice where expression of functional TCF4 has been manipulated (Kennedy et al., 2016; Li et al., 2019; Mesman et al., 2020; Phan et al., 2020; Schoof et al., 2020; Sarkar et al., 2021; Wittmann et al., 2021). In *Tcf4* heterozygous knockout mice, RNA-seq of hippocampal neurons has revealed 402 differentially expressed genes. These genes are associated with neuronal plasticity, axon guidance, cell adhesion, calcium signalling and neuroreceptors. More specifically, upregulated genes include genes necessary for dopamine (*Drd1a, Cckbr, Chrm4*), oxytocin (*Oxtr*), serotonin (*Htr2c*), glycine (*Glra2, Glra3*) and neuromedin B (*Nmbr*) signalling. Downregulated genes included *Grin2a, Npy2r, Lpar1* and *S1pr5* involved in learning and memory, and genes associated with myelination. In addition, individual genes which were differentially expressed included upregulation of *Klotho* (enhancer of LTP) and downregulation of *Arc* (involved in synaptic plasticity and memory formation). Interestingly, *Lefty1* which is predominantly expressed in the left hemisphere was also downregulated. As the authors described weak front right paws for *Tcf4* heterozygous animals, it may be that *Lefty1* plays a role in that phenotype (Kennedy et al., 2016).

A combined meta-analysis of previous RNA-seq studies has been done by Sarkar *et al*. (Sarkar et al., 2021) of *Tcf4* haploinsufficient mouse data from adult hippocampus (Kennedy et al., 2016) and cortex (Phan et al., 2020). The results suggest an overall bi-directional role of TCF4 on transcription regulation meaning that depending on context, TCF4 can either activate or repress transcription. In addition, RNA-seq data of brain tissue from *Tcf4* heterozygous (Phan et al., 2020) and homozygous knock-out mice (Li et al., 2019) show minimal overlap of differentially expressed genes (Sarkar et al., 2021). The RNA expression-based studies on potential TCF4 target genes are summarized in table 2.

Table 2. *Summary of RNA based genome wide analyses on potential TCF4 target genes. iPSC, induced pluripotents stem cell; kd, knock-down; P, postnatal day; NPC, neural progenitor cell; NSC, neural stem cell; scRNA-seq, single cell RNA sequencing; shRNA, single hairpin RNA.*

Reference	Method	Organism	Tissue/ Stage
Forrest et al., 2013	Microarray, shRNA TCF4 kd	Human SH-SY5Y cells	Cell line
Kennedy et al., 2016	RNA-seq	*Tcf4*$^{+/-}$ mouse	Hippocampus/ Adult
Hill et al., 2017	Microarray, shRNA TCF4 kd	Human NPCs	Cell line
Hennig et al., 2017	Microarray, shRNA TCF4 kd	Human iPSCs	Differentiated NPCs
Moen et al., 2017	RNA-seq, shRNA Tcf4 kd	Mouse	NSCs
Li et al., 2019	RNA-seq	*Tcf4*$^{-/-}$ and $^{+/-}$ mouse	Dorsal telencephalon/ P0

Table 2. *Continued*

Reference	Method	Organism	Tissue/ Stage
Doostparast Torshizi et al., 2019	RNA-seq, shRNA TCF4 kd	SCZ patient iPSCs	Neurons/ P3, P14
Phan et al., 2020	RNA-seq	*Tcf4*$^{+/-}$	Prefrontal cortex/ P1, Adult
Phan et al., 2020	RNA-seq	*Tcf4*$^{+/\triangle 574-579}$ *Tcf4*$^{+/R579W}$ *Actin-Cre*::*Tcf4*$^{+/floxed}$ *Nestin-Cre*::*Tcf4*$^{+/floxed}$	Hemibrain/ P0-2, Adult
Schoof et al., 2020	RNA-seq	*Tcf4*$^{-/-}$ mouse	Forebrain/ Newborn
Mesman et al., 2020	RNA-seq	*Tcf4*$^{-/-}$ mouse	Cortex/ E14.5
Wittmann et al., 2021	scRNA-seq	*Tcf4*$^{-/-}$ long isoforms	Neocortex/ E18.5
Sarkar et al., 2021	RNA-seq	*Tcf4*$^{floxed/floxed}$	Hippocampus, Adult
Papes et al., 2022	scRNA-seq	PTHS patient iPSCs	Organoids

Additionally, TCF4 has been shown to repress *KCNQ1* and *SCN10a* expression (Rannals et al., 2016). Interestingly, *KCNQ1* expression is upregulated in PTHS patient iPSC-derived neurons (Papes et al., 2022). Another target gene of TCF4 is *GADD45G* (Sepp et al., 2017), which is also downregulated in PTHS NPCs (Papes et al., 2022). In addition, TCF4 also binds to promoter areas and intronic enhancer region of brain derived neurotrophic factor (Tuvikene et al., 2021; Esvald et al., 2022).

1.2.5 Regulation of activity

Binding of calmodulin inhibits the transcriptional activity of all the E-proteins (Saarikettu et al., 2004). It is also suggested that the activity of TCF4 and all the other E-proteins can be controlled by a post-translational mechanism affecting E-protein dimerization partners or E-proteins themselves. For example, phosphorylation of serines or threonines in the second helix of class II bHLH proteins can inhibit DNA binding (Quan et al., 2016). Additionally, phosphorylation of TCF4 at S448 by protein kinase A enhances neuronal-activity-dependent transcriptional activity of TCF4 (Sepp et al., 2017).

The E-box CANNTG target sequence of TCF4, its flanking area and possible single strand DNA modifications have been studied to understand the complex mechanism behind TCF4 mediated transcriptional regulation. First of all, the presence of dimerization partners of TCF4 affects the target binding sequence (Bertrand et al., 2002). In SH-SY5Y

cells, the main binding sequence for TCF4 is E-box motif CATCTG and the palindromic ATOH1 motif CAGCTG (Forrest et al., 2018). The preferred E-box motif of TCF4 dimerization partner ASCL1 is suggested to be CAGCTG (Castro et al., 2011). Secondly, protein binding studies have shown highest overall affinity of TCF4 for **C**AGGTGGT E-box sequence with methylation of the first cytosine (in bold) reducing DNA binding. Interestingly, while 5-methylated cytosines in E-box decreases TCF4 DNA binding, the presence of unmodified or 5-hydroxy methylated cytosines increases DNA binding of TCF4 (Khund-Sayeed et al., 2016). In addition, 5-carboxylation of a cytosine in the flanking area of E-box (in bold, **C**GCAGGTG) increases binding of TCF4 heterodimers with ASCL1 (Golla et al., 2014). These results have been confirmed by a later study which concluded that modifications in the first two flanking nucleotides of an E-box (in bold, **CG**CACGTG) increase TCF4 binding, modifications in the first two E-box nucleotides (in bold, CG**CA**CGTG) decrease TCF4 binding and modifications affecting the middle E-box nucleotides (in bold, CGCA**CG**TG) have little to no effect on TCF4 binding (Yang et al., 2019).

1.2.6 Gene structure and functional protein domains

Human and rodent *TCF4* gene structure is complex and results in many different transcripts due to the use of numerous promoters and alternative splicing (Sepp et al., 2011; Nurm et al., 2021). The human *TCF4* gene comprises of 41 exons out of which 21 are alternative 5' exons (Figure 1A) (Sepp et al., 2011). Meanwhile, the mouse *Tcf4* gene contains 33 exons out of which 14 are alternative 5' exons (Nurm et al., 2021). At least 18 and 7 N-terminally distinct TCF4 protein isoforms are encoded in human and mouse, respectively. The presence of different protein isoform encoding transcripts in the human and rodent nervous system is similar with one exception – translation of 5' exon 7bI containing transcripts results in protein isoform TCF4-G in humans (Figure 1B) but TCF4-D in mice due to the addition of a nucleotide in the mouse 5' exon 7bI which causes a frameshift and results in the use of more downstream translation initiation site in internal exon 8 (Sepp et al., 2011; Nurm et al., 2021).

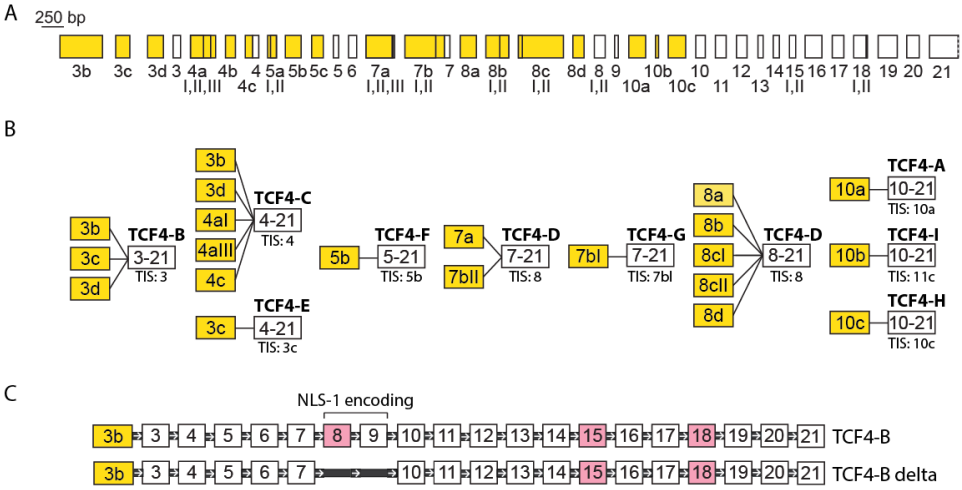

Figure 1. Human TCF4 gene structure. (A) *TCF4 exons expressed in the human central nervous system shown in scale with the scale bar marked at the top. The name of each exon is shown below with roman numerals indicating multiple splice sites. White boxes represent internal and 3' exons and yellow boxes represent 5' exons.* ***(B)*** *Different transcripts of human TCF4 arising from the use of many transcription start sites before the self-exclusive 5' exons. Yellow boxes are 5' exons with their names inside and the lines indicate individual splicing of the respective 5' exon to the internal exon. Internal exons are shown as white boxes with the numbers indicating the first and last internal exon present in the transcript. Translation initiation sites (TIS) are shown below the internal exons together with the respective translated TCF4 protein isoform shown above.* ***(C)*** *Splicing of TCF4 internal exons 8 and 9 results in transcripts encoding NLS-1 or lacking NLS-1 (also known as delta protein isoforms). Splicing of TCF4-B encoding transcript arising from the use of 5' exon 3b is shown for reference. Internal exons with alternative splicing sites are marked in pink, others in white. Based on data from Sepp et al.(2011). TIS, translation initiation site.*

The overall number of transcripts from the *TCF4* gene is much higher than the number of N-terminally distinct protein isoforms due to non-coding 5' exons and alternative splicing of internal exons. Transcripts from the *TCF4* gene can be grouped as "+" and "-" isoforms based on whether they include the Arg-Ser-Arg-Ser (RSRS) coding sequence ("+" isoforms) or not ("-" isoforms). Skipping of exons 8 and 9 results in delta isoforms which lack nuclear localization signal (NLS-1) (Figure 1C) (Sepp et al., 2011).

TCF4 protein isoforms contain three activation domains (AD1-3) out of which AD1 is present only in longer TCF4 isoforms (TCF4-B, -J, -K and -L, partially in TCF4-C). The remaining activation domain 2 (AD2) and 3 (AD3) are present in all the TCF4 isoforms. AD1 binds transcriptional co-activators p300/CBP and STAGA, and co-repressor ETO (Bayly et al., 2004; Zhang et al., 2004; Guo et al., 2009; Denis et al., 2012). p300/CBP interaction has also been shown for AD2 (Bayly et al., 2004; Denis et al., 2012). AD3 interacts with the TAF4 subunit of transcription factor IID (Chen et al., 2013). A conserved element located between AD1 and AD3 regulates the activity of AD1 (Herbst and Kolligs, 2008) and a repression domain between AD2 and bHLH domain can repress both AD1 and AD2 (Markus et al., 2002). The bHLH domain contains a basic sequence which mediates DNA binding and a HLH region which is necessary for dimerization (reviewed in Teixeira et al., 2021). The C domain is involved in dimerization (Goldfarb et al., 1998).

In addition to the previously mentioned NLS-1, the bHLH region of TCF4 contains a second NLS-2 and two nuclear export signals (NES-1 and NES-2) (Greb-Markiewicz et al., 2019). All the functional protein domains of TCF4 are shown in Figure 2.

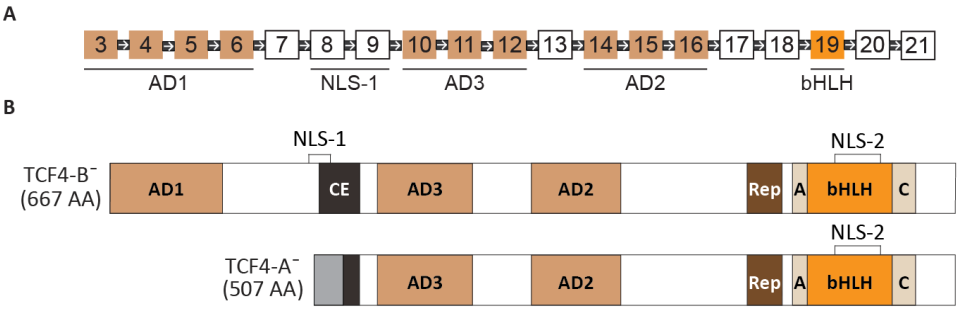

Figure 2. Schematic representation of TCF4 protein domains*. (A) TCF4 internal exons with the functional domain-encoding exons marked below the exon names. (B) Schematic representation of protein domains for TCF4 isoforms B and A. The colored areas represent functional protein domains – activation domains are light brown, conserved element is black, repression domain is dark brown, A and C domains are beige, bHLH domain is in orange and the unique region for TCF4-A encoded by exon 10a is shown in grey. The nuclear localization signals are shown above and the names of the respective TCF4 protein isoforms with their lengths in amino acids is on the left. AD, activation domain; NLS, nuclear localization signal; bHLH, basic helix-loop-helix domain; CE, conserved element; Rep, repression domain; AA, amino acid.*

Due to the presence of so many functional protein domains, the many TCF4 protein isoforms exhibit differing transcription activation capabilities in *in vitro* reporter assays (Sepp et al., 2011, 2017; Nurm et al., 2021). In addition, *in vivo* experiments have revealed that while overexpression of TCF4-B disrupts the distributions of pyramidal cells in the developing rat cortex, overexpression of TCF4-B lacking AD2 does not (Page et al., 2018). The distinct functions of all the TCF4 isoforms remain to be studied.

1.2.7 Expression

The first studies on *TCF4* expression used northern blot analysis and *in situ* hybridization to confirm that *TCF4* is expressed in neural and nonneural tissues in both rodents and human (Soosaar et al., 1994; Pscherer et al., 1996). More recent studies have used quantitative reverse-transcription PCR, digital-droplet PCR and RNA-seq to describe *Tcf4* expression through development. By now it is known that total *Tcf4* expression is highest in the mouse cerebral cortex during late prenatal and early postnatal development (Li et al., 2019; Phan et al., 2020). In human, *in situ* hybridization and quantitative reverse-transcription PCR have shown that *TCF4* mRNA is widely expressed in both neural and nonneural tissues (Pontual et al., 2009; Sepp et al., 2011). Further RNA-seq analysis of human tissues showed that *TCF4* expression peaks during late prenatal development of the cerebral cortex (Ma et al., 2018).

TCF4 protein expression in mouse has been characterised with immunostaining using antibodies specific for TCF4 (Jung et al., 2018; Sarkar et al., 2021) and with the use of mice expressing GFP-tagged TCF4 (Kim et al., 2020). Taken together, these studies showed that TCF4 protein is expressed in multiple brain regions of the developing and adult mouse with highest expression levels in the olfactory bulb, cerebral cortex, hippocampus and cerebellum (Jung et al., 2018; Kim et al., 2020). At the cellular level,

TCF4 is highly expressed in the inhibitory (GABAergic) and excitatory (glutamatergic) neurons of the adult cortex, hippocampus, striatum and cerebellum (Kim et al., 2020; Sarkar et al., 2021). In addition, TCF4 protein expression is high in cortical astrocytes and oligodendrocytes (Kim et al., 2020).

Several regulatory mechanisms have been described that control TCF4 expression. The expression of *TCF4* is negatively regulated by miRNA-137 (Ripke et al., 2011). Meanwhile, administration of suberoylanilide hydroxamic acid (SAHA), a histone deacetylase inhibitor, upregulates *TCF4* expression in both human NPCs (Hennig et al., 2017) and *Tcf4* haploinsufficient mice (Kennedy et al., 2016).

1.3 TCF4-related diseases

It is well known that mutations which cause *TCF4* haploinsufficiency result in a rare but severe autism spectrum disorder Pitt-Hopkins syndrome (Zweier et al., 2007; Zollino et al., 2019). Aberrations within *TCF4* have also been tied to many other neurocognitive disorders such as schizophrenia (Stefansson et al., 2009), mild-moderate intellectual disability (Kharbanda et al., 2016; Maduro et al., 2016) and autism (Stessman et al., 2017). In addition, *TCF4* has been associated with the development of posttraumatic stress disorder (Gelernter et al., 2019), major depression (Wray et al., 2018), Fuchs Endothelial corneal dystrophy (Wieben et al., 2012) and cancer (Kolligs et al., 2002).

1.3.1 Pitt-Hopkins Syndrome

Pitt-Hopkins syndrome (PTHS) is a rare (prevalence 1:300 000) neurodevelopmental disorder caused by *de novo* autosomal dominant mutations in *TCF4*. Symptoms of PTHS include developmental delay, intellectual disability, breathing anomalies, limited speech, motor delay, epilepsy, gastrointestinal disturbances and distinct facial features (Zollino et al., 2019). PTHS-related mutations usually involve large deletions and translocations but also frameshift, nonsense and missense mutations which result in *TCF4* haploinsufficiency, meaning that only one allele of the *TCF4* gene is functional but is not sufficient to produce enough TCF4 protein. In about 20% of cases, just a single nucleotide mutation which affects one amino acid can cause the expression of an unfunctional TCF4 protein with dominant-negative effects (Brockschmidt et al., 2007; Zweier et al., 2007, 2008; Zollino et al., 2019). Pathogenic single nucleotide mutations cluster in the bHLH region of TCF4, also known as a hotspot for missense mutations associated with PTHS (Whalen et al., 2012), and these mutations can impair or completely abrogate transcriptional activity, DNA binding or heterodimerization capability of the protein (Sepp et al., 2012). Mutations in *CNTNAP2* and *NRXN1* cause a disorder with a similar phenotype to PTHS (Zweier et al., 2009) and more importantly, data suggests that *NRXN1* is possibly a target gene of TCF4 (Moen et al., 2017).

Skin fibroblasts of PTHS patient have been used to generate iPSCs to study changes in PTHS NPCs and organoids. PTHS patient fibroblasts and NPCs have reduced *TCF4* expression and downregulated Wnt signaling pathway genes (Hennig et al., 2017; Papes et al., 2022). PTHS organoids have aberrant structure, morphology, neuronal content and transcriptome, show a higher percentage of NPCs with possibly impaired ability to proliferate and differentiate into neurons (Papes et al., 2022).

PTHS model mice have been studied extensively using different behavioural, histological and sequencing techniques (Grubišić et al., 2015; Kennedy et al., 2016; Thaxton et al., 2018; Li et al., 2019; Mesman et al., 2020; Phan et al., 2020; Wittmann et al., 2021). Even the effect of the most prevalent PTHS missense mutation in human

(R580W, mouse R579W) have been studied in a heterozygous mouse background. Compared to *Tcf4* heterozygous mouse, the R579W mutant mouse displays similar phenotypes including smaller body and brain weight, hyperactivity, reduced anxiety, deficits in memory and learning, and N-methyl-D-aspartate receptor mediated enhanced hippocampal LTP (Thaxton et al., 2018). However, the R579W mutant does not display deficits in habituation which have been described for *Tcf4* heterozygous knock-out mice (Kennedy et al., 2016; Thaxton et al., 2018). Comparison of these *Tcf4* mutant mice revealed that the R579W mutant mice show reduced intrinsic excitability but increased prepulse inhibition in 7-11 week old mice compared to *Tcf4* heterozygous mice (Thaxton et al., 2018). Overall, the current PTHS mouse models seem to mimic the human disease quite well as even the most common non-neurological symptom – abnormal gut function – is present in *Tcf4* haploinsufficient mice (Grubišić et al., 2015). Homozygous mice lacking only longer TCF4 isoforms show reduced cortical thickness and a smaller dentate gyrus, and agenesis of the splenium of corpus callosum which resembles the anomalies seen in human PTHS patients (Jung et al., 2018).

I*n vitro* studies suggest that the use of histone deacetylase inhibitors other than SAHA combined with the activation of Wnt signaling can increase *TCF4* expression and rescue the aberrant phenotype of PTHS organoids (Hennig et al., 2017; Papes et al., 2022). In addition for potential therapeutic applications, it is known that administration of SAHA alleviates the deficits in memory and learning of PTHS mice (Kennedy et al., 2016). In addition, blocking Nav1.8 channels or silencing *Scn10* expression improves the PTHS phenotype of abnormal breathing and locomotion of *Tcf4* haploinsufficient mice (Ekins et al., 2019; Cleary et al., 2021).

1.3.2 Schizophrenia

Schizophrenia is a severe psychiatric disorder characterized by delusions, cognitive deficits and affective retraction. The association between SCZ and TCF4 was first revealed by a genome wide association study which revealed a single nucleotide polymorphism located in intron three of TCF4 (rs9960767) as a risk allele for the development of SCZ (Stefansson et al., 2009). To date, more SCZ associated mutations have been located in introns, exons and intragenic regions of TCF4 (Ripke et al., 2011; Steinberg et al., 2011; Hu et al., 2014; Basmanav et al., 2015; Li et al., 2016).

SCZ-associated single nucleotide polymorphisms in *TCF4* influence auditory sensory gating (only in heavy smoking individuals) (Quednow et al., 2012) and verbal memory (Lennertz et al., 2011) of SCZ patients. These findings are supported by studies in mice showing that both overexpression of *Tcf4* in the postnatal brain (Brzózka et al., 2010) and *Tcf4* haploinsufficiency (Kennedy et al., 2016) causes SCZ-specific defects in fear memory formation and sensorimotor gating. In addition, SCZ-associated missense mutations in TCF4 alter the transcriptional activity of TCF4 *in vitro* (Sepp et al., 2017).

Association between TCF4 with the development of SCZ is also supported by the connection to miRNA-137, which is an important regulator of neuronal maturation and one of the top risk genes for SCZ. Studies on miRNA-137 target genes have revealed that *TCF4* expression is negatively regulated by miRNA-137 (Ripke et al., 2011; reviewed in Wright et al., 2013).

Earlier studies on *TCF4* expression in SCZ have shown that *TCF4* mRNA expression levels are about 55% lower in the adult post-mortem cerebellum of SCZ patients compared to healthy controls (Mudge et al., 2008). In addition, a slightly higher (2.22%) *TCF4* mRNA expression has also been described in the blood of SCZ patients (Wirgenes et al., 2012).

Studies with neurons derived from human iPSCs of SCZ patients have revealed highly increased *TCF4* mRNA expression (>2 fold) (Brennand et al., 2011).

More recent studies confirm that TCF4 is a major risk factor in the development of SCZ (Doostparast Torshizi et al., 2019; Ruzicka et al., 2020) and it has been shown that TCF4 has binding sites in several SCZ risk loci (Xia et al., 2018). Bulk and scRNA-seq of the human prefrontal cortex from SCZ patients show that TCF4 mRNA expression is upregulated in at least 14 different cell types including inhibitory and excitatory neurons, oligodendrocytes and microglia (Ruzicka et al., 2020).

As a potential cure for TCF4 associated SCZ, it has been suggested that administration of spironolactone or aripiprazole may help as administration of these drugs to a *Tcf4* transgenic SCZ model mouse reduced the SCZ-like cognitive deficits. However, in combination therapy where mice are treated with both drugs simultaneously, the positive effects described for single drug treatments were reduced (Stephan et al., 2022).

1.3.3 Mild to moderate intellectual disability

Mutations in *TCF4* which affect only the longer TCF4 isoforms cause mild to moderate intellectual disability (MMID). The resulting phenotype is less severe compared to PTHS–patients can live independently with only little support. The symptoms include dysmorphic features, developmental delay and learning difficulties. The known mutations affecting *TCF4* in patients with MMID are a translocation between chromosome 20 and 18 (Kalscheuer et al., 2008) and between chromosome 14 and 18 (Maduro et al., 2016), and a heritable deletion which affects *TCF4* exons 1a-4 (Kharbanda et al., 2016).

1.3.4 Fuchs endothelial corneal dystrophy

Fuchs endothelial corneal dystrophy (FECD) was described in the beginning of the 20th century by professor Ernst Fuchs. The disease is characterized by loss of corneal endothelial cells and accumulation of extracellular matrix and formation of guttae in the descemet membrane. It culminates with corneal edema and heavily decreased visibility. Currently, the only treatment for FECD is surgical corneal transplantation which relies on the availability of donor material. The occurrence of FECD is around 5% among persons over 40 in Europe and United states. Interestingly, the male-to-female prevalence ratio is about 1:3 (Fautsch et al., 2020).

FECD can be categorized into early- and late-onset form. While early-onset FECD is inheritable in an autosomal dominant fashion and associated with mutations in the *COL8A2* gene(Gottsch et al., 2005), the late-onset form has been associated with defects in *TCF4*. Linkage between chromosome 18q21 (location of human *TCF4*) and late-onset FECD was first described in 2006 (Sundin et al., 2006). A later study by Wieben and others showed that the CTG trinucleotide repeat expansion in intron 3 of *TCF4* is the causative mutation of late-onset FECD. When measured from blood, FECD patients have CTG repeat lengths >50 while healthy controls carry around 12-18 CTG repeats (Wieben et al., 2012).

1.3.5 Tumorigenesis

The first indication that TCF4 may be involved in cancer came from a study showing that TCF4 is a downstream target of the Wnt/b-catenin pathway (Kolligs et al., 2002) – mutations in that pathway have been associated with many different cancer types (Wang et al., 2021). However, different studies have categorized TCF4 as both an oncogene and a tumour suppressor, depending on the cellular context.

As an oncogene, *Tcf4* overexpression results in tumour like migratory and invasive phenotype of cancer cells *in vitro* (Sobrado et al., 2009; Appaiah et al., 2010).

The invasive phenotype is believed to arise as a result of indirect repression of the expression of the cell-cell adhesion protein E-cadherin by TCF4, providing further evidence for the involvement of TCF4 in the process of EMT (Sobrado et al., 2009). Silencing of *TCF4*, however, reduced the tumour like invasive phenotype of breast cancer cells (Appaiah et al., 2010). Progression of tumours is regularly studied by injecting cancer cells to rodent. Injection of colon cancer cells where both b-catenin and TCF4 were silenced resulted in almost no tumour growth compared to the injection of control cancer cells. In addition, activation of b-catenin and silencing of TCF4 in an already formed induced colon tumour caused almost total regression of the tumour within a few months (Mologni et al., 2010).

As a tumour suppressor, TCF4 is known to be negative regulator of Wnt/β-catenin signaling and can thus negatively regulate cell proliferation. This happens through a regulatory loop which controls expression of TCF4 via the Wnt/β-catenin pathway. While the b-catenin/TCF7L2 complex induces TCF4 expression, TCF4 itself interferes with the formation of the b-catenin/TCF7L2 complex (Shin et al., 2014).

TCF4 is frequently mutated in sporadic sonic-hedgehog associated medulloblastomas (Kool et al., 2014) and a study suggests that TCF4 is involved in downregulation cell proliferation in such sonic hedgehog positive subtype of medulloblastomas (Hellwig et al., 2019). Reduced *TCF4* expression has been described in many cancer types including lung, gastric and ovarian cancer (Kim et al., 2008; Pernía et al., 2020).

2 Aims of the study

The aim of this thesis was to study the expression and functioning of TCF4. More specific aims of the study were as follows:

- Study mRNA and protein expression of the different TCF4 isoforms during rodent and human development.
- Study how FECD-associated CTG trinucleotide repeat in the third intron of the *TCF4* gene modulates *TCF4* expression.
- Study how SCZ, MMID, RTT-like syndrome and PTHS related missense mutations in *TCF4* impact the functionality of TCF4 protein.

3 Materials and Methods

The following methods were used in this study and are described in more detail in the respective publications:

- Cell culture of cell lines (HEK293, Neuro2a, SH-SY5Y) – Publications I, II and III.
- Cell culture of primary cells (rat primary cortical and hippocampal neurons) – Publications II and III.
- Molecular cloning – Publications I, II and III
- Site-directed mutagenesis – Publication II
- *In vitro* protein translation – Publications I and II
- RNA extraction, cDNA synthesis, PCR – Publication II
- Western blot analysis – Publication I, II and III
- Transfection of cells – Publications I, II and III
- Luciferase reporter assay – Publications II and III
- Direct TCF4 RNA sequencing – Publication I
- Analysis of publicly available RNA-seq datasets – Publications I and II
- Bioinformatic analysis of gene structure, cap sites, sequence variation – Publications I, II and III
- Bioinformatical analysis – Publication I and II
- CRISPR/ Cas9-mediated gene mutation – Publication I
- Animal husbandry – Publication I
- Collection of neuronal and nonneuronal tissues throughout rodent development – Publication I
- Protein extraction from rodent and human tissues – Publication I
- 5' Rapid amplification of cDNA ends– Publication II
- Electrophoretic mobility shift assay – Publication III
- Immunocytochemistry – Publication III

4 Results

4.1 Results obtained in publication I

- In mouse cerebral cortex, mRNAs transcribed from *Tcf4* gene encode isoforms TCF4-B, -C, -D, -A and -I;
- *Tcf4* mRNA expression peaks around birth in mouse and rat;
- *Tcf4* mRNA expression is highest in the mouse and rat cerebral cortex, hippocampus and cerebellum;
- The majority of *Tcf4* transcripts in mouse and rat neural tissues encode TCF4-A.
- *Tcf4* mRNA expression levels are similar in mouse and rat nonneural tissues except for the liver, where *Tcf4* levels are almost non existant;
- TCF4-A encoding transcripts account for the majority of *Tcf4* transcripts in mouse nonneural tissues;
- *Tcf4* protein expression peaks around birth in mouse and rat;
- Expression of TCF4 protein is highest in mouse and rat cerebral cortex, hippocampus, cerebellum and olfactory bulb;
- Long and short TCF4 protein isoforms are expressed in all rat and mouse neural tissues;
- TCF4-D protein expression is high in the cerebral cortex and hippocampus and low or undetectable in other neural tissues of mouse and rat;
- *TCF4* mRNA expression peaks during human embryonic development and decreases after birth;
- *TCF4* mRNA expression is higher in human brain compared to nonneural tissues;
- Transcripts encoding TCF4-A are expressed at highest levels in human tissues, except for the testis, where TCF4-J encoding transcripts are expressed at highest levels starting from adolescence;
- Expression of *TCF4* long, medium and short protein isoforms can be detected in adult human cerebral cortex and hippocampus.

4.2 Results obtained in publication II

- There are numerous transcription start sites in the intron between *TCF4* internal exons 3 and 4;
- The CTG trinucleotide repeat is located upstream of the 5' UTR coding region of *TCF4* exons 4a, 4b and 4c;
- Activity of TCF4 promoters immediately downstream of the CTG trinucleotide repeat decreases with increasing CTG repeat length – significant decreases were observed from promoters with >50 CTG repeats;
- In FECD patients, an expanded CTG trinucleotide repeat has contrasting effects on the expression of different *TCF4* transcripts – expression levels of transcripts under the control of promoters near the repeat region decline while expression of certain transcripts starting further downstream of the repeat region increase.

4.3 Results obtained in publication III

- PTHS missense mutations alter intranuclear localization of TCF4 in a cell type dependent manner:
    - R569W and N585D mutants form condensed intranuclear puncta in neuron cultures but not in HEK293 cells;
- PTHS missense mutations impair or completely abrogate DNA binding:
    - R569W mutant homodimers show reduced DNA binding while N585D and A587P mutants show no DNA binding as homodimers;
    - Heterodimerization with ASCL1 can fully (R569W and N585D) or partly (A587P) alleviate the negative effects of PTHS mutations on DNA binding;
- PTHS missense mutations modulate transcriptional activity of TCF4;
    - R569W mutation:
        - Increases the transcriptional activity of TCF4-B homo- and heterodimers in HEK293 cells;
        - Decreases the activity of TCF4-A in all the studied conditions;
        - Increases the activity of TCF4-B heterodimers with ASCL1 in neuron cultures;
    - N585D and A587P mutations:
        - Decrease the transcriptional activity of TCF4 homodimers (N585D, A587P) and heterodimers (A587P) in HEK293 cells;
        - Almost completely abrogate the transcriptional activity of TCF4-A (N585D, A587P) and TCF4-B (A587P) in neuron cultures;
        - Increases the transcriptional activity of TCF4-B (N585D) heterodimers in basal conditions in neuron cultures;
- SCZ, MMID and RTT-like syndrome associated missense mutations and variations do not affect the functionality of TCF4 or have very mild effects in the *in vitro* cell and molecular biology assays used.

5 Discussion

5.1 Expression of TCF4 in rodent and human

The human and rodent *TCF4* gene structures are based on short read sequencing data (Sepp et al., 2011; Nurm et al., 2021). However, for a complex gene like *TCF4*, it can be complicated to describe all the 5' exons and alternative splicing effects using only short read data. Here, we used a long-read direct RNA sequencing approach to describe *Tcf4* transcripts in the rodent cerebral cortex. Overall, our data confirmed the previous results by Nurm et al. that transcription from the rodent *Tcf4* gene results in transcripts encoding at least 5 N-terminally different protein isoforms – TCF4-B, -C, -D, -A and -I (Nurm et al., 2021).

Most of the available RNA-seq data from the development of different organs and tissues is based on short reads which makes it difficult to quantify all the transcripts of *TCF4*. This is one of the reasons why only a few of the previous studies on *TCF4* mRNA expression have focused on describing the expression of transcripts encoding the distinct TCF4 protein isoforms, mostly by using quantitative reverse transcription PCR (Sepp et al., 2011). To distinguish and quantify *TCF4* transcripts from short read data, we developed a splice-site based analysis method which quantifies transcripts overlapping the various TCF4 splice-sites (Publication I). Our results show that the expression ratios between different *TCF4* isoform encoding transcripts remain relatively stable throughout rodent and human development. The only exception was seen in the human testis, where the expression of transcripts encoding TCF4-J is initiated during adolescence.

Studies on TCF4 protein expression have only focused on total TCF4 (Kim et al., 2020) or on long TCF4 isoforms(Jung et al., 2018). However, functional differences between TCF4 isoforms have been described. First of all, it is long known that different TCF4 protein isoforms have contrasting transactivation capabilities (Sepp et al., 2011, 2017). Previous explanations have attributed this phenomenon to the differential presence or absence of functional protein domains in the long and short isoforms of TCF4. A more recent study has shed more light on the topic, suggesting that longer and shorter TCF4 isoforms may interact with different interaction partners (Wittmann et al., 2021). The importance of studying the many TCF4 isoforms separately cannot be over-emphasised. For example, during rodent development, knock-out of longer TCF4 isoforms results in loss of forebrain commissure system that cannot be rescued by the remaining endogenous expression of shorter isoforms (Mesman et al., 2020; Wittmann et al., 2021). First of all, this means that the previously suggested compensatory mechanisms between E-proteins probably depends on context (Ravanpay and Olson, 2008). Secondly, it would be of interest to study whether the expression of shorter TCF4 isoforms increases in response to the loss of expression of long protein isoforms to confirm the suggested feedback loop which regulates TCF4 expression. This would also confirm that the loss of commissure system is not due to the reduced dosage of TCF4. Thirdly, it would be interesting to artificially enhance the expression of short TCF4 isoforms when longer isoforms are knocked-out to see whether it can rescue the negative phenotype.

We performed a comprehensive analysis of the expression of different TCF4 protein isoforms. We were able to distinguish the expression of long (TCF4-B, -C), medium (TCF4-D) and short (TCF4-A, -I) TCF4 isoforms, and describe the spatiotemporal expression pattern of these isoforms throughout rodent development. Interestingly, while the long and short protein isoforms were seen in both neural and nonneural tissues, the expression

of medium isoforms was only seen in the cerebral cortex, hippocampus and olfactory bulb. The medium isoforms are differentiated from the long isoforms by the absence of AD1, and from the short isoforms by the presence of NLS-1. Future studies on the expression of *Tcf4* isoforms at single cell level may help to understand the role of TCF4-D in neural tissues and different cell types.

As we now have extensive information about isoform-specific expression of TCF4 in neural and nonneural tissues (Publication I) we can suggest that the different isoforms may have different functions in different tissues and developmental stages. This is partly supported by previous studies where *Tcf4* reduction in embryonic development resulted in reduced neuronal firing of cortical neurons (Rannals et al., 2016), whereas a similar experiment in adult hippocampal neurons resulted in hyperexcitability (Sarkar et al., 2021). Also, *in vitro* experiments suggest that during differentiation of NPCs to neurons, the expression of TCF4-B, -C, -F, -G and -A encoding transcripts changes while the total TCF4 levels do not change (Hennig et al., 2017).

For a more detailed analysis of the function of TCF4 isoforms it would be possible to generate *Tcf4* null or tagged animals for distinct *Tcf4* protein isoforms that have their translation start sites located in alternative 5' exons, for example Tcf4-A (exon 10a). This would allow to study the functions, target genes and binding sites of each isoform separately. However, it should be taken into consideration that our experiments in Neuro2a cells show that silencing TCF4-A by causing a frameshift mutation in exon 10a leads to increased expression of TCF4-I.

It must be noted that results of different RNA-seq experiments can completely differ based on experimental conditions. For example, Sarkar et al. have compared RNA-seq data from *Tcf4* heterozygous mice which exhibit *Tcf4* haploinsufficiency unconditionally (Kennedy et al., 2016; Phan et al., 2020) with RNA-seq data from adult mice where total TCF4 knock-out has been induced in about two month old animals. In conclusion, these two datasets shared only about 15% differentially expressed genes (Sarkar et al., 2021).

When studying target genes using knock-down of *Tcf4* expression followed by RNA-seq, differentially expressed genes may also rise due to reduced expression from TCF4 target genes which in turn have their own target genes. The use of isoform-specifically tagged TCF4 combined with ChIP-seq is the best available option. This potentially reveals TCF4 isoform-specific target genes and confirms whether all the TCF4 isoforms bind the same or different target genes. If all the TCF4 isoforms bind the same target genes, then it remains to be studied how the transcriptional activity of TCF4 isoforms is regulated and whether there are any isoform-specific interaction partners or post translational modifications.

5.2 The effect of FECD associated CTG trinucleotide repeat expansion on the expression of *TCF4*

More than 40 diseases have been associated with the expansion of nucleotide repeat regions in both coding and noncoding regions of the genome (Nelson et al., 2013; Paulson, 2018). The effects of repeat region expansions have been extensively studied for several diseases. Expansion of the CGG repeat region in 5' UTR of *FMR1* gene is used for the diagnosis of fragile X-associated ataxia syndrome (FXTAS). The repeat expansion in FXTAS can have two outcomes depending on the length of the repeat expansion – a CGG expansion between 55-200 repeats causes increased expression of *FMR1*, while a CGG expansion >200 repeats results in hypermethylation and full transcriptional and translational silencing (Salcedo-Arellano et al., 2020). Friedrich ataxia is caused by an

expanded GAA repeat in the intron of *FXN* gene which results in reduced expression of the gene in patient-derived cells possibly due to hypermethylation – a longer repeat expansion causes a more severe reduction in mRNA expression levels of *FXN* (Castaldo et al., 2008; Chutake et al., 2014). A hexamer repeat expansion (GGGGCC) in the 5' region of the *C9ORF72* gene causes frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Experiments indicate that the GGGGCC repeat expansion reduces promoter activity in cell lines (Gijselinck et al., 2016). Taken together, previous data indicates that repeat expansions can both increase or reduce transcription of genes, which led us to study whether the FECD related CTG repeat expansion may cause changes in *TCF4* expression in patients with FECD (Publication II).

The CTG repeat is in the third intron of human TCF4 gene, upstream of the promoters regulating expression of *TCF4* 5' exons 4a, 4b and 4c. It has been shown that a CTG repeat expansion >40 increases the risk of developing FECD (Wieben et al., 2012; reviewed in Ong Tone et al., 2020). To study the CTG repeat, we first generated reporter constructs with the *TCF4* exon 4a, 4b and 4c promoter regions with differing CTG repeat lengths to drive the expression of a luciferase reporter gene. We report that *TCF4* CTG repeat affects the activity of nearby promoters in a length dependent manner – the longer the CTG repeat the lower the promoter activity.

Previous analyses on *TCF4* expression in patients with FECD have been conflicting. According to two studies, the expression of total TCF4 does not change significantly in patients with FECD (Ołdak et al., 2015; Mootha et al., 2017). However, Okumura et al. reported increased total *TCF4* expression and Foja et al. detected a decrease in the levels of *TCF4* transcripts beginning near the CTG repeat region (Foja et al., 2017; Okumura et al., 2019). Here, we set out to solve these contradictory results by analysing two previously published RNA-seq datasets from patients with FECD to study the expression of *TCF4*. Our data showed a slight increase in total *TCF4* expression as also reported by Okumura and others(Okumura et al., 2019). Next, we showed that indeed, the expression of *TCF4* transcripts beginning near the repeat region is reduced in FECD just as described by Foja and others (Foja et al., 2017). However, we also observed an increase in the expression of transcripts beginning more downstream of the repeat region. Another study of MMID patients showed that in the blood, where the expression of longer isoforms is absent due to a translocation in the 5' region of *TCF4*, the expression of medium and short isoform encoding transcripts is increased (Maduro et al., 2016). Such bidirectional regulation of different *TCF4* transcripts illustrates the importance of studying the expression of all the possible transcripts separately when working with genes with a complex gene structure like *TCF4*. This phenomenon also suggests that the upregulation of shorter TCF4 isoforms is likely to compensate for the loss of longer TCF4 isoforms. Thus, there may exist a feedback loop which controls TCF4 expression and tries to activate all the TCF4 promoters if the total TCF4 levels drop.

The CTG repeat expansion in TCF4 has also associated with vulnerability to bipolar disorder. A single study has shown that TCF4 CTG repeat expansion >40 is frequent in patients with a severe type of bipolar disorder (Del-Favero et al., 2002). Interestingly, severity of FECD correlates with the CTG repeat length as patients with longer repeat regions tend to have a more severe form of FECD (Soliman et al., 2015). It is possible that the CTG repeat expansion may also cause *TCF4* expression aberrations in the brain like the effect seen in the cornea of FECD patients but in the case of bipolar disorder, it only influences severity of bipolar disorder and is not a necessity for the generation of the disease.

5.3 The effect of disease-related missense variations and mutations in *TCF4* on the functionality of the protein

PTHS-causing missense mutations cluster in the bHLH region of *TCF4*, resulting in changes in transcription activation, dimerization and DNA binding (Zweier et al., 2007; Pontual et al., 2009; Forrest et al., 2012; Sepp et al., 2012). Amino acid substitutions caused by single nucleotide variations and mutations in *TCF4* have also been described in SCZ, MMID and Rett-like syndrome (Basmanav et al., 2015; Kharbanda et al., 2016; Srivastava et al., 2018). In publication III, we studied the effects of novel SCZ and MMID single nucleotide variations and PTHS and Rett-like syndrome single nucleotide mutations on transcription activation, nuclear localization and DNA binding of TCF4 protein.

Changes in *TCF4* gene sequence and function are considered to be a major risk factor in the development of SCZ (Stefansson et al., 2009; Doostparast Torshizi et al., 2019). However, little information is known how TCF4 mediates development of the disease. As there are six SCZ-associated missense variations in TCF4, we decided to study whether we can see a direct impact on the functionality of TCF4. Our results indicated that none of the six SCZ associated missense mutations altered DNA binding or nuclear localization of TCF4. However, three of the SCZ related missense variations (P299S, A315V and G428V) increased the transcriptional activity of TCF4. This result falls in line with previous observations that P299S and G428V variants increased activity of TCF4 (Sepp et al., 2017). It is interesting to note that SCZ-related P156T variant, which is in the beginning of NLS-1, did not cause aberrations of TCF4 functioning. P156T was the only SCZ missense variant that is located in a described functional domain of TCF4 – NLS-1.

Only one mutation (S253R) has been found in a patient with RTT-like syndrome. The symptoms described for the patient include intellectual disability and facial dysmorphisms like seen in PTHS patients (Srivastava et al., 2018). S253R variant was the only missense mutation outside the bHLH region which modulated transcription and DNA binding of TCF4. S253R is in AD3, which is important for assembling the transcriptional machinery by mediating the binding of TFIID (Chen et al., 2013). TFIID can modulate the activity of RNA polymerase II and stabilize E-proteins which helps to bind coactivators and -repressors of TCF4 (Juven-Gershon et al., 2008; Chen et al., 2013). As S253R reduced transcriptional activity of TCF4 in HEK293 cells but had no effect on the transcriptional activity of TCF4 in cultured neurons, it is possible that the effect rises from differential expression of co-activators and -repressors in the studied cell types. The binding of co-activators may be impaired, or the binding of co-repressors may be increased due to the mutation. As there is evidence that longer TCF4 isoforms bind different interaction partners compared to shorter isoforms (Wittmann et al., 2021), it would be interesting to study whether mutation S253R mediates such TCF4 long isoform specific interactions.

PTHS missense mutations had the most severe effects on the functionality of TCF4. For mutants R569W and N585D we saw aberrant intranuclear aggregations in neuron cultures which may arise from the fact that these mutations are in NLS-2 (Greb-Markiewicz et al., 2019). If these mutations cause dysfunction of NLS-2, then our results indicate that the presence of a functional NLS-1 is necessary for nuclear localization while NLS-2 may be involved in intranuclear localization of TCF4. This is also supported by experiments with TCF4-A which does not carry NLS-1 and is not strictly located to the nucleus in contrast to TCF4-B (Sepp et al., 2012). However, it is also possible that the nuclear aggregates may arise from protein misfolding or destabilization as suggested by previous studies (Sepp et al., 2012). Interestingly, the formation of

nuclear aggregates by mutant TCF4 proteins was not seen in HEK293 cells meaning that the phenomena is caused by cell-type specific effects.

PTHS mutations had more severe effects on transcription activation in the context of TCF4-A compared to TCF4-B. Expression from the TCF4 gene results in numerous transcripts which encode at least 18 N-terminally distinct TCF4 protein isoforms. However, we have only studied the mutations in the context of TCF4-B and TCF4-A. Such an approach is acceptable in the case for PTHS mutations which are almost always clustered in the bHLH region and for mutations in TCF4 common exons 10-21 which are present in all the TCF4 isoforms. However, some of the studied mutations (N90S, R114K, P156T) are only present in a subset of longer TCF4 isoforms – TCF4-B, -C, -E, -F, -D and -G (Sepp et al., 2011). Currently, it is unknown whether these mutations may have isoform specific effects on the functioning of TCF4 and it would be interesting to study the effects of these mutations in the context of other major TCF4 isoforms – widely expressed TCF4-C and brain specific TCF4-D. In addition, as TCF4 expression levels are highest around birth (Jung et al., 2018; Ma et al., 2018), it would be interesting to study the effects of missense mutations *in vivo* when *TCF4* expression levels are highest.

When studying changes in amino acid sequences *in vitro* it is necessary to keep in mind that all the effects seen can be dependent on the experimental conditions. *In vivo* studies can always provide more pronounced results when considering the presence of all the different cell types and developmental stages. In the case of SCZ, the changes in TCF4 functionality necessary for the development of the disease may be very small but still lead to complex unknown downstream effects (Ripke et al., 2011). This may also be the case for the MMID associated variations in *TCF4 (*N90S, R114K) located in the 5' end of *TCF4* (Kalscheuer et al., 2008; Kharbanda et al., 2016) as they had no effect on the functionality of TCF4 in our experiments.

It is suggested that the function of TCF4 is regulated by the expression pattern of the many interaction partners (Quednow et al., 2014) meaning that the effect of dimerization partners on the functionality of TCF4 cannot be underestimated. Our reporter assays showed that dimerization of TCF4 PTHS mutants with ASCL1 can enhance TCF4 mediated transcription activation to much higher levels compared to the wt protein. This indicates that PTHS missense mutations may also result in gain of function effects contrary to the usual belief that PTHS develops in response to loss of function of TCF4. The complex network of TCF4 regulation by its interaction partners remains to be studied in more detail.

6 Conclusions

- 5 N-terminally distinct *Tcf4* protein isoforms are expressed in rodent cerebral cortex (TCF4-B, -C, -D, -A and -I);
- *Tcf4* mRNA and protein expression in rodent is much lower in nonneural tissues compared to the brain;
- Medium TCF4 protein isoforms (TCF4-D) are highly expressed in rodent brain and almost undetectable in nonneural tissues;
- *TCF4* expression levels in human are highest in the brain and very low in the liver with most of the transcripts (>40%) encoding TCF4-A;
- The CTG trinucleotide repeat expansion in *TCF4* exon 3 decreases the activity of immediate downstream *TCF4* promoters;
- The CTG trinucleotide repeat does not significantly reduce the expression of total TCF4 due to the simultaneous decrease in the expression of longer isoforms and increase in the expression of shorter isoforms;
- PTHS-associated missense mutations alter TCF4 intranuclear localization and can completely impair DNA binding and transcriptional activity of TCF4. These effects cannot be rescued by dimerization partners;
- SCZ, MMID and RTT-like syndrome-associated missense mutations and variations have mild to no effects on the functioning of TCF4.

7 References

Appaiah, H., Bhat-Nakshatri, P., Mehta, R., Thorat, M., Badve, S., and Nakshatri, H. (2010). ITF2 is a target of CXCR4 in MDA-MB-231 breast cancer cells and is associated with reduced survival in estrogen receptor-negative breast cancer. *Cancer Biol. Ther.* 10, 600–614. doi: 10.4161/cbt.10.6.12586.

Basmanav, F. B., Forstner, A. J., Fier, H., Herms, S., Meier, S., Degenhardt, F., et al. (2015). Investigation of the role of TCF4 rare sequence variants in schizophrenia. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* 168B, 354–362. doi: 10.1002/ajmg.b.32318.

Bayly, R., Chuen, L., Currie, R. A., Hyndman, B. D., Casselman, R., Blobel, G. A., et al. (2004). E2A-PBX1 interacts directly with the KIX domain of CBP/p300 in the induction of proliferation in primary hematopoietic cells. *J. Biol. Chem.* 279, 55362–55371. doi: 10.1074/jbc.M408654200.

Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990). The protein Id: A negative regulator of helix-loop-helix DNA binding proteins. *Cell* 61, 49–59. doi: 10.1016/0092-8674(90)90214-Y.

Bergqvist, I., Eriksson, M., Saarikettu, J., Eriksson, B., Corneliussen, B., Grundström, T., et al. (2000). The basic helix-loop-helix transcription factor E2-2 is involved in T lymphocyte development. *Eur. J. Immunol.* 30, 2857–2863. doi: 10.1002/1521-4141(200010)30:10<2857::AID-IMMU2857>3.0.CO;2-G.

Bertrand, N., Castro, D. S., and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* 3, 517–530. doi: 10.1038/nrn874.

Blake, D. J., Forrest, M., Chapman, R. M., Tinsley, C. L., O'Donovan, M. C., and Owen, M. J. (2010). TCF4, schizophrenia, and Pitt-Hopkins Syndrome. *Schizophr. Bull.* 36, 443–447. doi: 10.1093/schbul/sbq035.

Brennand, K. J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., et al. (2011). Modelling schizophrenia using human induced pluripotent stem cells. *Nature* 473, 221–225. doi: 10.1038/nature09915.

Brockschmidt, A., Todt, U., Ryu, S., Hoischen, A., Landwehr, C., Birnbaum, S., et al. (2007). Severe mental retardation with breathing abnormalities (Pitt–Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. *Hum. Mol. Genet.* 16, 1488–1494. doi: 10.1093/hmg/ddm099.

Brzózka, M. M., Radyushkin, K., Wichert, S. P., Ehrenreich, H., and Rossner, M. J. (2010). Cognitive and sensorimotor gating impairments in transgenic mice overexpressing the schizophrenia susceptibility gene Tcf4 in the brain. *Biol. Psychiatry* 68, 33–40. doi: 10.1016/j.biopsych.2010.03.015.

Castaldo, I., Pinelli, M., Monticelli, A., Acquaviva, F., Giacchetti, M., Filla, A., et al. (2008). DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. *J. Med. Genet.* 45, 808–812. doi: 10.1136/jmg.2008.058594.

Castro, D. S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., et al. (2011). A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev.* 25, 930–945. doi: 10.1101/gad.627811.

Chen, T., Wu, Q., Zhang, Y., Lu, T., Yue, W., and Zhang, D. (2016). Tcf4 Controls Neuronal Migration of the Cerebral Cortex through Regulation of Bmp7. *Front. Mol. Neurosci.* 9. doi: 10.3389/fnmol.2016.00094.

Chen, W.-Y., Zhang, J., Geng, H., Du, Z., Nakadai, T., and Roeder, R. G. (2013). A TAF4 coactivator function for E proteins that involves enhanced TFIID binding. *Genes Dev.* 27, 1596–1609. doi: 10.1101/gad.216192.113.

Chutake, Y. K., Lam, C., Costello, W. N., Anderson, M., and Bidichandani, S. I. (2014). Epigenetic Promoter Silencing in Friedreich Ataxia is Dependent on Repeat Length. *Ann. Neurol.* 76, 522–528. doi: 10.1002/ana.24249.

Cisse, B., Caton, M. L., Lehner, M., Maeda, T., Scheu, S., Locksley, R., et al. (2008). Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell* 135, 37–48. doi: 10.1016/j.cell.2008.09.016.

Cleary, C. M., James, S., Maher, B. J., and Mulkey, D. K. (2021). Disordered breathing in a Pitt-Hopkins syndrome model involves Phox2b-expressing parafacial neurons and aberrant Nav1.8 expression. *Nat. Commun.* 12, 5962. doi: 10.1038/s41467-021-26263-2.

Corneliussen, B., Thornell, A., Hallberg, B., and Grundström, T. (1991). Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. *J. Virol.* 65, 6084–6093.

Crux, S., Herms, J., and Dorostkar, M. M. (2018). Tcf4 regulates dendritic spine density and morphology in the adult brain. *PLoS ONE* 13. doi: 10.1371/journal.pone.0199359.

Del-Favero, J., Gestel, S. V., Børglum, A. D., Muir, W., Ewald, H., Mors, O., et al. (2002). European combined analysis of the CTG18.1 and the ERDA1 CAG/CTG repeats in bipolar disorder. *Eur. J. Hum. Genet. EJHG* 10, 276–280. doi: 10.1038/sj.ejhg.5200803.

Denis, C. M., Chitayat, S., Plevin, M. J., Wang, F., Thompson, P., Liu, S., et al. (2012). Structural basis of CBP/p300 recruitment in leukemia induction by E2A-PBX1. *Blood* 120, 3968–3977. doi: 10.1182/blood-2012-02-411397.

Doostparast Torshizi, A., Armoskus, C., Zhang, H., Forrest, M. P., Zhang, S., Souaiaia, T., et al. (2019). Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. *Sci. Adv.* 5. doi: 10.1126/sciadv.aau4139.

Ekins, S., Gerlach, J., Zorn, K. M., Antonio, B. M., Lin, Z., and Gerlach, A. (2019). Repurposing Approved Drugs as Inhibitors of Kv7.1 and Nav1.8 to Treat Pitt Hopkins Syndrome. *Pharm. Res.* 36, 137. doi: 10.1007/s11095-019-2671-y.

Esvald, E.-E., Tuvikene, J., Moistus, A., Rannaste, K., Kõomägi, S., and Timmusk, T. (2022). Differential regulation of the BDNF gene in cortical and hippocampal neurons. *J. Neurosci.* doi: 10.1523/JNEUROSCI.2535-21.2022.

Fautsch, M. P., Wieben, E. D., Baratz, K. H., Bhattacharyya, N., Sadan, A. N., Hafford-Tear, N. J., et al. (2020). TCF4-mediated Fuchs endothelial corneal dystrophy: Insights into a common trinucleotide repeat-associated disease. *Prog. Retin. Eye Res.*, 100883. doi: 10.1016/j.preteyeres.2020.100883.

Fischer, B., Azim, K., Hurtado-Chong, A., Ramelli, S., Fernández, M., and Raineteau, O. (2014). E-proteins orchestrate the progression of neural stem cell differentiation in the postnatal forebrain. *Neural Develop.* 9, 23. doi: 10.1186/1749-8104-9-23.

Flora, A., Garcia, J. J., Thaller, C., and Zoghbi, H. Y. (2007). The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15382–15387. doi: 10.1073/pnas.0707456104.

Foja, S., Luther, M., Hoffmann, K., Rupprecht, A., and Gruenauer-Kloevekorn, C. (2017). CTG18.1 repeat expansion may reduce TCF4 gene expression in corneal endothelial cells of German patients with Fuchs' dystrophy. *Graefes Arch. Clin. Exp. Ophthalmol.* 255, 1621–1631. doi: 10.1007/s00417-017-3697-7.

Forrest, M., Chapman, R. M., Doyle, A. M., Tinsley, C. L., Waite, A., and Blake, D. J. (2012). Functional analysis of TCF4 missense mutations that cause Pitt–Hopkins syndrome. *Hum. Mutat.* 33, 1676–1686. doi: 10.1002/humu.22160.

Forrest, M. P., Hill, M. J., Kavanagh, D. H., Tansey, K. E., Waite, A. J., and Blake, D. J. (2018). The Psychiatric Risk Gene Transcription Factor 4 (TCF4) Regulates Neurodevelopmental Pathways Associated With Schizophrenia, Autism, and Intellectual Disability. *Schizophr. Bull.* 44, 1100–1110. doi: 10.1093/schbul/sbx164.

Forrest, M. P., Waite, A. J., Martin-Rendon, E., and Blake, D. J. (2013). Knockdown of human TCF4 affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. *PloS One* 8, e73169. doi: 10.1371/journal.pone.0073169.

Gelernter, J., Sun, N., Polimanti, R., Pietrzak, R., Levey, D. F., Bryois, J., et al. (2019). Genome-wide Association Study of Posttraumatic Stress Disorder (PTSD) Re-Experiencing Symptoms in >165,000 US Veterans. *Nat. Neurosci.* 22, 1394–1401. doi: 10.1038/s41593-019-0447-7.

Ghosh, H. S., Cisse, B., Bunin, A., Lewis, K. L., and Reizis, B. (2010). Continuous expression of the transcription factor e2-2 maintains the cell fate of mature plasmacytoid dendritic cells. *Immunity* 33, 905–916. doi: 10.1016/j.immuni.2010.11.023.

Gijselinck, I., Van Mossevelde, S., van der Zee, J., Sieben, A., Engelborghs, S., De Bleecker, J., et al. (2016). The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Mol. Psychiatry* 21, 1112–1124. doi: 10.1038/mp.2015.159.

Goldfarb, A. N., Lewandowska, K., and Pennell, C. A. (1998). Identification of a Highly Conserved Module in E Proteins Required for in Vivo Helix-loop-helix Dimerization *. *J. Biol. Chem.* 273, 2866–2873. doi: 10.1074/jbc.273.5.2866.

Golla, J. P., Zhao, J., Mann, I. K., Sayeed, S. K., Mandal, A., Rose, R. B., et al. (2014). Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. *Biochem. Biophys. Res. Commun.* 449, 248–255. doi: 10.1016/j.bbrc.2014.05.018.

Gottsch, J. D., Sundin, O. H., Liu, S. H., Jun, A. S., Broman, K. W., Stark, W. J., et al. (2005). Inheritance of a novel COL8A2 mutation defines a distinct early-onset subtype of fuchs corneal dystrophy. *Invest. Ophthalmol. Vis. Sci.* 46, 1934–1939. doi: 10.1167/iovs.04-0937.

Greb-Markiewicz, B., Kazana, W., Zarębski, M., and Ożyhar, A. (2019). The subcellular localization of bHLH transcription factor TCF4 is mediated by multiple nuclear localization and nuclear export signals. *Sci. Rep.* 9. doi: 10.1038/s41598-019-52239-w.

Grubišić, V., Kennedy, A. J., Sweatt, J. D., and Parpura, V. (2015). Pitt-Hopkins Mouse Model has Altered Particular Gastrointestinal Transits In Vivo. *Autism Res. Off. J. Int. Soc. Autism Res.* 8, 629–633. doi: 10.1002/aur.1467.

Guo, C., Hu, Q., Yan, C., and Zhang, J. (2009). Multivalent Binding of the ETO Corepressor to E Proteins Facilitates Dual Repression Controls Targeting Chromatin and the Basal Transcription Machinery. *Mol. Cell. Biol.* 29, 2644–2657. doi: 10.1128/MCB.00073-09.

Hellwig, M., Lauffer, M. C., Bockmayr, M., Spohn, M., Merk, D. J., Harrison, L., et al. (2019). TCF4 (E2-2) harbors tumor suppressive functions in SHH medulloblastoma. *Acta Neuropathol. (Berl.)* 137, 657–673. doi: 10.1007/s00401-019-01982-5.

Hennig, K. M., Fass, D. M., Zhao, W.-N., Sheridan, S. D., Fu, T., Erdin, S., et al. (2017). WNT/β-Catenin Pathway and Epigenetic Mechanisms Regulate the Pitt-Hopkins Syndrome and Schizophrenia Risk Gene TCF4. *Mol. Neuropsychiatry* 3, 53–71. doi: 10.1159/000475666.

Henthorn, P., Kiledjian, M., and Kadesch, T. (1990). Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. *Science* 247, 467–470.

Herbst, A., and Kolligs, F. T. (2008). A conserved domain in the transcription factor ITF-2B attenuates its activity. *Biochem. Biophys. Res. Commun.* 370, 327–331. doi: 10.1016/j.bbrc.2008.03.081.

Hill, M. J., Killick, R., Navarrete, K., Maruszak, A., McLaughlin, G. M., Williams, B. P., et al. (2017). Knockdown of the schizophrenia susceptibility gene TCF4 alters gene expression and proliferation of progenitor cells from the developing human neocortex. *J. Psychiatry Neurosci. JPN* 42, 181–188. doi: 10.1503/jpn.160073.

Hu, X., Zhang, B., Liu, W., Paciga, S., He, W., Lanz, T. A., et al. (2014). A survey of rare coding variants in candidate genes in schizophrenia by deep sequencing. *Mol. Psychiatry* 19, 858–859. doi: 10.1038/mp.2013.131.

Jung, M., Häberle, B. M., Tschaikowsky, T., Wittmann, M.-T., Balta, E.-A., Stadler, V.-C., et al. (2018). Analysis of the expression pattern of the schizophrenia-risk and intellectual disability gene TCF4 in the developing and adult brain suggests a role in development and plasticity of cortical and hippocampal neurons. *Mol. Autism* 9, 20. doi: 10.1186/s13229-018-0200-1.

Juven-Gershon, T., Hsu, J.-Y., Theisen, J. W. M., and Kadonaga, J. T. (2008). The RNA Polymerase II Core Promoter – the Gateway to Transcription. *Curr. Opin. Cell Biol.* 20, 253–259. doi: 10.1016/j.ceb.2008.03.003.

Kalscheuer, V. M., Feenstra, I., Van Ravenswaaij-Arts, C. M. A., Smeets, D. F. C. M., Menzel, C., Ullmann, R., et al. (2008). Disruption of the TCF4 gene in a girl with mental retardation but without the classical Pitt–Hopkins syndrome. *Am. J. Med. Genet. A.* 146A, 2053–2059. doi: 10.1002/ajmg.a.32419.

Kantzer, C. G., Yang, W., Grommisch, D., Patil, K. V., Mak, K. H.-M., Shirokova, V., et al. (2022). ID1 and CEBPA coordinate epidermal progenitor cell differentiation. *Development* 149, dev201262. doi: 10.1242/dev.201262.

Kennedy, A. J., Rahn, E. J., Paulukaitis, B. S., Savell, K. E., Kordasiewicz, H. B., Wang, J., et al. (2016). Tcf4 Regulates Synaptic Plasticity, DNA Methylation, and Memory Function. *Cell Rep.* 16, 2666–2685. doi: 10.1016/j.celrep.2016.08.004.

Kharbanda, M., Kannike, K., Lampe, A., Berg, J., Timmusk, T., and Sepp, M. (2016). Partial deletion of TCF4 in three generation family with non-syndromic intellectual disability, without features of Pitt-Hopkins syndrome. *Eur. J. Med. Genet.* 59, 310–314. doi: 10.1016/j.ejmg.2016.04.003.

Khund-Sayeed, S., He, X., Holzberg, T., Wang, J., Rajagopal, D., Upadhyay, S., et al. (2016). 5-hydroxymethylcytosine in E-Box motifs ACAT|GTG and ACAC|GTG increases DNA-binding of the B-HLH transcription factor TCF4. *Integr. Biol. Quant. Biosci. Nano Macro* 8, 936–945. doi: 10.1039/c6ib00079g.

Kim, H., Berens, N. C., Ochandarena, N. E., and Philpot, B. D. (2020). Region and Cell Type Distribution of TCF4 in the Postnatal Mouse Brain. *Front. Neuroanat.* 14, 42. doi: 10.3389/fnana.2020.00042.

Kim, S.-K., Jang, H.-R., Kim, J.-H., Kim, M., Noh, S.-M., Song, K.-S., et al. (2008). CpG methylation in exon 1 of transcription factor 4 increases with age in normal gastric mucosa and is associated with gene silencing in intestinal-type gastric cancers. *Carcinogenesis* 29, 1623–1631. doi: 10.1093/carcin/bgn110.

Kolligs, F. T., Nieman, M. T., Winer, I., Hu, G., Van Mater, D., Feng, Y., et al. (2002). ITF-2, a downstream target of the Wnt/TCF pathway, is activated in human cancers with β-catenin defects and promotes neoplastic transformation. *Cancer Cell* 1, 145–155. doi: 10.1016/S1535-6108(02)00035-1.

Kool, M., Jones, D. T. W., Jäger, N., Northcott, P. A., Pugh, T. J., Hovestadt, V., et al. (2014). Genome Sequencing of SHH Medulloblastoma Predicts Genotype-Related Response to Smoothened Inhibition. *Cancer Cell* 25, 393–405. doi: 10.1016/j.ccr.2014.02.004.

Lennertz, L., Rujescu, D., Wagner, M., Frommann, I., Schulze-Rauschenbach, S., Schuhmacher, A., et al. (2011). Novel Schizophrenia Risk Gene TCF4 Influences Verbal Learning and Memory Functioning in Schizophrenia Patients. *Neuropsychobiology* 63, 131–136. doi: 10.1159/000317844.

Li, H. S., Yang, C. Y., Nallaparaju, K. C., Zhang, H., Liu, Y.-J., Goldrath, A. W., et al. (2012). The signal transducers STAT5 and STAT3 control expression of Id2 and E2-2 during dendritic cell development. *Blood* 120, 4363–4373. doi: 10.1182/blood-2012-07-441311.

Li, H., Zhu, Y., Morozov, Y. M., Chen, X., Page, S. C., Rannals, M. D., et al. (2019). Disruption of TCF4 regulatory networks leads to abnormal cortical development and mental disabilities. *Mol. Psychiatry*. doi: 10.1038/s41380-019-0353-0.

Li, J., Chen, Z., Wang, F., Ouyang, Y., Zhang, N., Yang, M., et al. (2016). Polymorphisms of the TCF4 gene are associated with the risk of schizophrenia in the Han Chinese. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* 171, 1006–1012. doi: 10.1002/ajmg.b.32449.

Li, N., Liu, B., Wu, W., Hong, Y., Zhang, J., Liu, Y., et al. (2020). Upregulation of transcription factor 4 downregulates NaV1.8 expression in DRG neurons and prevents the development of rat inflammatory and neuropathic hypersensitivity. *Exp. Neurol.* 327, 113240. doi: 10.1016/j.expneurol.2020.113240.

Ma, C., Gu, C., Huo, Y., Li, X., and Luo, X.-J. (2018). The integrated landscape of causal genes and pathways in schizophrenia. *Transl. Psychiatry* 8. doi: 10.1038/s41398-018-0114-x.

Maduro, V., Pusey, B. N., Cherukuri, P. F., Atkins, P., du Souich, C., Rupps, R., et al. (2016). Complex translocation disrupting TCF4 and altering TCF4 isoform expression segregates as mild autosomal dominant intellectual disability. *Orphanet J. Rare Dis.* 11. doi: 10.1186/s13023-016-0439-6.

Markus, M., Du, Z., and Benezra, R. (2002). Enhancer-specific modulation of E protein activity. *J. Biol. Chem.* 277, 6469–6477. doi: 10.1074/jbc.M110659200.

Massari, M. E., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* 20, 429–440.

Mesman, S., Bakker, R., and Smidt, M. P. (2020). Tcf4 is required for correct brain development during embryogenesis. *Mol. Cell. Neurosci.* 106, 103502. doi: 10.1016/j.mcn.2020.103502.

Moen, M. J., Adams, H. H. H., Brandsma, J. H., Dekkers, D. H. W., Akinci, U., Karkampouna, S., et al. (2017). An interaction network of mental disorder proteins in neural stem cells. *Transl. Psychiatry* 7, e1082. doi: 10.1038/tp.2017.52.

Mologni, L., Dekhil, H., Ceccon, M., Purgante, S., Lan, C., Cleris, L., et al. (2010). Colorectal Tumors Are Effectively Eradicated by Combined Inhibition of β-Catenin, KRAS, and the Oncogenic Transcription Factor ITF2. *Cancer Res.* 70, 7253–7263. doi: 10.1158/0008-5472.CAN-10-1108.

Mootha, V. V., Hansen, B., Rong, Z., Mammen, P. P., Zhou, Z., Xing, C., et al. (2017). Fuchs' Endothelial Corneal Dystrophy and RNA Foci in Patients With Myotonic Dystrophy. *Invest. Ophthalmol. Vis. Sci.* 58, 4579–4585. doi: 10.1167/iovs.17-22350.

Mudge, J., Miller, N. A., Khrebtukova, I., Lindquist, I. E., May, G. D., Huntley, J. J., et al. (2008). Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PloS One* 3, e3625. doi: 10.1371/journal.pone.0003625.

Murre, C. (2019). Helix–loop–helix proteins and the advent of cellular diversity: 30 years of discovery. *Genes Dev.* 33, 6–25. doi: 10.1101/gad.320663.118.

Nelson, D. L., Orr, H. T., and Warren, S. T. (2013). The Unstable Repeats - Three Evolving Faces of Neurological Disease. *Neuron* 77, 825–843. doi: 10.1016/j.neuron.2013.02.022.

Nurm, K., Sepp, M., Castany-Pladevall, C., Creus-Muncunill, J., Tuvikene, J., Sirp, A., et al. (2021). Isoform-Specific Reduction of the Basic Helix-Loop-Helix Transcription Factor TCF4 Levels in Huntington's Disease. *eNeuro* 8, ENEURO.0197-21.2021. doi: 10.1523/ENEURO.0197-21.2021.

Oh, T.-I., Lee, M., Lee, Y.-M., Kim, G.-H., Lee, D., You, J. S., et al. (2021). PGC1α Loss Promotes Lung Cancer Metastasis through Epithelial-Mesenchymal Transition. *Cancers* 13, 1772. doi: 10.3390/cancers13081772.

Okumura, N., Hayashi, R., Nakano, M., Yoshii, K., Tashiro, K., Sato, T., et al. (2019). Effect of Trinucleotide Repeat Expansion on the Expression of TCF4 mRNA in Fuchs' Endothelial Corneal Dystrophy. *Invest. Ophthalmol. Vis. Sci.* 60, 779–786. doi: 10.1167/iovs.18-25760.

Ołdak, M., Ruszkowska, E., Udziela, M., Oziębło, D., Bińczyk, E., Ścieżyńska, A., et al. (2015). Fuchs Endothelial Corneal Dystrophy: Strong Association with rs613872 Not Paralleled by Changes in Corneal Endothelial TCF4 mRNA Level. *BioMed Res. Int.* 2015. doi: 10.1155/2015/640234.

Ong Tone, S., Kocaba, V., Böhm, M., Wylegala, A., White, T. L., and Jurkunas, U. V. (2020). Fuchs endothelial corneal dystrophy: The vicious cycle of Fuchs pathogenesis. *Prog. Retin. Eye Res.*, 100863. doi: 10.1016/j.preteyeres.2020.100863.

Page, S. C., Hamersky, G. R., Gallo, R. A., Rannals, M. D., Calcaterra, N. E., Campbell, M. N., et al. (2018). The schizophrenia and autism associated gene, Transcription Factor 4 (TCF4) regulates the columnar distribution of layer 2/3 prefrontal pyramidal neurons in an activity-dependent manner. *Mol. Psychiatry* 23, 304–315. doi: 10.1038/mp.2017.37.

Papes, F., Camargo, A. P., de Souza, J. S., Carvalho, V. M. A., Szeto, R. A., LaMontagne, E., et al. (2022). Transcription Factor 4 loss-of-function is associated with deficits in progenitor proliferation and cortical neuron content. *Nat. Commun.* 13, 2387. doi: 10.1038/s41467-022-29942-w.

Paulson, H. (2018). Repeat expansion diseases. *Handb. Clin. Neurol.* 147, 105–123. doi: 10.1016/B978-0-444-63233-3.00009-9.

Pernía, O., Sastre-Perona, A., Rodriguez-Antolín, C., García-Guede, A., Palomares-Bralo, M., Rosas, R., et al. (2020). A Novel Role for the Tumor Suppressor Gene ITF2 in Tumorigenesis and Chemotherapy Response. *Cancers* 12, 786. doi: 10.3390/cancers12040786.

Pfurr, S., Chu, Y.-H., Bohrer, C., Greulich, F., Beattie, R., Mammadzada, K., et al. (2017). The E2A splice variant E47 regulates the differentiation of projection neurons via p57(KIP2) during cortical development. *Development* 144, 3917–3931. doi: 10.1242/dev.145698.

Phan, B. N., Bohlen, J. F., Davis, B. A., Ye, Z., Chen, H.-Y., Mayfield, B., et al. (2020). A myelin-related transcriptomic profile is shared by Pitt–Hopkins syndrome models and human autism spectrum disorder. *Nat. Neurosci.* 23, 375–385. doi: 10.1038/s41593-019-0578-x.

Pontual, L. de, Mathieu, Y., Golzio, C., Rio, M., Malan, V., Boddaert, N., et al. (2009). Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. *Hum. Mutat.* 30, 669–676. doi: 10.1002/humu.20935.

Powell, L. M., and Jarman, A. P. (2008). Context dependence of proneural bHLH proteins. *Curr. Opin. Genet. Dev.* 18, 411–417. doi: 10.1016/j.gde.2008.07.012.

Pscherer, A., Dörflinger, U., Kirfel, J., Gawlas, K., Rüschoff, J., Buettner, R., et al. (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 J.* 15, 6680–6690.

Quan, X.-J., Yuan, L., Tiberi, L., Claeys, A., De Geest, N., Yan, J., et al. (2016). Post-translational Control of the Temporal Dynamics of Transcription Factor Activity Regulates Neurogenesis. *Cell* 164, 460–475. doi: 10.1016/j.cell.2015.12.048.

Quednow, B. B., Brinkmeyer, J., Mobascher, A., Nothnagel, M., Musso, F., Gründer, G., et al. (2012). Schizophrenia risk polymorphisms in the TCF4 gene interact with smoking in the modulation of auditory sensory gating. *Proc. Natl. Acad. Sci. U. S. A.* 109, 6271–6276. doi: 10.1073/pnas.1118051109.

Quednow, B. B., Brzózka, M. M., and Rossner, M. J. (2014). Transcription factor 4 (TCF4) and schizophrenia: integrating the animal and the human perspective. *Cell. Mol. Life Sci.* 71, 2815–2835. doi: 10.1007/s00018-013-1553-4.

Quevedo, M., Meert, L., Dekker, M. R., Dekkers, D. H. W., Brandsma, J. H., van den Berg, D. L. C., et al. (2019). Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nat. Commun.* 10, 2669. doi: 10.1038/s41467-019-10502-8.

Rannals, M. D., Hamersky, G. R., Page, S. C., Campbell, M. N., Briley, A., Gallo, R. A., et al. (2016). Psychiatric Risk Gene Transcription Factor 4 Regulates Intrinsic Excitability of Prefrontal Neurons via Repression of SCN10a and KCNQ1. *Neuron* 90, 43–55. doi: 10.1016/j.neuron.2016.02.021.

Ravanpay, A. C., and Olson, J. M. (2008). E protein dosage influences brain development more than family member identity. *J. Neurosci. Res.* 86, 1472–1481. doi: 10.1002/jnr.21615.

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., et al. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* 43, 969–976. doi: 10.1038/ng.940.

Rosato, M., Stringer, S., Gebuis, T., Paliukhovich, I., Li, K. W., Posthuma, D., et al. (2021). Combined cellomics and proteomics analysis reveals shared neuronal morphology and molecular pathway phenotypes for multiple schizophrenia risk genes. *Mol. Psychiatry* 26, 784–799. doi: 10.1038/s41380-019-0436-y.

Ruzicka, W. B., Mohammadi, S., Davila-Velderrain, J., Subburaju, S., Tso, D. R., Hourihan, M., et al. (2020). Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. *medRxiv*, 2020.11.06.20225342. doi: 10.1101/2020.11.06.20225342.

Saarikettu, J., Sveshnikova, N., and Grundström, T. (2004). Calcium/calmodulin inhibition of transcriptional activity of E-proteins by prevention of their binding to DNA. *J. Biol. Chem.* 279, 41004–41011. doi: 10.1074/jbc.M408120200.

Salcedo-Arellano, M. J., Dufour, B., McLennan, Y., Martinez-Cerdeno, V., and Hagerman, R. (2020). Fragile X Syndrome and associated disorders: clinical aspects and pathology. *Neurobiol. Dis.* 136, 104740. doi: 10.1016/j.nbd.2020.104740.

Sallee, M. D., and Greenwald, I. (2015). Dimerization-driven degradation of C. elegans and human E proteins. *Genes Dev.* 29, 1356–1361. doi: 10.1101/gad.261917.115.

Sarkar, D., Shariq, M., Dwivedi, D., Krishnan, N., Naumann, R., Bhalla, U. S., et al. (2021). Adult brain neurons require continual expression of the schizophrenia-risk gene Tcf4 for structural and functional integrity. *Transl. Psychiatry* 11, 1–11. doi: 10.1038/s41398-021-01618-x.

Schoof, M., Hellwig, M., Harrison, L., Holdhof, D., Lauffer, M. C., Niesen, J., et al. (2020). The basic helix-loop-helix transcription factor TCF4 impacts brain architecture as well as neuronal morphology and differentiation. *Eur. J. Neurosci.* 51, 2219–2235. doi: 10.1111/ejn.14674.

Sepp, M., Kannike, K., Eesmaa, A., Urb, M., and 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. doi: 10.1371/journal.pone.0022138.

Sepp, M., Pruunsild, P., and Timmusk, T. (2012). Pitt–Hopkins syndrome-associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant-negative effects. *Hum. Mol. Genet.* 21, 2873–2888. doi: 10.1093/hmg/dds112.

Sepp, M., Vihma, H., Nurm, K., Urb, M., Page, S. C., Roots, K., et al. (2017). The Intellectual Disability and Schizophrenia Associated Transcription Factor TCF4 Is Regulated by Neuronal Activity and Protein Kinase A. *J. Neurosci.* 37, 10516–10527. doi: 10.1523/JNEUROSCI.1151-17.2017.

Shariq, M., Sahasrabuddhe, V., Krishna, S., Radha, S., Nruthyathi, null, Bellampalli, R., et al. (2021). Adult neural stem cells have latent inflammatory potential that is kept suppressed by Tcf4 to facilitate adult neurogenesis. *Sci. Adv.* 7, eabf5606. doi: 10.1126/sciadv.abf5606.

Shin, H., Choi, H., So, D., Kim, Y., Cho, K., Chung, H., et al. (2014). ITF2 Prevents Activation of the β-Catenin–TCF4 Complex in Colon Cancer Cells and Levels Decrease With Tumor Progression. *Gastroenterology* 147, 430-442.e8. doi: 10.1053/j.gastro.2014.04.047.

Sobrado, V. R., Moreno-Bueno, G., Cubillo, E., Holt, L. J., Nieto, M. A., Portillo, F., et al. (2009). The class I bHLH factors E2-2A and E2-2B regulate EMT. *J. Cell Sci.* 122, 1014–1024. doi: 10.1242/jcs.028241.

Soliman, A. Z., Xing, C., Radwan, S. H., Gong, X., and Mootha, V. V. (2015). Correlation of Severity of Fuchs Endothelial Corneal Dystrophy With Triplet Repeat Expansion in TCF4. *JAMA Ophthalmol.* 133, 1386–1391. doi: 10.1001/jamaophthalmol.2015.3430.

Soosaar, A., Chiaramello, A., Zuber, M. X., and 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. *Mol. Brain Res.* 25, 176–180. doi: 10.1016/0169-328X(94)90297-6.

Srivastava, S., Desai, S., Cohen, J., Smith-Hicks, C., Barañano, K., Fatemi, A., et al. (2018). Monogenic disorders that mimic the phenotype of Rett syndrome. *neurogenetics* 19, 41–47. doi: 10.1007/s10048-017-0535-3.

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., et al. (2009). Common variants conferring risk of schizophrenia. *Nature* 460, 744–747. doi: 10.1038/nature08186.

Steinberg, S., de Jong, S., Andreassen, O. A., Werge, T., Børglum, A. D., Mors, O., et al. (2011). Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum. Mol. Genet.* 20, 4076–4081. doi: 10.1093/hmg/ddr325.

Stephan, M., Schoeller, J., Raabe, F. J., Schmitt, A., Hasan, A., Falkai, P., et al. (2022). Spironolactone alleviates schizophrenia-related reversal learning in Tcf4 transgenic mice subjected to social defeat. *Schizophrenia* 8, 1–13. doi: 10.1038/s41537-022-00290-4.

Stessman, H. A. F., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., et al. (2017). Targeted sequencing identifies 91 neurodevelopmental disorder risk genes with autism and developmental disability biases. *Nat. Genet.* 49, 515–526. doi: 10.1038/ng.3792.

Sundin, O. H., Broman, K. W., Chang, H. H., Vito, E. C. L., Stark, W. J., and Gottsch, J. D. (2006). A Common Locus for Late-Onset Fuchs Corneal Dystrophy Maps to 18q21.2-q21.32. *Invest. Ophthalmol. Vis. Sci.* 47, 3919–3926. doi: 10.1167/iovs.05-1619.

Sweatt, J. D. (2013). Pitt-Hopkins Syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. *Exp. Mol. Med.* 45, e21. doi: 10.1038/emm.2013.32.

Tamberg, L., Sepp, M., Timmusk, T., and Palgi, M. (2015). Introducing Pitt-Hopkins syndrome-associated mutations of TCF4 to Drosophila daughterless. *Biol. Open* 4, 1762–1771. doi: 10.1242/bio.014696.

Teixeira, J. R., Szeto, R. A., Carvalho, V. M. A., Muotri, A. R., and Papes, F. (2021). Transcription factor 4 and its association with psychiatric disorders. *Transl. Psychiatry* 11, 19. doi: 10.1038/s41398-020-01138-0.

Thaxton, C., Kloth, A. D., Clark, E. P., Moy, S. S., Chitwood, R. A., and Philpot, B. D. (2018). Common Pathophysiology in Multiple Mouse Models of Pitt–Hopkins Syndrome. *J. Neurosci.* 38, 918–936. doi: 10.1523/JNEUROSCI.1305-17.2017.

Tuvikene, J., Esvald, E.-E., Rähni, A., Uustalu, K., Zhuravskaya, A., Avarlaid, A., et al. (2021). Intronic enhancer region governs transcript-specific Bdnf expression in rodent neurons. *eLife* 10, e65161. doi: 10.7554/eLife.65161.

Wang, Y., Lu, Z., Zhang, Y., Cai, Y., Yun, D., Tang, T., et al. (2020). Transcription Factor 4 Safeguards Hippocampal Dentate Gyrus Development by Regulating Neural Progenitor Migration. *Cereb. Cortex* 30, 3102–3115. doi: 10.1093/cercor/bhz297.

Wang, Z., Zhao, T., Zhang, S., Wang, J., Chen, Y., Zhao, H., et al. (2021). The Wnt signaling pathway in tumorigenesis, pharmacological targets, and drug development for cancer therapy. *Biomark. Res.* 9, 68. doi: 10.1186/s40364-021-00323-7.

Wedel, M., Fröb, F., Elsesser, O., Wittmann, M.-T., Lie, D. C., Reis, A., et al. (2020). Transcription factor Tcf4 is the preferred heterodimerization partner for Olig2 in oligodendrocytes and required for differentiation. *Nucleic Acids Res.* 48, 4839–4857. doi: 10.1093/nar/gkaa218.

Whalen, S., Héron, D., Gaillon, T., Moldovan, O., Rossi, M., Devillard, F., et al. (2012). Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: clinical score and further delineation of the TCF4 mutational spectrum. *Hum. Mutat.* 33, 64–72. doi: 10.1002/humu.21639.

Wieben, E. D., Aleff, R. A., Tosakulwong, N., Butz, M. L., Highsmith, W. E., Edwards, A. O., et al. (2012). A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. *PloS One* 7, e49083. doi: 10.1371/journal.pone.0049083.

Wirgenes, K. V., Sønderby, I. E., Haukvik, U. K., Mattingsdal, M., Tesli, M., Athanasiu, L., et al. (2012). TCF4 sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders. *Transl. Psychiatry* 2, e112. doi: 10.1038/tp.2012.39.

Wittmann, M.-T., Katada, S., Sock, E., Kirchner, P., Ekici, A. B., Wegner, M., et al. (2021). scRNA sequencing uncovers a TCF4-dependent transcription factor network regulating commissure development in mouse. *Dev. Camb. Engl.* 148, dev196022. doi: 10.1242/dev.196022.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* 50, 668. doi: 10.1038/s41588-018-0090-3.

Wright, C., Turner, J., Calhoun, V., and Perrone Bizzozero, N. (2013). Potential Impact of miR-137 and Its Targets in Schizophrenia. *Front. Genet.* 4. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2013.00058 [Accessed January 2, 2023].

Xia, H., Jahr, F. M., Kim, N.-K., Xie, L., Shabalin, A. A., Bryois, J., et al. (2018). Building a schizophrenia genetic network: transcription factor 4 regulates genes involved in neuronal development and schizophrenia risk. *Hum. Mol. Genet.* 27, 3246–3256. doi: 10.1093/hmg/ddy222.

Yang, J., Horton, J. R., Li, J., Huang, Y., Zhang, X., Blumenthal, R. M., et al. (2019). Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine. *Nucleic Acids Res.* 47, 8375–8387. doi: 10.1093/nar/gkz381.

Yoon, S. O., and Chikaraishi, D. M. (1994). Isolation of two E-box binding factors that interact with the rat tyrosine hydroxylase enhancer. *J. Biol. Chem.* 269, 18453–18462.

Zhang, J., Kalkum, M., Yamamura, S., Chait, B. T., and Roeder, R. G. (2004). E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science* 305, 1286–1289. doi: 10.1126/science.1097937.

Zhuang, Y., Barndt, R. J., Pan, L., Kelley, R., and Dai, M. (1998). Functional Replacement of the Mouse E2A Gene with a Human HEB cDNA. *Mol. Cell. Biol.* 18, 3340–3349.

Zhuang, Y., Cheng, P., and Weintraub, H. (1996). B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. *Mol. Cell. Biol.* 16, 2898–2905.

Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. *Cell* 79, 875–884. doi: 10.1016/0092-8674(94)90076-0.

Zollino, M., Zweier, C., Van Balkom, I. D., Sweetser, D. A., Alaimo, J., Bijlsma, E. K., et al. (2019). Diagnosis and management in Pitt-Hopkins syndrome: First international consensus statement. *Clin. Genet.* 95, 462–478. doi: 10.1111/cge.13506.

Zweier, C., de Jong, E. K., Zweier, M., Orrico, A., Ousager, L. B., Collins, A. L., et al. (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. *Am. J. Hum. Genet.* 85, 655–666. doi: 10.1016/j.ajhg.2009.10.004.

Zweier, C., Peippo, M. M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., et al. (2007). Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am. J. Hum. Genet.* 80, 994–1001. doi: 10.1086/515583.

Zweier, C., Sticht, H., Bijlsma, E. K., Clayton-Smith, J., Boonen, S. E., Fryer, A., et al. (2008). Further delineation of Pitt-Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. *J. Med. Genet.* 45, 738–744. doi: 10.1136/jmg.2008.060129.

Acknowledgements

First of all, I am forever thankful to my supervisor professor Tõnis Timmusk for granting me the ability to explore and work in the field of molecular neurobiology. You did everything possible to help me succeed – pushed me when needed and always supported me. There are not many people in this world as kind and supportive as you.

My greatest thanks go to all the members of neurolab including Epp Väli, Indrek Koppel, Richard Tamme, Mari Palgi, Olga Jasnovidova, Florencia Cabrera Cabrera, and all the previous members Kaisa Roots, Kristian Costa Dos Santos Leite, Kaur Jaanson, Kaja Nurm and Hanna Vihma. Special thanks goes to Mari Sepp who guided me a lot during the writing phases of our co-authored manuscripts.

Jürgen Tuvikene and Eli-Eelika Esvald, thank you so much for all the great discussions and help with my experiments. Guidance received from you is priceless.

Laura Tamberg, Annela Avarlaid, Anastassia Šubina and Carl Sander Kiir, thank you so much for always being there for me. You always had my back – helped me through my toughest times and were there to share my greatest moments. I really hope that our bond will never die.

I would also like to thank members of the institute Tiit Lukk, Pirjo Spuul, Vello Tõugu and Peep Palumaa for all the great discussions and help that I have received from you.

I am deeply thankful to my parents, my sister and my brother for always supporting me and giving me the chance to fully focus on my studies. Without you this journey would had been much more difficult.

I thank all the funding agencies that supported this work: Estonian Research Council (grants MJD341, IUT19-18 and PRG805), National R&D program “Health” (grant AR12098), Estonian Academy of Sciences, the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) and H2020-MSCA-RISE-2016 (EU734791), the Pitt Hopkins Research Foundation (grants 8 and 21), the Million Dollar Bike Ride Pilot Grant Program for Rare Disease Research at UPenn Orphan Disease Center (grants MDBR-16-122-PHP and MDBR-17-127-Pitt Hopkins). I thank the “TUT Institutional Development Program for 2016−2022” Graduate School in Clinical Medicine, which received funding from the European Regional Development Fund under program ASTRA 2014-2020.4.01.16-0032 in Estonia.

Abstract

Molecular Characterization of Basic Helix-Loop-Helix Transcription Factor TCF4: From Expression to Function

Transcription factor 4 (TCF4) is a member of the bHLH family of transcription factors that mediates its function through homo- and heterodimerization with interaction partners such as ASCL1, NEUROD and ID proteins. After dimerization, TCF4 binds to its target sequence (CANNTG) also known as the Ephrussi box. Overall TCF4 plays a very important role in neurogenesis. More specifically, TCF4 is involved in the development of B- and T-cells, maturation of plasmacytoid dendritic cells, epithelial-mesenchymal transition and differentiation and migration of neurons.

Due to the complex gene structure transcription from the TCF4 gene can result in at least 18 N-terminally distinct protein isoforms. In rodent, 7 N-terminally different TCF4 protein isoforms have been described. Earlier studies of TCF4 expression have shown that total TCF4 mRNA and protein expression is highest just around birth. As none of the previous studies of TCF4 expression have focused on the distinct TCF4 transcripts, we studied the expression pattern of the many *TCF4* transcripts and protein isoforms throughout development. For that we did long read mRNA sequencing, analysed previously published RNA-seq data and ran Western blot analysis of neural and nonneural tissue lysates. Our results confirmed that *Tcf4* mRNA and protein expression is highest just around birth. The main *TCF4* isoforms expressed in rodent and human are TCF4-B, -C, -D, -A and -I. While TCF4-A encoding transcripts account for the majority of TCF4 transcripts, TCF4-B, -C and -D encoding transcripts are also highly expressed in rodent and human tissues. All these isoforms were also detected in Western blot analysis. Long and short TCF4 protein isoforms are expressed in all the studied rodent neural and nonneural tissues, however, medium protein isoforms are only expressed in the brain.

The CTG trinucleotide repeat expansion in the third intron of *TCF4* has been tied to the development of late-onset Fuchs endothelial corneal dystrophy (FECD). A *TCF4* allele with >50 CTG repeats confers an increased risk of developing FECD. FECD can be described by the formation of guttae in the corneal endothelium which starts spreading from the middle of the eye to the side and if left untreated results in loss of vision. Studies suggest that FECD affects about 5% of middle-aged people. The only treatment for FECD is replacement of the corneal endothelium with donor tissue.

Human *TCF4* has a total of 41 exons with 21 being alternative self-exclusive 5' exons. The FECD associated CTG repeat region is located in intron three, just before exon 4a. One of the goals of this thesis was to study, how the CTG repeat expansion can mediate the development of FECD. First, we confirmed that the CTG repeat region is in close proximity to several transcription start sites which initiate the expression of *TCF4* exon 4a, 4b and 4c encoding transcripts. Our results showed that expansion of the CTG repeat >50 significantly reduces nearby promoter activity. Next, we studied previously published RNA-sequencing data from cornea of FECD patients to describe the expression of *TCF4*. For that, we developed a method which allows to quantify the expression of not just total TCF4, but also the expression of the many TCF4 transcripts from short-read RNA-seq data. Collectively, our results displayed that expression of *TCF4* transcripts linked to promoters near the repeat region declined while expression of certain transcripts beginning further downstream of the CTG repeat increased.

Missense mutations and variations in TCF4 have been associated with the generation of several neurocognitive disorders such as schizophrenia, intellectual disability, autism, depression and Rett-like syndrome. In addition, mutations in *TCF4* which include large deletions, translocations and also frameshift, nonsense and missense mutations cause Pitt-Hopkins syndrome (PTHS), a rare but severe autism spectrum disorder described by developmental delay and mental retardation. The third goal of this thesis was to study the effects of 12 previously described disease-related missense mutations and variations on the functionality of *TCF4*. Our results showed that schizophrenia, intellectual disability and Rett-like syndrome related mutations and variations had very little or no effect on the functionality of TCF4. However, all the PTHS associated mutant TCF4 proteins displayed impaired DNA binding activity. In addition, PTHS mutations can either increase or decrease transcription activation of TCF4 depending on the cell type or specific TCF4 isoform.

Collectively, our results show that:

- *TCF4* expression is highest around birth in rodents and humans;
- *TCF4* expression is highest in neural tissues and much lower in nonneural tissues;
- The main TCF4 protein isoforms expressed in neural tissues are TCF4-B, -C, -D, -A and -I;
- The CTG repeat expansion in third intron of *TCF4* gene reduces the activity of nearby *TCF4* promoters;
- Pitt-Hopkins syndrome-related missense mutations in the basic helix-loop-helix region of TCF4 reduce the abilty of TCF4 to bind DNA and also alter transcriptional activation depending on cell type.

Lühikokkuvõte

Aluselise heeliks-ling heeliks transkriptsiooniteguri TCF4 ekspressiooni ja funktsiooni kirjeldamine

Transkriptsioonifaktor 4 (TCF4) on bHLH transkriptsioonifaktorite perekonna liige, mis vahendab oma funktsiooni homo- ja heterodimerisatsiooni kaudu interaktsioonipartneritega, nagu ASCL1, NEUROD ja ID valgud. Pärast dimeriseerumist seondub TCF4 oma sihtjärjestusega (CANNTG), mida tuntakse ka kui *Ephrussi box*. Üldiselt mängib TCF4 väga olulist rolli neurogeneesis. Täpsemalt osaleb TCF4 B- ja T-rakkude arengus, plasmatsütoidsete dendriitrakkude küpsemises, epiteeli-mesenhümaalses transformatsioonis ning neuronite diferentseerumises ja migratsioonis.

Keerulise geenistruktuuri tõttu võib inimese *TCF4* geeni transkriptsioon anda tulemuseks vähemalt 18 N-terminaalselt erinevat valgu isovormi. Närilistel on kirjeldatud 7 N-terminaalselt erinevat TCF4 valgu isovormi. Varasemalt on näidatud, et TCF4 mRNA ja valgu ekspressioon on kõige suurem just sünni paiku. Siiski, keegi pole täpsemalt kirjeldanud, kuidas on erinevad TCF4 geeni transkriptid ja valgu isovormid ekspresseeritud ning kuidas muutub nende ekspressiooni dünaamika läbi arengu. Sellest tulenevalt kasutasime uudset pikkade transkriptide RNA-sekveneerimise meetodit, analüüsisime varasemalt avaldatud RNA-sekveneerimise andmeid ning valmistasime valgu lüsaadid erinevatest närilise ja inimese kudedest, et uurida TCF4 mRNA ning valgu ekspressiooni läbi arengu pannes rõhku just erinevate transkrpitide ja isovormide ekspressioonile. Meie tulemused kinnitasid, et TCF4 mRNA ja valgu ekspressioon on kõrgeim just sündimise ajal. Peamised närilistel ja inimestel ekspresseeritud TCF4 isovormid on TCF4-B, -C, -D, -A ja -I. Kuigi TCF4-A kodeerivad transkriptid moodustavad suurema osa TCF4 transkriptidest, on TCF4-B, -C, -D ja -I kodeerivad transkriptid samuti tugevalt ekspresseeritud näriliste ja inimese kudedes. Kõik need TCF4 isovormid tuvastati ka Western blot analüüsis. Kui pikki ja keskmiseid TCF4 valgu isovorme oli näha kõikides uuritud kudedes, siis, keskmise suurusega valgu isovorme oli näha vaid ajukoores ning hippokampuses.

Kõrvalekaldeid *TCF4* geenis on seosatatud mitmete haigustega millest üks on hiline Fuchsi endoteliaalne sarvkesta düstroofia (FECD). FECD tekke riskiks peetakse *TCF4* geenis paikneva CTG kolmenukleotiidiline kordusjärjestuse arvu suurenemist >50 korduse. FECD tekib pea 5%-l keskealistest inimestest ning selle korral halveneb patsiendi nägemine silmale moodustunud kaelaadse kogumi tõttu, mis hakkab levima silma keskosast äärtesse. Ainus FECD ravimeetod on sarvkesta endoteeli asendamine doonorkoega.

Inimese *TCF4* geen koosneb 41-st eksonist, millest 21 on alternatiivsed üksteist välistavad 5’ eksonid. FECD-ga seotud CTG korduspiirkond asub kolmandas intronis, vahetult enne eksonit 4a. Üks käesoleva lõputöö eesmärke oli uurida, kuidas CTG korduse suurenemine võib vahendada FECD tekkimist. Esimesena kinnitasime, et CTG korduspiirkond on mitme transkriptsiooni alguspunkti vahetus läheduses, mis aktiveerivad TCF4 eksoneid 4a, 4b ja 4c kodeerivate transkriptide ekspressiooni. Meie tulemused näitasid, et CTG korduse laienemine >50 vähendab oluliselt läheduses asuva promootori aktiivsust. Sellest tulenevalt, otsustasime uurida kas varasemalt publitseeritud FECD patsientide RNA sekveneerimise tulemustest oleks võimalik uurida, kuidas on muutunud erinevate *TCF4* mRNAde tasemed. Selliseks analüüsiks töötasime

välja meetodi, mis kvantiseerib RNA-sekveneerimise tulemusi üle splaiss-saitide. Kokkuvõtvalt näitasid meie tulemused, et CTG korduspiirkonna lähedal olevate promootorite kontrolli all olevate *TCF4* mRNAde transkriptsioon vähenes ja teatud *TCF4* transkriptide ekspressioon, mille promootorid paiknevad CTG kordusest kaugel allavoolu, suurenes.

Mutatsioone ja -variatsioone *TCF4* geenis on seostatud mitmete neurokognitiivsete häirete tekkega nagu skisofreenia, vaimne alaareng, autism, depressioon ja Rett-i sarnane sündroom. Lisaks, on teada et *TCF4* haplopuudulikkus põhjustab harvaesinevat aga tõsist haigust Pitt-Hopkinsi sündroom (PTHS). PTHS-i korral on üks *TCF4* geeni alleelidest muteerunud ja sealt ei ekspresseeru funktsionaalset valku.

Selle lõputöö kolmas eesmärk oli uurida varem kirjeldatud haigusega seotud asendusmutatsioonide ja -variatsioonide mõju TCF4 funktsionaalsusele. Meie tulemused näitasid, et skisofreenia, intellektipuude ja Retti-sarnase sündroomiga seotud mutatsioonid ja variatsioonid mõjutasid TCF4 funktsionaalsust väga vähe või üldse mitte. Siiski, kõik PTHS-ga seotud mutatsioonid, mis paiknesid heeliks-ling-heeliks domeenis, vähendasid TCF4 võimet seonduda DNA-le. Lisaks võivad PTHS-i mutatsioonid sõltuvalt rakutüübist või spetsiifilisest TCF4 isovormist kas suurendada või vähendada TCF4 transkriptsiooni aktivatsiooni.

Kokkuvõtvalt näitasid meie tulemused:

- TCF4 ekspressioon on närilistel ja inimestel kõrgeim sünni paiku;
- TCF4 ekspressioon on kõrgeim närvisüsteemis ja palju madalam mitteneuraalsetes kudedes;
- Peamised närvisüsteemis ja mitteneuraalsetes kudedes ekspresseeritud TCF4 isovormid on TCF4-B, -C, -D, -A ja -I;
- Erinevate TCF4 isovormide suhtelised tasemed on läbi arengu stabiilsed. Suuremaid muutuseid, kus ühe TCF4 isovormi ekspressioon väheneb ja teise TCF4 isovormi ekspressioon suureneb ei esine.
- CTG korduse laienemine TCF4 kolmandas intronis vähendab lähedalasuvate TCF4 promootorite aktiivsust;
- Pitt-Hopkinsi sündroomiga seotud asendusmutatsioonid TCF4 heeliks-ling-heeliks piirkonnas vähendavad TCF4 võimet siduda DNA-d ja muudavad ka TCF4 vahendatud transkriptsiooni aktivatsiooni sõltuvalt rakutüübist.

Appendix

Publication I

Sirp, A.*, Shubina, A.*, Tuvikene, J., Tamberg, L., Kiir, C.S., Kranich, L., Timmusk, T.
Expression of alternative transcription factor 4 mRNAs and protein isoforms in the developing and adult rodent and human tissues.
Front. Mol. Neurosci., 15. 2022 Nov. DOI: 10.3389/fnmol.2022.1033224.

Publication II

Sirp, A.*, Leite, K.*, Tuvikene, J.*, Nurm, K., Sepp, M., Timmusk, T.
The Fuchs corneal dystrophy-associated CTG repeat expansion in the TCF4 gene affects transcription from its alternative promoters.
Sci. Rep., 2020 Oct; 10 (1), #18424. DOI: 10.1038/s41598-020-75437-3

Publication III

Sirp, A.*, Roots, K.*, Nurm, K., Tuvikene, J., Sepp, M., Timmusk, T.
Functional consequences of TCF4 missense substitutions associated with Pitt-Hopkins syndrome, mild intellectual disability, and schizophrenia.
J. Biol. Chem. 2021 Dec; 297, 101381. doi: 10.1016/j.jbc.2021.101381.

Curriculum vitae

Personal data

Name	Alex Sirp
Date of birth	04.05.1992
Place of birth	Tallinn, Estonia

Contact data

E-mail	alex.sirp@taltech.ee

Education

2017–2023	Tallinn University of Technology, Gene Technology (PhD)
2014–2016	Tallinn University of Technology, Gene Technology (MSc)
2011–2014	Tallinn University of Technology, Gene Technology (BSc)
2008–2011	Estonian Business School High school

Professional employment

2017–	Tallinn University of Technology, School of Science, Department of Chemistry and Biotechnology, PhD student – junior researcher (0,10)

Managerial and administrative work

2015–2016	Tallinn University of Technology, member of the Student Union

Academic degrees

Alex Sirp, Master's Degree, 2016, (sup) Jürgen Tuvikene; Tõnis Timmusk, Screening of Transcription Factors Regulating BDNF Positive Feedback Loop in Cortical Neurons, Tallinn University of Technology Faculty of Science, Department of Gene Technology, Chair of Molecular Biology.

Alex Sirp, Bachelor's Degree, 2014, (sup) Teet Velling, The Role of Epidermal Growth Factor Receptor, Integrin-Type Collagen Receptors and Filamin A in Etoposide- and Staurosporine-Induced Cell Death, Tallinn University of Technology Faculty of Science, Department of Gene Technology, Chair of Molecular Biology.

Scholarships and awards

2022	Erkki Truve Doctoral Study Scholarship, Tallinn University of Technology Development Fund.

Courses and conferences

October 2022	Participation in EstSHG 22 annual conference, Pärnu, Estonia
November 2021	Participation in EstSHG 21 annual conference, Haapsalu, Estonia

February 2020	Participation in winter school for PhD students "Writing Process Reengineering", Rakvere, Estonia
November 2019	Participation in EstSHG 20 annual conference, Pärnu, Estonia
June 2019	Poster presentation at "Nordic Neuroscience 3" conference, "Regulation of the basic helix-loop-helix transcription factor TCF4 activity in neuronal cells", Helsinki, Finland
November 2018	Participation in EstSHG 19 annual conference, Viljandi, Estonia
November 2018	Poster presentation at "Neuroscience 2018" conference, "Regulation of the basic helix-loop-helix transcription factor TCF4 activity in neuronal cells", San Diego, USA
August 2018	Participation in "Frontiers in Delivery of Therapeutics" conference, Tartu, Estonia
May 2018	Participation in science day for PhD students, poster presentation "Regulation of the basic helix-loop-helix transcription factor TCF4 activity by disease-related missense variations", Vehendi, Estonia
February 2018	Participation in winter school for PhD students "Life after PhD", Pärnu, Estonia
November 2017	Participation in EstSHG 18 annual conference, Rakvere, Estonia

Supervised dissertations

Carl Sander Kiir, Master's Degree, 2021, (sup) Alex Sirp; Mari Palgi, Neuronal activity-dependent regulation of daughterless expression in the central nervous system of Drosophila melanogaster, Tallinn University of Technology School of Science, Department of Chemistry and Biotechnology.

Britt-Anett Kristelštein, Bachelor's Degree, 2021, (sup) Alex Sirp; Tõnis Timmusk, Tagging of transcription factor TCF-4 in mouse Neuro-2a cells and mouse embryonic stem cells by CRISPR-Cas9 method, Tallinn University of Technology Faculty of Science, Department of Gene Technology, Chair of Molecular Biology.

Anastassia Šubina, magistrikraad, 2020, (sup) Alex Sirp; Tõnis Timmusk, Expression of transcription factor TCF4 in the developing and adult rodent brain, Tallinn University of Technology Faculty of Science, Department of Gene Technology, Chair of Molecular Biology.

Laura Kranich, bakalaureusekraad, 2020, (sup) Alex Sirp; Tõnis Timmusk, Silencing of Transcription Factor TCF4 in Mouse Neuro-2a and Human SH-SY5Y Cells Using the Crispr-Cas9 Method, Tallinn University of Technology Faculty of Science, Department of Gene Technology, Chair of Molecular Biology.

Paul Mark Tammiste, 12th grade research paper, 2019, (sup) Alex Sirp; Martin Saar, Subcellular localization of the basic helix-loop-helix transcription factor TCF4 in Huntington's disease.

Publications

Sirp A,* Shubina A,* Tuvikene J, Tamberg L, Kiir CS, Kranich L, Timmusk T (2022) Expression of alternative transcription factor 4 mRNAs and protein isoforms in the developing and adult rodent and human tissues. Front. Mol. Neurosci., 15. 2022 Nov. DOI: 10.3389/fnmol.2022.1033224.

Sirp, A.,* Roots, K.,* Nurm, K., Tuvikene, J., Sepp, M., and Timmusk, T. (2021). Functional consequences of TCF4 missense substitutions associated with Pitt-Hopkins syndrome, mild intellectual disability, and schizophrenia. *J. Biol. Chem.* 297, 101381. doi: 10.1016/j.jbc.2021.101381.

Nurm, K.; Sepp, M.; Castany-Pladevall, C.; Creus-Muncunill, J.; Tuvikene, J.; **Sirp, A**; Vihma, H.; Blake, DJ..; Perez-Navarro, E.; Timmusk, T. (2021). Isoform-specific reduction of the basic helix-loop-helix transcription factor TCF4 levels in Huntington's disease. eNeuro, ENEURO.0197-21.2021. DOI: 10.1523/eneuro.0197-21.2021.

Sirp, A.*; Leite, K.*; Tuvikene, J.*; Nurm, K.; Sepp, M.; Timmusk, T. (2020). The Fuchs corneal dystrophy-associated CTG repeat expansion in the TCF4 gene affects transcription from its alternative promoters. Sci. Rep., 10 (1), #18424. DOI: 10.1038/s41598-020-75437-3.

Tamberg, L.; Jaago, M.; Saalik, K.; **Sirp, A**.; Tuvikene, J.; Shubina, A.; Kiir, C. S.; Nurm, K.; Sepp, M.; Timmusk, T.; Palgi, M. (2020). Daughterless, the Drosophila orthologue of TCF4, is required for associative learning and maintenance of the synaptic proteome. Disease Models & Mechanisms, 13 (7), #dmm042747. DOI: 10.1242/dmm.042747.

Esvald, E. E.; Tuvikene, J.; **Sirp, A**.; Patil, S.; Bramham, C. R.; Timmusk, T. (2020). CREB Family Transcription Factors Are Major Mediators of BDNF Transcriptional Autoregulation in Cortical Neurons. Journal of Neuroscience, 40 (7), 1405−1426. DOI: 10.1523/JNEUROSCI.0367-19.2019.

*****- equal contribution

November 2019	Osalemine Eesti Inimesegeneetikaühingu 20 aastakonverentsil, Pärnu
Juuni 2019	Posterettekanne „Nordic Neuroscience“ konverentsil, “Regulation of the basic helix-loop-helix transcription factor TCF4 activity in neuronal cells”, Helsingi, Soome
November 2018	Osalemine Eesti Inimesegeneetikaühingu 19 aastakonverentsil, Viljandi
November 2018	Posterettekanne “Neuroscience 2018” konverentsil, “Regulation of the basic helix-loop-helix transcription factor TCF4 activity in neuronal cells”, San Diego, Ameerika Ühendriigid
August 2018	Osalemine “Frontiers in Delivery of Therapeutics” konverentsil, Tartu, Estonia
Mai 2018	Posterettekanne doktorantide teaduspäeval, “Regulation of the basic helix-loop-helix transcription factor TCF4 activity by disease-related missense variations”, Vehendi
Veebruar 2018	Osalemine doktorantide talvekoolis “Elu pärast PhD-d”
November 2017	Osalemine Eesti Inimesegeneetikaühingu 18 aastakonverentsil, Rakvere

Juhendatud väitekirjad

Carl Sander Kiir, magistrikraad, 2021, (juh) Alex Sirp; Mari Palgi, Neuronal activity-dependent regulation of daughterless expression in the central nervous system of Drosophila melanogaster (Neuraalsest aktiivsusest sõltuv daughterless-i ekspressiooni regulatsioon äädikakärbse kesknärvisüsteemis), Tallinna Tehnikaülikool, Loodusteaduskond, Keemia ja biotehnoloogia instituut

Britt-Anett Kristelštein, bakalaureusekraad, 2021, (juh) Alex Sirp; Tõnis Timmusk, Transkriptsiooniteguri TCF4 märgistamine hiire Neuro-2a rakkudes ja hiire embrüonaalsetes tüvirakkudes CRISPR-Cas9 meetodiga

Anastassia Šubina, magistrikraad, 2020, (juh) Alex Sirp; Tõnis Timmusk, Expression of transcription factor TCF4 in the developing and adult rodent brain (Transkriptsioonifaktori TCF4 ekspressioon arenevas ja täiskasvanud närilise peaajus), Tallinna Tehnikaülikool, Loodusteaduskond, Keemia ja biotehnoloogia instituut

Laura Kranich, bakalaureusekraad, 2020, (juh) Alex Sirp; Tõnis Timmusk, Silencing of Transcription Factor TCF4 in Mouse Neuro-2a and Human SH-SY5Y Cells Using the Crispr-Cas9 Method

Paul Mark Tammiste, uurimustöö, 2019, (juh) Alex Sirp; Martin Saar, Transkriptsioonifaktori TCF4 rakusisese lokalisatsiooni uurimine Huntingtoni tõves

Publikatsioonid

Sirp A,* Shubina A,* Tuvikene J, Tamberg L, Kiir CS, Kranich L, Timmusk T (2022) Expression of alternative transcription factor 4 mRNAs and protein isoforms in the developing and adult rodent and human tissues. Front. Mol. Neurosci., 15. 2022 Nov. DOI: 10.3389/fnmol.2022.1033224.

Sirp, A.,* Roots, K.,* Nurm, K., Tuvikene, J., Sepp, M., and Timmusk, T. (2021). Functional consequences of TCF4 missense substitutions associated with Pitt-Hopkins syndrome, mild intellectual disability, and schizophrenia. *J. Biol. Chem.* 297, 101381. doi: 10.1016/j.jbc.2021.101381.

Nurm, K.; Sepp, M.; Castany-Pladevall, C.; Creus-Muncunill, J.; Tuvikene, J.; **Sirp, A**; Vihma, H.; Blake, DJ..; Perez-Navarro, E.; Timmusk, T. (2021). Isoform-specific reduction of the basic helix-loop-helix transcription factor TCF4 levels in Huntington's disease. eNeuro, ENEURO.0197-21.2021. DOI: 10.1523/eneuro.0197-21.2021.

Sirp, A.*; Leite, K.*; Tuvikene, J.*; Nurm, K.; Sepp, M.; Timmusk, T. (2020). The Fuchs corneal dystrophy-associated CTG repeat expansion in the TCF4 gene affects transcription from its alternative promoters. Scientific Reports, 10 (1), #18424. DOI: 10.1038/s41598-020-75437-3.

Tamberg, L.; Jaago, M.; Saalik, K.; **Sirp, A**.; Tuvikene, J.; Shubina, A.; Kiir, C. S.; Nurm, K.; Sepp, M.; Timmusk, T.; Palgi, M. (2020). Daughterless, the Drosophila orthologue of TCF4, is required for associative learning and maintenance of the synaptic proteome. Disease Models & Mechanisms, 13 (7), #dmm042747. DOI: 10.1242/dmm.042747.

Esvald, E. E.; Tuvikene, J.; **Sirp, A**.; Patil, S.; Bramham, C. R.; Timmusk, T. (2020). CREB Family Transcription Factors Are Major Mediators of BDNF Transcriptional Autoregulation in Cortical Neurons. Journal of Neuroscience, 40 (7), 1405−1426. DOI: 10.1523/JNEUROSCI.0367-19.2019.

*- võrdne panus