

DOCTORAL THESIS

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

Alex Sirp

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
25/2023

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

ALEX SIRP



TALLINN UNIVERSITY OF TECHNOLOGY

School of Science

Department of Chemistry and Biotechnology

This dissertation was accepted for the defence of the degree 23/05/2023

Supervisor: Professor Tõnis Timmusk, PhD
Department of Chemistry and Biotechnology
Tallinn University of Technology
Tallinn, Estonia

Opponents: Prof. Benjamin D. Philpot
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

Prof. Toivo Maimets
University of Tartu
Tartu, Estonia

Defence of the thesis: 21/08/2023, Tallinn

Declaration:

Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.

Alex Sirp

Signature



European Union
European Regional
Development Fund



Investing
in your future

Copyright: Alex Sirp, 2023

ISSN 2585-6898 (publication)

ISBN 978-9916-80-000-3 (publication)

ISSN 2585-6901 (PDF)

ISBN 978-9916-80-001-0 (PDF)

Printed by Auratrükk

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
25/2023

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

ALEX SIRP



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 <i>TCF4</i>	29
5.3 The effect of disease-related missense variations and mutations in <i>TCF4</i> 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, MTF, 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 lineage 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 suppresses 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 phenomenon resulting from *Tcf4* overexpression in rats is the reduction of inflammatory and neuropathic pain sensitivity. This is achieved by suppressing neuronal activity of dorsal root ganglion neurons via downregulation of *Nav1.8* 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 (atoh1 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 *Mcp1* 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 (*Gla2*, *Gla3*) 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	<i>Tcf4</i> ^{+/-} 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	<i>Tcf4</i> ^{-/-} 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	<i>Tcf4</i> ^{+/-}	Prefrontal cortex/ P1, Adult
Phan et al., 2020	RNA-seq	<i>Tcf4</i> ^{+/Δ574-579} <i>Tcf4</i> ^{+/^{R579W}} <i>Actin-Cre::Tcf4</i> ^{+/^{floxed}} <i>Nestin-Cre::Tcf4</i> ^{+/^{floxed}}	Hemibrain/ P0-2, Adult
Schoof et al., 2020	RNA-seq	<i>Tcf4</i> ^{-/-} mouse	Forebrain/ Newborn
Mesman et al., 2020	RNA-seq	<i>Tcf4</i> ^{-/-} mouse	Cortex/ E14.5
Wittmann et al., 2021	scRNA-seq	<i>Tcf4</i> ^{-/-} long isoforms	Neocortex/ E18.5
Sarkar et al., 2021	RNA-seq	<i>Tcf4</i> ^{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 CAGGTGGT 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, CGCAGGTG) 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, CGCACGTG) increase TCF4 binding, modifications in the first two E-box nucleotides (in bold, CGCACGTG) decrease TCF4 binding and modifications affecting the middle E-box nucleotides (in bold, CGCACGTG) 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 7b1 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 7b1 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).

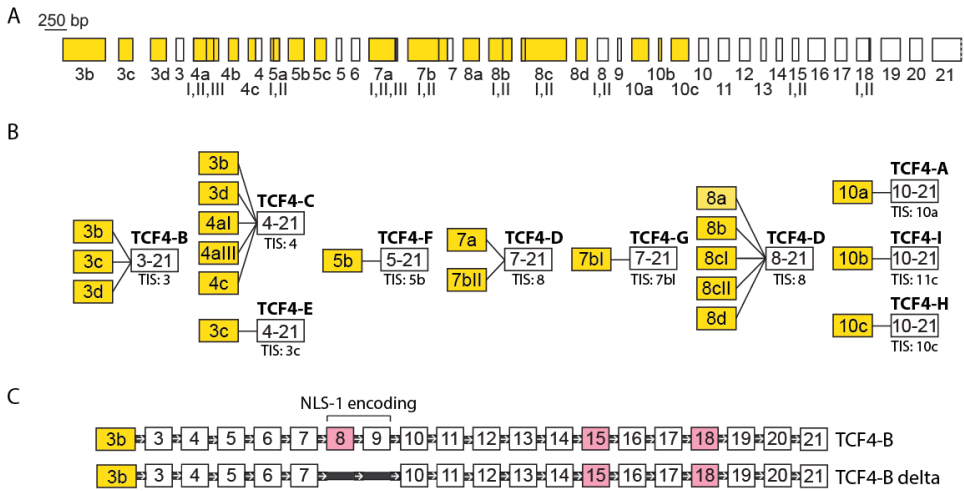


Figure 1. Human TCF4 gene structure. (A) TCF4 exons expressed in the human central nervous system shown in scale with the scale bar marked at the top. The name of each exon is shown below with roman numerals indicating multiple splice sites. White boxes represent internal and 3' exons and yellow boxes 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.

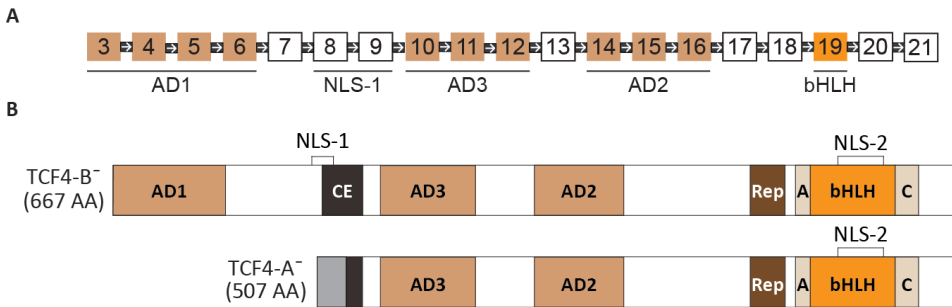


Figure 2. Schematic representation of TCF4 protein domains. (A) TCF4 internal exons with the functional domain-encoding exons marked below the exon names. (B) Schematic representation of protein domains for TCF4 isoforms B and A. The colored areas represent functional protein domains – activation domains 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; Nurme 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).

In 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
- Bioinformatic 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-existent;
- 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 (Gijssels 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.
- Brennan, 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., Børghlum, 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.
- Gijselincx, I., Van Mossevelde, S., van der Zee, J., Sieben, A., Engelborghs, S., De Bleeker, 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.
- Gottsche, 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 Ozyhar, 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., Timmus, 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 Smoothed 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). PGC1 α 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., Sahasrabudhe, 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., Fencikova, 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., Sønnderby, 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 have 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 (PHS), 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 PHS associated mutant *TCF4* proteins displayed impaired DNA binding activity. In addition, PHS 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 ability 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, epiteelimesenhümaalses 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 transkriptide 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 ajukoos ja hippokampus.

Kõrvalekaldeid *TCF4* geenis on seostatatud mitmete haigustega millest üks on hiline Fuchi endotelialne 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 Rett-i 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.



OPEN ACCESS

EDITED BY

Enrico Tongiorgi,
University of Trieste,
Italy

REVIEWED BY

Marc Forrest,
Northwestern University,
United States
Brady J. Maher,
Lieber Institute for Brain Development,
United States
Germana Meroni,
University of Trieste,
Italy

*CORRESPONDENCE

Tõnis Timmusk
tonis.timmusk@taltech.ee

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Neuroplasticity and Development, a section of the Frontiers in Molecular Neuroscience

RECEIVED 31 August 2022

ACCEPTED 05 October 2022

PUBLISHED 02 November 2022

CITATION

Sirp A, Shubina A, Tuvikene J, Tamberg L, Kiir CS, Kranich L and 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. 15:1033224.

doi: 10.3389/fnmol.2022.1033224

COPYRIGHT

© 2022 Sirp, Shubina, Tuvikene, Tamberg, Kiir, Kranich and Timmusk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Expression of alternative transcription factor 4 mRNAs and protein isoforms in the developing and adult rodent and human tissues

Alex Sirp^{1†}, Anastassia Shubina^{1†}, Jürgen Tuvikene^{1,2}, Laura Tamberg¹, Carl Sander Kiir¹, Laura Kranich¹ and Tõnis Timmusk^{1,2*}

¹Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia,

²Protobios LLC, Tallinn, Estonia

Transcription factor 4 (TCF4) belongs to the class I basic helix–loop–helix family of transcription factors (also known as E-proteins) and is vital for the development of the nervous system. Aberrations in the *TCF4* gene are associated with several neurocognitive disorders such as schizophrenia, intellectual disability, post-traumatic stress disorder, depression, and Pitt-Hopkins Syndrome, a rare but severe autism spectrum disorder. Expression of the human *TCF4* gene can produce at least 18 N-terminally distinct protein isoforms, which activate transcription with different activities and thus may vary in their function during development. We used long-read RNA-sequencing and western blot analysis combined with the analysis of publicly available short-read RNA-sequencing data to describe both the mRNA and protein expression of the many distinct TCF4 isoforms in rodent and human neural and nonneural tissues. We show that TCF4 mRNA and protein expression is much higher in the rodent brain compared to nonneural tissues. TCF4 protein expression is highest in the rodent cerebral cortex and hippocampus, where expression peaks around birth, and in the rodent cerebellum, where expression peaks about a week after birth. In human, highest *TCF4* expression levels were seen in the developing brain, although some nonneural tissues displayed comparable expression levels to adult brain. In addition, we show for the first time that out of the many possible TCF4 isoforms, the main TCF4 isoforms expressed in the rodent and human brain and other tissues are TCF4-B, -C, -D, -A, and -I. Taken together, our isoform specific analysis of TCF4 expression in different tissues could be used for the generation of gene therapy applications for patients with TCF4-associated diseases.

KEYWORDS

transcription factor TCF4, basic helix–loop–helix transcription factor, western blot analysis, neurodevelopment, long-read RNA sequencing, brain tissue, peripheral tissue

Introduction

Transcription factor 4 (TCF4) is a member of the class I basic helix–loop–helix transcription factor family (also known as E-proteins) and is the main E-protein expressed in the adult mouse brain (Massari and Murre, 2000; Fischer et al., 2014). TCF4 regulates numerous genes involved in neurodevelopment (Forrest et al., 2018) and has been shown to mediate its function by forming either homo- or heterodimers with proneural interaction partners such as achaete-scute homolog 1 (ASCL1; Persson et al., 2000) and neurogenic differentiation factor 2 (NEUROD2; Brzózka et al., 2010) as well as negative regulators known as inhibitor of DNA binding (ID) proteins (Chiaramello et al., 1995; Einarson and Chao, 1995). The expression of TCF4 interaction partners is strictly regulated, allowing TCF4 to possibly exert different functions during the development of the nervous system (Quednow et al., 2011).

Changes in the *TCF4* gene are linked to the development of many severe neurocognitive disorders such as schizophrenia (Stefansson et al., 2009; Ripke et al., 2014; Doostparast Torshizi et al., 2019), intellectual disability (Kharbanda et al., 2016), post-traumatic stress disorder (Gelernter et al., 2019), and depression (Wray et al., 2018). In addition, *de novo* mutations in one of the *TCF4* alleles cause Pitt-Hopkins syndrome (Brockschmidt et al., 2007; Zweier et al., 2007)—an autism spectrum disorder described by severe cognitive impairment, breathing abnormalities, motor delay, and distinctive facial features (Zollino et al., 2019). Interestingly, in addition to deletions and translocations, just a single nucleotide mutation in the basic helix–loop–helix encoding domain can completely impair the normal functionality of the TCF4 protein (Amiel et al., 2007; Zweier et al., 2007; Sepp et al., 2012; Sirp et al., 2021). *Tcf4* heterozygous mutant mice exhibit memory deficits, impaired motor control, and social isolation (Kennedy et al., 2016; Thaxton et al., 2018). Similar results have been noted in *Drosophila melanogaster*, where downregulation of Daughterless, the orthologue of TCF4, impairs memory and learning (Tamberg et al., 2020). Overexpression of *Tcf4* in mouse brain causes impairments in cognition and sensorimotor gating (Brzózka et al., 2010), and increased long term depression at synapses (Badowska et al., 2020). Homozygous *Tcf4* knockout mice have low viability and usually die around birth (Zhuang et al., 1996).

Expression of *Tcf4* was first described during late embryonic and early postnatal development in different mouse brain regions using northern blot analysis and *in situ* hybridization (Soosaar et al., 1994; Pscherer et al., 1996; Ravanpay and Olson, 2008). More recent studies have used quantitative droplet digital PCR

and reverse-transcription quantitative PCR to show that in the cerebral cortex *TCF4* mRNA expression peaks around birth and declines rapidly in the following 2 weeks (Li et al., 2019; Phan et al., 2020). Expression of TCF4 protein in the developing and adult mouse brain has been described in detail by Jung et al. (2018) using immunostaining with antibodies specific for longer TCF4 protein isoforms. During embryonic development of the brain, expression levels of long TCF4 isoforms are high in the areas which will develop into the cortex and the hippocampus. More specifically, long TCF4 isoforms are largely expressed in the germinal regions that will give rise to GABAergic and glutamatergic neurons of the cortex (Jung et al., 2018). In the brain of adult mice, TCF4 expression of long TCF4 isoforms is high in the cortex, hippocampus, and cerebellum (Jung et al., 2018). Similar results have been obtained by Kim and colleagues who used TCF4-GFP mice to characterize total TCF4 expression (Kim et al., 2020).

In human, TCF4 is expressed broadly, with the expression of different TCF4 isoforms varying between tissues (de Pontual et al., 2009; Sepp et al., 2011). Further analysis of human RNA sequencing (RNA-seq) data has revealed that the mRNA expression dynamics of *TCF4* in human and mouse appear to be conserved in the cerebral cortex—*TCF4* mRNA is highly expressed during fetal stages of development and reaches the maximum before birth, rapidly declines around birth until entering a relatively stable expression level from the early postnatal period to adulthood (Ma et al., 2018). As *TCF4* remains stably expressed in adult humans and rodents alike (de Pontual et al., 2009; Jung et al., 2018; Ma et al., 2018; Li et al., 2019) its expression is probably important for the normal functioning of the organism (Sarkar et al., 2021).

To date, none of the previous studies of *TCF4* mRNA (Ma et al., 2018; Li et al., 2019; Phan et al., 2020) and protein (Jung et al., 2018; Kim et al., 2020) expression have described expression of the variety of TCF4 isoforms. Expression of the mouse and human *TCF4* gene results in many different transcripts encoding N-terminally distinct protein isoforms which vary in their intracellular localization, transactivation capability (Sepp et al., 2011, 2017; Nurm et al., 2021) and possibly mediate their function depending on dosage (Ravanpay and Olson, 2008). Here, we investigated the complex expression dynamics of different TCF4 mRNAs and protein isoforms in the developing and adult rodent and human tissues. Our results can be used to estimate which TCF4 isoforms and at which proportions should be introduced into different tissues during development to generate gene therapy applications of the TCF4-associated diseases.

Materials and methods

Direct TCF4 RNA sequencing

Total RNA was extracted from a mixture of cerebral cortices from three P3 BALB/c mice using the RNeasy lipid tissue mini kit

Abbreviations: TCF4, Transcription factor 4; RNA-seq, RNA sequencing; RT, Room temperature; RT-qPCR, Reverse transcription quantitative PCR; gRNA, Guide RNA; P, Postnatal; E, Embryonic; CTX, Cerebral cortex; HC, Hippocampi; CB, Cerebellum; OB, Olfactory bulb; HTH, Hypothalamus; MB, Midbrain; TH, Thalamus; STRT, Striatum.

(Qiagen). Genomic DNA was digested on-column using RNase-Free DNase Set (Qiagen). Concentration of the purified RNA was determined with BioSpec-nano spectrophotometer (Shimadzu).

Before RNA-seq library preparation, 20 µg of P3 BALB/c cortical RNA was enriched for *Tcf4* transcripts in wash/binding buffer (0.5 M NaCl, 20 mM Tris-HCl pH 7.5, and 1 mM EDTA) with 8 µM each of three 5' biotinylated oligonucleotides (Microsynth AG)—two oligonucleotides were complementary to the 3' untranslated region of *Tcf4*, and one was complementary to the bHLH region of *Tcf4* (Supplementary Table S1). As there are no known *Tcf4* transcripts that lack the bHLH region or exon 21, we expect that our *Tcf4* mRNA enrichment strategy is unbiased and enriches all possible TCF4 transcripts (Sepp et al., 2011; Nurm et al., 2021). First the mixture was incubated at 70°C for 2 min and then cooled to room temperature in about 30 min in a heating block. When the heating block reached 60°C, 100 units of RiboLock RNase Inhibitor (Thermo Fisher Scientific) was added to the mixture.

Next, Pierce streptavidin magnetic beads (Thermo Fisher Scientific) were prepared for the binding reaction. For that, 40 µl of magnetic beads was washed in 800 µl wash/binding buffer and then suspended in 30 µl of wash/binding buffer. The magnetic beads were then added to the previously annealed oligonucleotide-RNA mixture for the binding reaction. The bead-oligonucleotide-RNA mixture was incubated at RT for 90 min with occasional agitation by hand. After 90 min, the flow-through sample was collected and after that the beads were washed twice with 100 µl of wash/bind buffer followed by three washes with 100 µl of ice-cold low salt buffer (0.15 M NaCl, 20 mM Tris-HCl pH 7.5, and 1 mM EDTA). For elution, the magnetic beads were incubated at 70°C for 5 min in 25 µl of nuclease free water (Qiagen) twice, with the total volume of eluted RNA being 50 µl. In total, two *Tcf4* RNA enrichments were done—the first *Tcf4* enriched RNA sample was sequenced twice and the second once.

RNA sequencing library was prepared separately for each of the three sequencing experiments according to the Sequence-specific direct RNA sequencing protocol SQK-RNA002 (Oxford Nanopore Technologies). The RNAClean XP beads (Agencourt) were substituted with the Mag-Bind total pure NGS magnetic beads (Omega Bio-tek). Sequencing was done three times with the MinION sequencer, FLO-MIN106 flow-cell (new flow cells were used for each experiment), and SQK-RNA002 kit using MinKNOW software (version 3.6.5; Oxford Nanopore Technologies).

Base-calling of the direct RNA sequencing data was performed using Guppy Basecalling Software (version 4.0.11 + f1071ceb, Oxford Nanopore Technologies) with high-accuracy basecalling algorithm. Failed reads were discarded and the passed reads were mapped to mouse GRCm38.p6 genome (obtained from Gencode) using Minimap2 (version 2.17-r941) with the following settings: -ax splice-uf -k14. The generated sam files were converted to bam format using Samtools (version 1.9), the alignments of all three replicates were combined and only reads mapping to the *Tcf4* gene locus were kept. The resulting merged and filtered bam file was

then converted to bed12 file format using bedtools (version 2.28.0) for easier visualization. All reads mapping to the *Tcf4* locus in bed12 format can be found in Data Sheet 1. Raw sequencing reads mapping to the *Tcf4* locus can be found in Data Sheet 2. The final data was visualized in Integrated Genomics Viewer and the transcripts encoding *Tcf4* isoforms were manually quantified. Aberrant transcripts were excluded from the analysis.

Guide RNA design and cloning

The University of California Santa Cruz Genome Browser Gateway¹ was used to define the genomic region of mouse exon 3 and exon 10a protein coding regions for guide RNA (gRNA) design. The genomic region for mouse exon 3 and exon 10a was chr18:69,347,299–69,347,369 and chr18:69,593,516–69,593,584, respectively, according to mouse GRCm38/mm10 (Dec. 2011) assembly. In total, three gRNAs were designed for exon 3 and two for exon 10a using Benchling Inc.² CRISPR guide design software.

To insert the gRNA targeting region-containing oligonucleotides effectively into the PX459 (Addgene #62988) expression vector, nucleotides were added to the 5' ends of gRNAs that were complementary to the sticky ends produced after restriction of the PX459 plasmid with the BbsI restriction enzyme (Thermo Scientific). In addition, a guanine nucleotide was added to the 5' end of each forward oligonucleotide sequence of gRNA as it has been found to increase targeting efficiency.³ The designed sequences are included in Supplementary Table S2. The oligonucleotides were ordered from Microsynth AG.

Cell culture and transfection

Mouse Neuro2a and human SH-SY5Y cells were grown in DMEM (Dulbecco's modified Eagle's medium, Thermo Scientific) medium, supplemented with 10% fetal bovine serum (Pan Biotech), 100 U/ml penicillin, and 0.1 mg/ml streptomycin (Thermo Scientific).

For transfection, cells were plated on a 12-well plate (Greiner) in 800 µl medium per well 24–48 h before transfection. At the time of the transfection, the cells were at 50–70% confluency. Neuro2a cells were transfected with 500 ng of the pEGFP plasmid and 500 ng of the PX459 plasmid expressing the respective gRNA, Cas9, and Puromycin resistance gene using Lipofectamine 2000 (Invitrogen). In each experiment, DNA to transfection reagent ratio was 1:2. For preparation of protein lysates, the cells were lysed in 1x Laemmli buffer [0.062 M Tris-HCl pH 6.8, 2% SDS, 5% 2-mercaptoethanol (Roth), 10% glycerol, and 0.01% bromophenol blue].

1 <https://genome.ucsc.edu>

2 <https://benchling.com>

3 <http://www.addgene.org/crispr/zhang/>

Reverse transcription PCR

Total RNA was extracted from Neuro2a cells using the RNeasy mini kit (Qiagen). Genomic DNA was digested on-column using RNase-Free DNase Set (Qiagen). Concentrations of the purified RNAs were determined with BioSpec-nano spectrophotometer (Shimadzu). cDNA was synthesized from Neuro2a total RNA using Superscript IV Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. Primers used for reverse transcription PCR are listed in [Supplementary Table S3](#).

Animal husbandry

The protocols involving animals were approved by the ethics committee of animal experiments at Ministry of Agriculture of Estonia (Permit Number: 45). All experiments were performed in accordance with the relevant guidelines and regulations. Wistar rats (RccHan:WI, Envigo) and C57BL/6 and BALB/c mouse strains (Envigo) were used in this study. Animals were maintained in conventional polycarbonate or H-TEMP polysulfone cages (2–4 animals per cage) with *ad libitum* access to clean water and food pellets (ssniff Spezialdiäten) under a 12-h light/dark cycle in humidity and temperature-controlled room (temperature $22 \pm 1^\circ\text{C}$ and humidity $50 \pm 10\%$).

To establish timed pregnancy for studying embryonic (E) development, the female mouse estrous cycle was monitored by visual observations of the vaginal opening of each female mouse based on the criteria described by [Champlin et al. \(1973\)](#). Mice in the proestrus or estrous phase of the cycle were selected for mating. Animals were bred in the evening and vaginal post-coitum protein plug was checked in the next morning no more than 12 h later. The morning that a plug was found was designated as E0.5 gestational stage. The day of the animal birth was designated as postnatal (P) 0 stage.

Tissue isolation and protein extraction

Mice and rats were euthanized by carbon dioxide inhalation and decapitated with a guillotine. Dissection of tissue samples was done in ice-cold 1x phosphate-buffered saline solution. Each sample contained tissues pooled together from three different animals for biological diversity and sufficient protein extraction at early developmental stages. The mouse and rat cerebral cortex, hippocampus, cerebellum, olfactory bulb, hypothalamus, and pons including medulla, midbrain, and thalamus were collected. Striatum was collected only for the BALB/c mouse strain. Tissue collection for mouse and rat brain regions occurred at P0, 3, 5, 7, 10, 14, 21, 60, and P0, 3, 5, 10, 14, 30, 60, respectively. In addition to postnatal days, collection of total mouse brain samples occurred at E13, 15, and 18. Mouse and rat peripheral tissues, skin, lung, kidney, heart, diaphragm, muscle, bladder, stomach, pancreas, thymus, spleen, liver, and blood, were collected at developmental stages

P0, 14, and 60. After collection, tissue samples were stored at -80°C until further processing.

Tissues were homogenized on ice with tissue grinder PELLET PESTLE® Cordless Motor (Kimble-Chase, DWK Life Sciences) in ice-cold Radioimmunoprecipitation assay buffer [RIPA, 50 mM Tris pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.5% sodium dodecyl sulfate (SDS), and 1x Roche Protease Inhibitor Cocktail Complete]. Lysates were sonicated for 15 s with Torbeo Ultrasonic probe sonicator (36810-series, Cole Parmer), and insoluble material was removed by centrifugation at 4°C for 20 min at 16,000 g. Protein concentration was measured with Pierce BCA Protein Assay Kit (Thermo Scientific).

Protein lysates from human post-mortem cerebral cortex and hippocampus were prepared like rodent lysates. All protocols using human tissue samples were approved by Tallinn Committee for Medical Studies, National Institute for Health Development (Permit Number 402). All experiments were performed in accordance with relevant guidelines and regulations.

In vitro protein translation

TCF4-A⁺, -A⁻, -B⁺, B⁻, -C⁻, -D⁻ and TCF4-I⁻ isoforms were translated *in vitro* using pcDNA3 plasmids encoding the respective TCF4 isoforms ([Sepp et al., 2011](#)) with ThT Quick Coupled Transcription/Translation System (Promega). Equal volumes of *in vitro* translated TCF4 protein mixtures were used for western blot analysis.

Western blot analysis

For western blot analysis, protein lysates in RIPA were diluted to the same concentration in 1x Laemmli buffer. 55 μg of each sample was electrophoretically separated by SDS-polyacrylamide gel electrophoresis in 8% gel and transferred to polyvinylidene difluoride membrane (Millipore) in Towbin buffer (25 mM Tris, 192 mM glycine, 20% methanol, 0.1% SDS, and pH 8.3) using wet transfer. Membranes were blocked with 5% skimmed milk (Sigma-Aldrich) in 1x Tris Buffered Saline with 0.1% Tween-20 (TBST, Sigma Aldrich) before incubating with primary [mouse monoclonal anti-ITF-2 (TCF4); C-8, 1:1,000 dilution, Santa Cruz] and secondary (goat polyclonal anti-mouse IgG HRP-conjugated antibody 1:5,000 dilution, Thermo Scientific) antibody in 2.5% skimmed milk-TBST solution overnight at 4°C and 1 h at room temperature, respectively. Specificity of the anti-ITF-2 (TCF4; C-8, Santa Cruz) antibody has been previously validated using tissue lysates from *Tcf4* knockout mice ([Nurm et al., 2021](#)). For peripheral tissues, mouse IgG kappa binding protein conjugated to HRP was used as a secondary antibody (1:5,000 dilution, Santa Cruz). Chemiluminescence signal detection with SuperSignal West Femto or Atto Chemiluminescence Substrate (Thermo Scientific). The chemiluminescence signal was visualized with ImageQuant LAS 4000 bioimager (GE Healthcare) and densitometric quantification was performed with ImageQuant TL

8.2 image analysis software (GE Healthcare). Membrane staining with Coomassie solution (0.1% Coomassie Brilliant Blue R-250, 25% ethanol, and 7% acetic acid) was used as a loading control and for total protein normalization.

Analysis of publicly available RNA-seq datasets

Raw RNA-seq datasets of human, mouse, and rat were obtained from EMBL-EBI European Nucleotide Archive database using www.sra-explorer.info (Keane et al., 2011; ENCODE Project Consortium, 2012; Schmitt et al., 2014; Yu et al., 2014; Vied et al., 2016; Li et al., 2017; Söllner et al., 2017; Cardoso-Moreira et al., 2019; Luo et al., 2020; Shafik et al., 2021; see Supplementary Table S4 for accession numbers and sample information). Adapter and quality trimming were done using BBDuk (part of BBDuk version 38.90) with the following parameters: ktrim = r k = 23 mink = 11 hdist = 1 tbo qtrim = lr trimq = 10 maq = 10 minlen = 25. Mouse reads were mapped to mm10 (primary assembly and annotation obtained from GENCODE, release M25, GRCm38), rat reads were mapped to rn6 (primary assembly and annotation obtained from Ensembl, release 104, RGSC 6.0/Rnor_6.0), and human reads were mapped to hg19 (primary assembly and annotation obtained from GENCODE, release 37, GRCh37) using STAR aligner (version 2.7.4a) with default parameters. To increase sensitivity for unannotated splice junctions, splice junctions obtained from the first pass were combined per dataset and filtered as follows: junctions on mitochondrial DNA and non-canonical intron motifs were removed; only junctions detected in at least 10% of samples (rounded up to the nearest integer) in the whole dataset were kept. Filtered junctions were added to the second pass mapping using STAR. Intron spanning reads were quantified using FeatureCounts (version 2.0.1). The following parameters were used for paired-end data: -p -B -C -J; and single-end data: -J. To count reads from *TCF4* extended exons (exons 4c and 7bII), reads crossing a 1 bp region 2 bp 5' from the internal exon 4 and 7, respectively, were quantified using FeatureCounts and a custom-made saf file. Splice junctions in the *TCF4* locus were manually curated and annotated to *TCF4* isoforms according to Sepp et al. (2011).

A custom R script⁴ was used to quantify the expression of different *TCF4* transcripts from analyzed RNA-seq datasets. Briefly, RNA-seq reads crossing the indicated *TCF4* splice-junctions were normalized using all splice-junction crossing reads in the respective sample. Then, the data were summarized by the Exon column (Supplementary Tables S5–S7). To acquire total *TCF4* mRNA levels, the mean value of exon-junctions from 10–11 to 19–20 was taken for the analysis. Aggregated mouse and rat data was meta-analyzed in tandem with human data to study *TCF4* expression during development. The expression of *TCF4*

transcripts encoding specific isoforms was assessed by quantifying the number of reads crossing the exon-junctions specific for the *TCF4* isoforms. Data were summarized in each tissue and age group by the Isoform column. Values for each isoform were then divided with the sum of all annotated isoforms to show isoform composition in percentages. The results were visualized using ggplot2 (version 3.3.5) in R (version 4.1.2). The *TCF4* exon-junction data used for the analysis of *TCF4* transcripts in mouse, rat, and human datasets can be found in Supplementary Tables S5–S7, respectively.

Data mining and visualization was also performed on human GTEx portal exon-exon junction dataset (dbGaP Accession phs000424.v8.p2) and human developmental transcriptome data from BrainSpan (RNA-Seq Gencode v10 summarized to genes). The human GTEx data used for the analyses described in this manuscript were obtained from the GTEx Portal⁵ on 12/01/2021 and the human BrainSpan data was obtained from the BrainSpan Atlas of the Developing Human Brain⁶ on 12/01/2021. For information about rodent and human developmental stages and the number of individual data points per developmental stage, see Supplementary Tables S8–S12.

Results

Five N-terminally distinct *TCF4* protein isoforms are expressed in the developing mouse cerebral cortex

The use of numerous alternative 5' exons results in the expression of many transcripts from the *Tcf4* gene, resulting in a variety of *TCF4* protein isoforms with different expression patterns between tissue types (Sepp et al., 2011; Nurm et al., 2021). We have previously described transcripts from the mouse *Tcf4* gene based on available mRNA and expressed sequence tag data from various tissues (Nurm et al., 2021) but characterizing all the isoform encoding transcripts using short-read sequencing is complicated. Here, we did long-read direct RNA-sequencing (RNA-seq) on the Oxford Nanopore Technologies platform from postnatal day 3 (P3) mouse cerebral cortex. Direct RNA-seq eliminates the bias which may result from complementary DNA synthesis used in conventional RNA-seq methods. Briefly, we extracted total RNA from the P3 mouse cerebral cortex which was then enriched for *Tcf4* transcripts using a combination of oligonucleotides complementary to the 3' untranslated region and the basic helix-loop-helix region of *Tcf4* (Figure 1A; Supplementary Table S1). The Oxford Nanopore Technologies platform begins sequencing from the 3' end of RNA meaning that early sequencing termination can result in reads which do not reach the 5' exons of *Tcf4* transcripts. Analysis of our results

4 https://github.com/CSKiir/Sirp_et_al_2022

5 <https://gtexportal.org>

6 <http://brainspan.org>

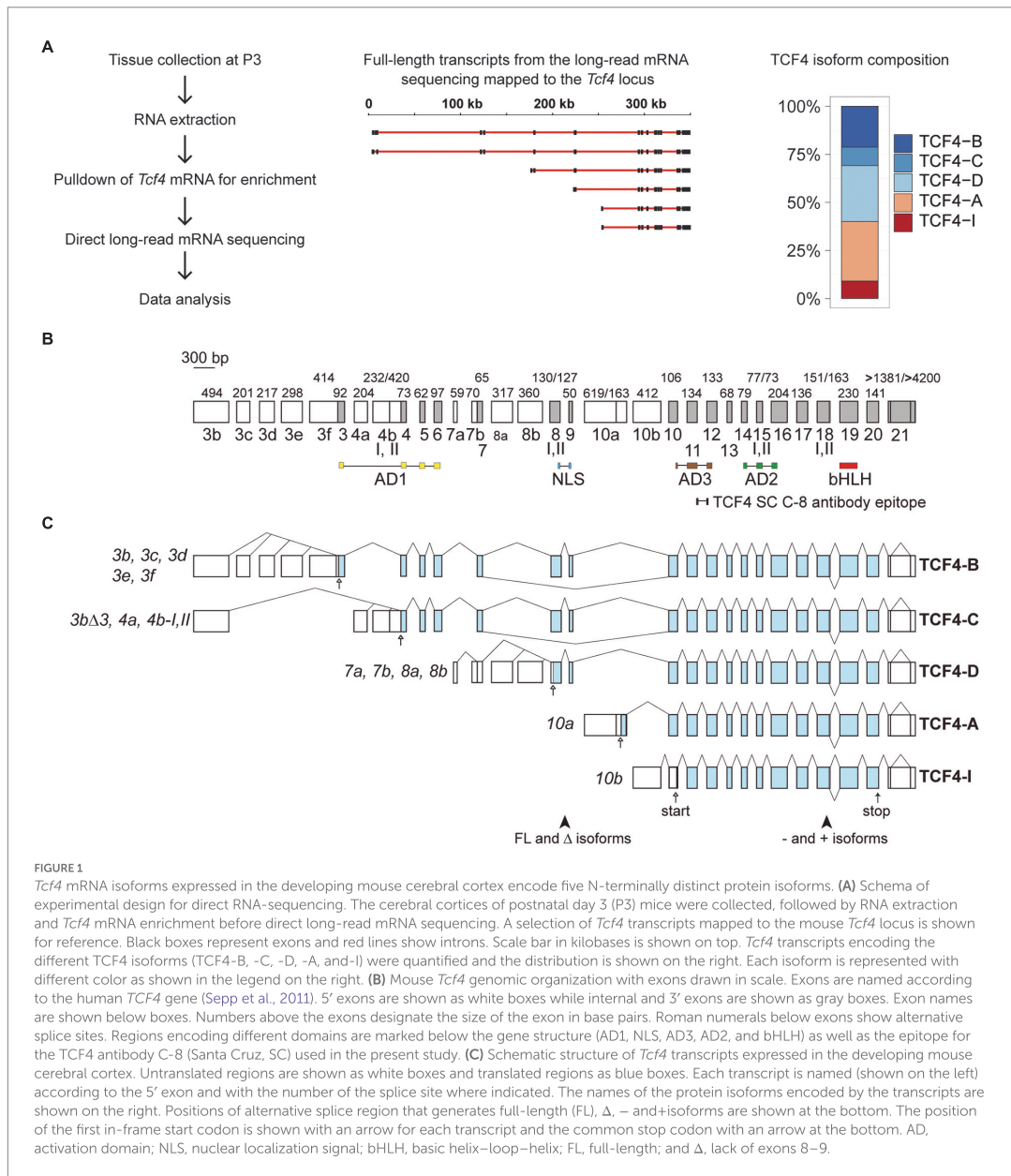


FIGURE 1

Tcf4 mRNA isoforms expressed in the developing mouse cerebral cortex encode five N-terminally distinct protein isoforms. (A) Schema of experimental design for direct RNA-sequencing. The cerebral cortices of postnatal day 3 (P3) mice were collected, followed by RNA extraction and *Tcf4* mRNA enrichment before direct long-read mRNA sequencing. A selection of *Tcf4* transcripts mapped to the mouse *Tcf4* locus is shown for reference. Black boxes represent exons and red lines show introns. Scale bar in kilobases is shown on top. *Tcf4* transcripts encoding the different TCF4 isoforms (TCF4-B, -C, -D, -A, and -I) were quantified and the distribution is shown on the right. Each isoform is represented with different color as shown in the legend on the right. (B) Mouse *Tcf4* genomic organization with exons drawn in scale. Exons are named according to the human *TCF4* gene (Sepp et al., 2011). 5' exons are shown as white boxes while internal and 3' exons are shown as gray boxes. Exon names are shown below boxes. Numbers above the exons designate the size of the exon in base pairs. Roman numerals below exons show alternative splice sites. Regions encoding different domains are marked below the gene structure (AD1, NLS, AD3, AD2, and bHLH) as well as the epitope for the TCF4 antibody C-8 (Santa Cruz, SC) used in the present study. (C) Schematic structure of *Tcf4* transcripts expressed in the developing mouse cerebral cortex. Untranslated regions are shown as white boxes and translated regions as blue boxes. Each transcript is named (shown on the left) according to the 5' exon and with the number of the splice site where indicated. The names of the protein isoforms encoded by the transcripts are shown on the right. Positions of alternative splice region that generates full-length (FL), Δ, - and +isoforms are shown at the bottom. The position of the first in-frame start codon is shown with an arrow for each transcript and the common stop codon with an arrow at the bottom. AD, activation domain; NLS, nuclear localization signal; bHLH, basic helix-loop-helix; FL, full-length; and Δ, lack of exons 8–9.

showed that most of the 1,336 RNAs which mapped to the mouse *Tcf4* gene were short and mapped only to the last exon of *Tcf4* gene. However, we obtained 163 *Tcf4* transcripts that reached from the 3'untranslated region to the 5' terminal exons and were thus considered full-length based on previous knowledge about the rodent *Tcf4* gene structure (Nurm et al., 2021). We then quantified the potential TCF4 protein isoforms encoded by these transcripts

(Figure 1A). The results showed that in the mouse P3 cerebral cortex ~20% of mRNAs transcribed from *Tcf4* gene encode isoform TCF4-B; 10% encode isoform TCF4-C; 30% encode isoform TCF4-D; 30% encode isoform TCF4-A and 10% encode isoform TCF4-I (Figure 1A). The presence of plus (containing the RSR amino acid sequence) and minus (without the RSR amino acid sequence; Corneliusen et al., 1991; Nurm et al., 2021) TCF4

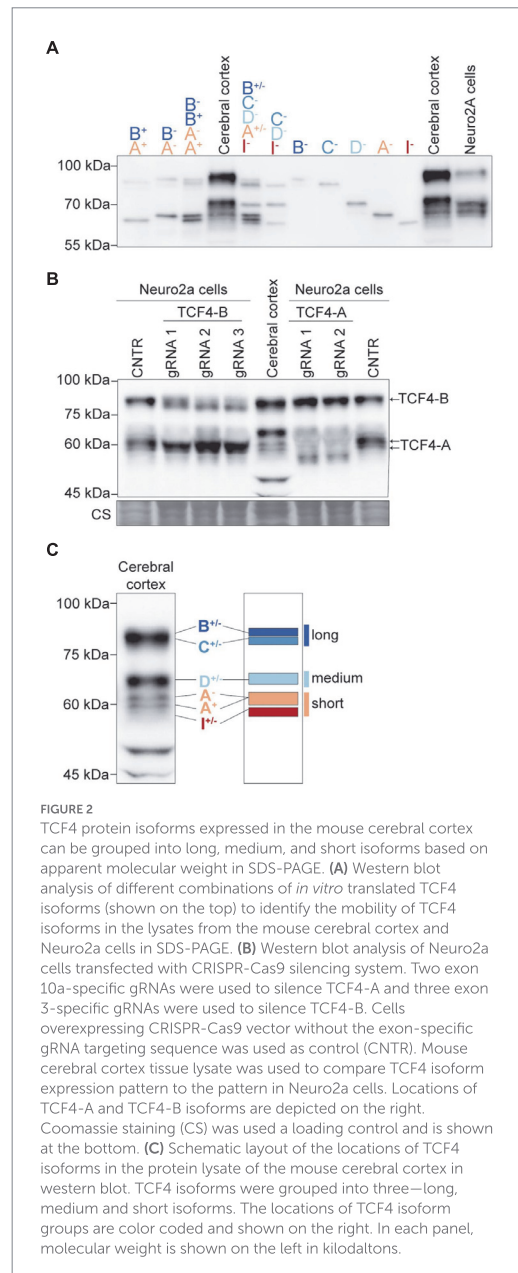
isoforms encoded by *Tcf4* mRNAs was roughly equal. Overall, our data show that the expression of mouse *Tcf4* gene leads to numerous transcripts due to the presence of 32 exons out of which 13 are alternative 5' exons and 18 are internal exons and one is a terminal 3' exon (Figures 1B,C). Using long-read sequencing, we discovered two novel 5' exons (exon 3e and 3f) that are included in *Tcf4* transcripts encoding TCF4-B isoform (Figures 1B,C).

We then described the expression pattern of TCF4 protein isoforms in SDS-PAGE/western blot analysis. For that, we used *in vitro* translated TCF4-B, -C, -D, -A, and -I plus or minus TCF4 isoforms. This allowed us to compare different combinations of *in vitro* translated TCF4 isoforms in western blot to the TCF4 protein pattern in the mouse cerebral cortex. In good agreement with our RNA-seq experiment, the combination of *in vitro* translated proteins TCF4-B, -C, -D, -A, and -I resembled the protein pattern of TCF4 in the cerebral cortex (Figure 2A). The variability in *in vitro* translated TCF4 isoform levels in western blot (Figure 2A) could arise from differential translation rate of the respective TCF4 isoform encoding plasmids. Next, we determined the apparent molecular weight and location of endogenously expressed TCF4 isoforms—TCF4-B and TCF4-A in SDS-PAGE/western blot analysis. To this end, we constructed a CRISPR-Cas9 system to inhibit the expression of these TCF4 isoforms in Neuro2a cell line by generating frameshift mutations in the unique exons encoding these isoforms (exons 3 and 10a, respectively). We could not specifically silence the expression of TCF4-D since its translation start site is in internal exon 8—a frameshift mutation in exon 8 would cause the silencing of not only TCF4-D, but also TCF4-B and -C isoforms. We used Neuro2a cells, which show high endogenous expression of TCF4 as confirmed by RT-PCR and western blot analysis (Figure 2B; Supplementary Figure S1). By expressing the generated CRISPR-Cas9 system in Neuro2a cells, we were able to inhibit the expression of TCF4-B and -A and thus confirm the location of these protein isoforms in western blot analysis (Figure 2B). Furthermore, western blot analysis showed that expression pattern of TCF4 isoforms was similar in Neuro2a cells and P3 mouse cerebral cortex (Figure 2B).

Taken together, the main N-terminally distinct TCF4 isoforms expressed in the early postnatal mouse cerebral cortex are TCF4-B, -C, -D, -A, and -I. We classified the detected TCF4 signals into three groups based on their molecular weight: long isoforms (TCF4-B and -C), medium isoforms (TCF4-D), and short isoforms (TCF4-A and -I; Figure 2C).

Expression of TCF4 protein in the mouse brain is highest around birth

Next, we studied the changes in TCF4 protein expression in the mouse brain throughout pre- and postnatal development. For that, we made mouse whole brain lysates from two strains



(BALB/C and C57BL/6) at 11 different developmental stages ranging from E13.5 to P60, and performed western blot analysis (Figures 3A,B). Both mouse strains exhibited expression of long, medium, and short TCF4 isoforms, with the highest expression of total TCF4 detected at late prenatal and early postnatal development. After peaking, TCF4 expression gradually declined

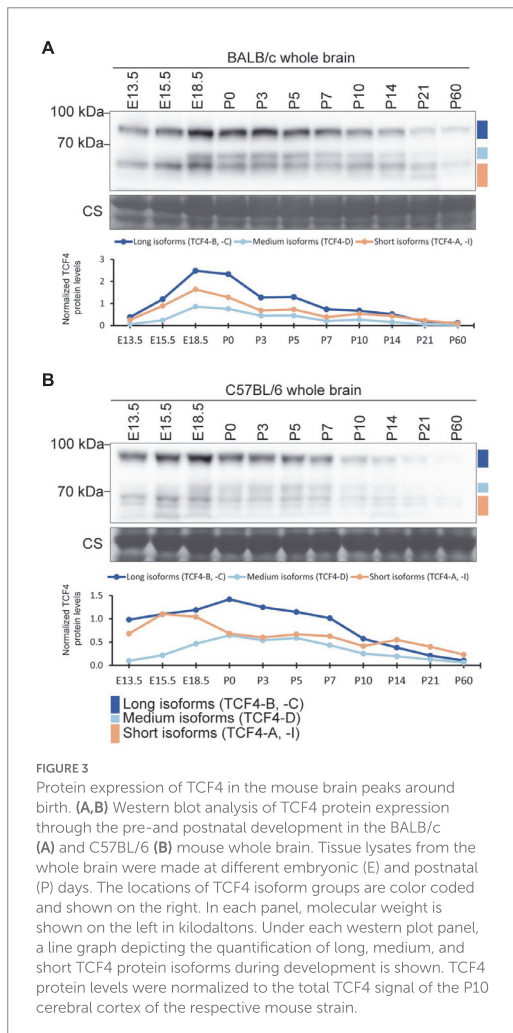


FIGURE 3

Protein expression of TCF4 in the mouse brain peaks around birth. (A,B) Western blot analysis of TCF4 protein expression through the pre- and postnatal development in the BALB/c (A) and C57BL/6 (B) mouse whole brain. Tissue lysates from the whole brain were made at different embryonic (E) and postnatal (P) days. The locations of TCF4 isoform groups are color coded and shown on the right. In each panel, molecular weight is shown on the left in kilodaltons. Under each western plot panel, a line graph depicting the quantification of long, medium, and short TCF4 protein isoforms during development is shown. TCF4 protein levels were normalized to the total TCF4 signal of the P10 cerebral cortex of the respective mouse strain.

during postnatal development of the brain (Figures 3A,B). While long and short TCF4 isoforms were detected at all stages, the medium-sized TCF4 isoforms became more apparent at later fetal stages and were almost undetectable before stage E18.5 (Figures 3A,B).

To better compare TCF4 total levels and isoform expression patterns between the two mouse strains, brain samples of the two strains from early postnatal development (P0–10) were analyzed in the same western blot experiment (Supplementary Figure S2). Our results revealed that the two mouse strains showed no major differences in TCF4 expression levels or expression patterns (Supplementary Figure S2). Altogether, these results indicated that TCF4 is expressed at both pre- and postnatal stages of the mouse brain development, with long and short TCF4 isoforms presented at all stages.

Expression of TCF4 is highest in the cerebral cortex, hippocampus, cerebellum, and olfactory bulb in the rodent brain

Next, we analyzed mRNA and protein expression of TCF4 in various rodent brain regions. First, we conducted a meta-analysis of available short-read RNA-seq data (Keane et al., 2011; ENCODE Project Consortium, 2012; Schmitt et al., 2014; Yu et al., 2014; Vied et al., 2016; Li et al., 2017; Söllner et al., 2017; Cardoso-Moreira et al., 2019; Luo et al., 2020; Shafik et al., 2021) to quantify the expression of total *Tcf4* mRNA and different *Tcf4* transcripts encoding distinct protein isoforms in rodents. Where possible, *Tcf4* expression dynamics was studied during different stages of pre- and postnatal development.

In the mouse brain, *Tcf4* mRNA expression was highest in the cerebral cortex, followed by the cerebellum, midbrain and hypothalamus (Figure 4A; Supplementary Figure S3A). A decrease in *Tcf4* mRNA expression after birth was seen for all the studied brain regions except for the cerebellum, which displayed relatively stable *Tcf4* mRNA levels during postnatal development. The majority of expressed *Tcf4* transcripts encoded isoforms TCF4-B, -C, -D, -A, and -I, with transcripts encoding TCF4-A showing the highest overall expression (Figure 4A; Supplementary Figure S4A) in the mouse brain. The cerebral cortex was the only brain region that displayed a notable change in the expression pattern of transcripts encoding different TCF4 isoforms—during development the expression of TCF4-A decreased and the expression of TCF4-D increased (Figure 4A).

We then sought to describe TCF4 expression at the protein level in mouse brain regions. For this, we dissected the cerebral cortex, hippocampus, cerebellum, striatum, pons, olfactory bulb, hypothalamus, thalamus, and midbrain at eight postnatal stages (P0, 3, 5, 7, 10, 14, 21, and 60) from BALB/C (Figure 4) and C57BL/6 (Supplementary Figure S5) mice, prepared protein lysates and analyzed TCF4 levels by western blot. First, we compared TCF4 protein expression across distinct brain regions at two postnatal stages, P0 and P10 by western blot analysis (Figures 4B,C; Supplementary Figures S5A,B). We observed high TCF4 expression levels in the cerebral cortex and hippocampus at P0, and in the cerebellum at P10 (Figures 4B,C; Supplementary Figures S5A,B). The long and short TCF4 protein isoforms were present in all studied brain regions (Figures 4B,C; Supplementary Figures S5A,B). The medium isoforms had more restricted patterns being detected at high levels in the cerebral cortex and hippocampus, at low levels in the cerebellum, olfactory bulb, and pons, and were below the detection limits in other brain regions (Figures 4B,C).

Next, we focused on the developmental dynamics of TCF4 protein expression in all the dissected mouse brain regions during postnatal development (P0–60; Figure 4D; Supplementary Figure S5C). To better compare TCF4 signals between individual brain regions across development, we used tissue lysate from the P10 cerebral cortex in each experiment as

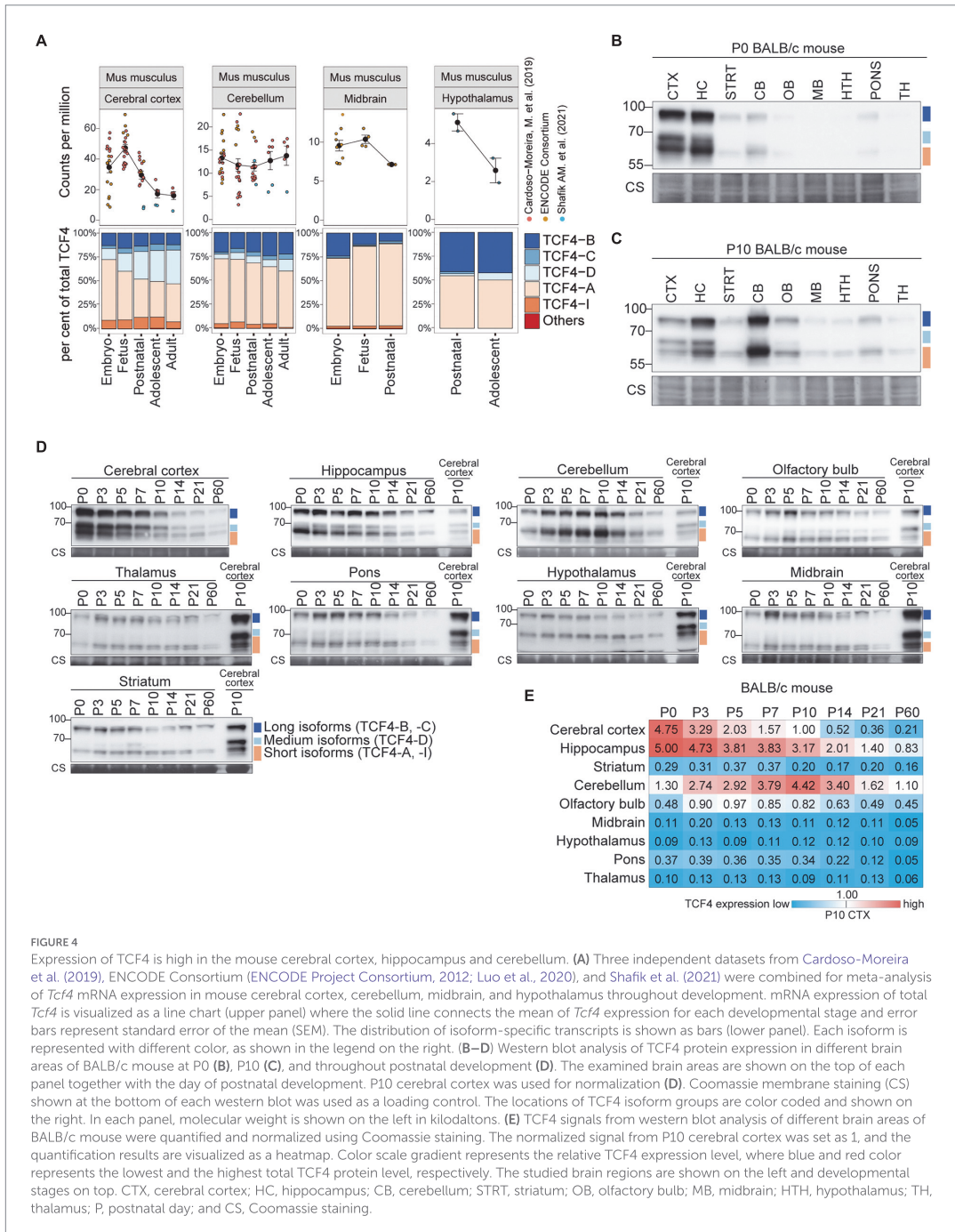


FIGURE 4

Expression of TCF4 is high in the mouse cerebral cortex, hippocampus and cerebellum. (A) Three independent datasets from Cardoso-Moreira et al. (2019), ENCODE Consortium (ENCODE Project Consortium, 2012; Luo et al., 2020), and Shafik et al. (2021) were combined for meta-analysis of *Tcf4* mRNA expression in mouse cerebral cortex, cerebellum, midbrain, and hypothalamus throughout development. mRNA expression of total *Tcf4* is visualized as a line chart (upper panel) where the solid line connects the mean of *Tcf4* expression for each developmental stage and error bars represent standard error of the mean (SEM). The distribution of isoform-specific transcripts is shown as bars (lower panel). Each isoform is represented with different color, as shown in the legend on the right. (B–D) Western blot analysis of TCF4 protein expression in different brain areas of BALB/c mouse at P0 (B), P10 (C), and throughout postnatal development (D). The examined brain areas are shown on the top of each panel together with the day of postnatal development. P10 cerebral cortex was used for normalization (D). Coomassie membrane staining (CS) shown at the bottom of each western blot was used as a loading control. The locations of TCF4 isoform groups are color coded and shown on the right. In each panel, molecular weight is shown on the left in kilodaltons. (E) TCF4 signals from western blot analysis of different brain areas of BALB/c mouse were quantified and normalized using Coomassie staining. The normalized signal from P10 cerebral cortex was set as 1, and the quantification results are visualized as a heatmap. Color scale gradient represents the relative TCF4 expression level, where blue and red color represents the lowest and the highest total TCF4 protein level, respectively. The studied brain regions are shown on the left and developmental stages on top. CTX, cerebral cortex; HC, hippocampus; CB, cerebellum; STRT, striatum; OB, olfactory bulb; MB, midbrain; HTH, hypothalamus; TH, thalamus; P, postnatal day; and CS, Coomassie staining.

a calibrator and quantified the results (Figure 4E; Supplementary Figure S5F). In agreement with our direct comparisons (Figures 4B,C; Supplementary Figures S5A,B), the highest levels of TCF4 expression were observed in the cerebral

cortex, hippocampus, and cerebellum, with all the other studied brain regions showing either moderate (olfactory bulb) or low (striatum, midbrain, hypothalamus, pons, and thalamus) TCF4 expression (Figure 4E; Supplementary Figure S5F). While in the

cerebral cortex and hippocampus TCF4 expression peaked just after birth (Figure 4E; Supplementary Figure S5F), cerebellum displayed a slightly delayed increase in TCF4 expression, peaking around a week after birth (Figure 4E; Supplementary Figure S5F). Notably, cerebellum and hippocampus retained a higher expression of TCF4 for a longer period compared to the cerebral cortex (Figure 4E; Supplementary Figure S5F).

To extend our observations for the mouse, we next characterized TCF4 expression in the brain of another important model organism, the rat (Figure 5). *Tcf4* mRNA expression was highest in the rat cortex, hippocampus, and cerebellum (Figure 5A; Supplementary Figure S3B). At P0 and P10, the highest TCF4 protein expression levels were seen in the cerebral cortex, hippocampus, and cerebellum (Figures 5B,C). The

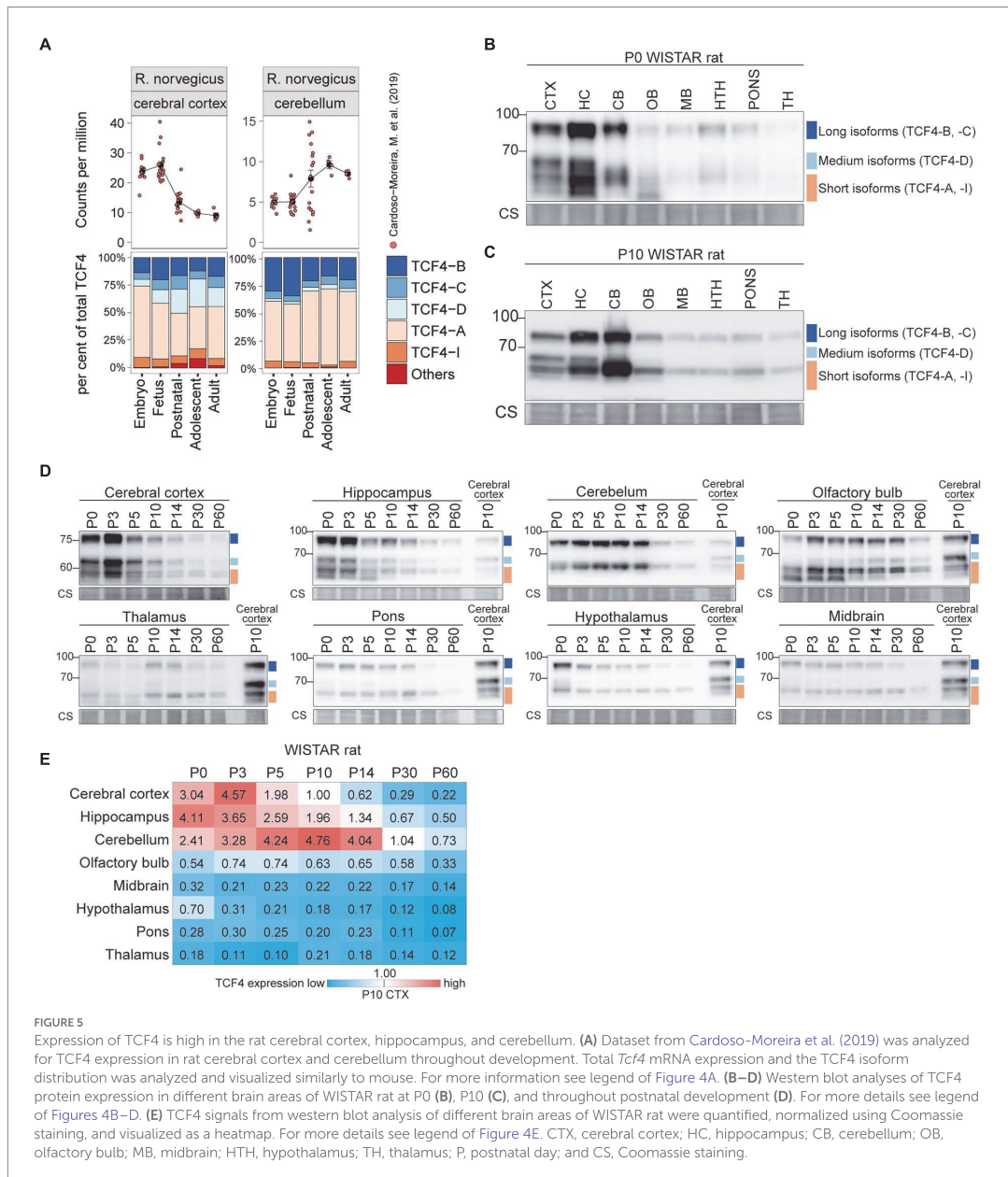


FIGURE 5

Expression of TCF4 is high in the rat cerebral cortex, hippocampus, and cerebellum. (A) Dataset from Cardoso-Moreira et al. (2019) was analyzed for TCF4 expression in rat cerebral cortex and cerebellum throughout development. Total *Tcf4* mRNA expression and the TCF4 isoform distribution was analyzed and visualized similarly to mouse. For more information see legend of Figure 4A. (B–D) Western blot analyses of TCF4 protein expression in different brain areas of WISTAR rat at P0 (B), P10 (C), and throughout postnatal development (D). For more details see legend of Figures 4B–D. (E) TCF4 signals from western blot analysis of different brain areas of WISTAR rat were quantified, normalized using Coomassie staining, and visualized as a heatmap. For more details see legend of Figure 4E. CTX, cerebral cortex; HC, hippocampus; CB, cerebellum; OB, olfactory bulb; MB, midbrain; HTH, hypothalamus; TH, thalamus; P, postnatal day; and CS, Coomassie staining.

expression dynamics of TCF4 protein isoforms across postnatal development were similar to mouse (Figures 5D,E). Notably, in P0-5 rat cerebral cortex, hippocampus, and olfactory bulb an isoform migrating faster than the TCF4-A isoforms was seen, potentially corresponding to TCF4-I (Figures 5B–D).

Taken together, our data showed that TCF4 expression pattern and dynamics were similar in the mouse and rat brain—the highest TCF4 protein expression was seen in the cerebral cortex and hippocampus around birth, and in the cerebellum 1–2 weeks after birth. TCF4 expression in other brain regions was relatively low. In addition to the differences in overall expression of TCF4, our data revealed that the composition of TCF4 isoforms expressed varies across brain regions in mouse and rat.

Expression of TCF4 is lower in rodent nonneural organs compared to the brain

We then focused on mouse nonneural organs and performed a meta-analysis of available short-read RNA-seq data (Keane et al., 2011; ENCODE Project Consortium, 2012; Schmitt et al., 2014; Yu et al., 2014; Vied et al., 2016; Li et al., 2017; Söllner et al., 2017; Cardoso-Moreira et al., 2019; Luo et al., 2020; Shafik et al., 2021). We analyzed *Tcf4* mRNA expression in the lung, kidney, thymus, spleen, liver, heart, and stomach (Figure 6A; Supplementary Figures S3A, S4A). These tissues displayed comparable *Tcf4* mRNA expression levels except for the liver, where almost no *Tcf4* mRNA expression was seen after birth (Figure 6A; Supplementary Figure S3A). Of the transcripts encoding different TCF4 protein isoforms in nonneural organs, the ones encoding TCF4-A were most prominently expressed, followed by TCF4-B-encoding transcripts (Figure 6A; Supplementary Figure S4A).

Next, we prepared protein lysates from BALB/c (Figure 6) and C57BL/6 (Supplementary Figure S5) mouse heart, diaphragm, muscle, skin, lung, kidney, bladder, stomach, pancreas, thymus, spleen, liver, and blood cells at P0, 14, and 60 for western blot analysis (Figures 6B,C, Supplementary Figure S5). In nonneural tissues, the composition of TCF4 protein isoforms was similar to the one in the brain—both long and short TCF4 protein isoforms were always present, whereas medium-sized TCF4 isoforms were not observed in any of the nonneural tissues (Figures 6B,C; Supplementary Figure S5D,E). Among the studied nonneural tissues, the highest levels of TCF4 protein were seen in the skin at P0 (Figure 6D; Supplementary Figure S5G). Very low TCF4 protein levels were detected in the pancreas, spleen, kidney and liver, and TCF4 protein expression was not seen in the blood cells (Figure 6D; Supplementary Figure S5G).

We also investigated TCF4 expression in rat nonneural tissues (Figure 7; Supplementary Figures S3B, S4B). Rat *Tcf4* mRNA expression was comparable in all the nonneural tissues except for the liver, where *Tcf4* expression was very low (Figure 7A; Supplementary Figure S3B). Different from mouse, transcripts encoding TCF4-A did not account for the majority of rat *Tcf4*

transcripts expressed in nonneural tissues, as also high expression of transcripts encoding TCF4-B and TCF4-C were present (Figure 7A; Supplementary Figure S4B).

Western blot analysis of rat nonneural tissues showed that different from mouse, TCF4 protein expression levels were more uniform between tissues (Figures 7B–D). In rat, TCF4 protein expression was highest in the thymus and was not observed in the pancreas (Figures 7B–D). The expression pattern of TCF4 isoforms in rat nonneural tissues was similar to mouse, i.e., mainly long and short TCF4 isoforms being present (Figures 7B,C).

Overall, the expression of TCF4 in the rodent nonneural tissues was much lower compared to the expression levels observed in the early postnatal development of the central nervous system. In addition, medium-sized TCF4 protein isoforms were almost non-existent in rodent nonneural tissues.

Expression of TCF4 in human tissues is highest around birth

Next, we analyzed available short-read RNA-seq data to describe *TCF4* total and isoform-specific mRNA expression in humans. We first analyzed the dataset published by Cardoso-Moreira and colleagues, which contained RNA-seq data from the human brain, heart, kidney, liver, and testis (Cardoso-Moreira et al., 2019). Of the noted tissues, the highest TCF4 mRNA expression was detected in the brain (Figure 8A). Human nonneural tissues showed detectable but lower *TCF4* mRNA levels compared to the brain, especially in the earlier stages of development (Figure 8A). Very low *TCF4* mRNA expression was noted for the liver (Figure 8A). In contrast to other tissues where *TCF4* mRNA levels were relatively stable throughout development, *TCF4* mRNA expression in the forebrain and kidney peaked during prenatal development (Figure 8A). We then used developmental transcriptome data from the BrainSpan project⁷ to describe the changes in total *TCF4* mRNA expression in different brain regions during human development (Supplementary Figure S6). Results were similar in all brain regions—*TCF4* mRNA expression peaked during embryonic development and decreased after birth (Supplementary Figure S6).

Next, we conducted a similar analysis for adult human RNA-seq data from the Genotype-Tissue Expression (GTEx) project.⁸ A selection of adult human tissues is shown in Figure 8B and all the studied tissues can be found in Supplementary Figure S7. When comparing different adult human tissues, the highest *TCF4* mRNA expression levels were seen in the adult human brain and adipose tissues (Figure 8B; Supplementary Figure S7). Almost no *TCF4* mRNA was detected in the human pancreas, liver, and whole blood (Figure 8B; Supplementary Figure S7).

7 <http://brainspan.org>

8 <https://gtexportal.org>

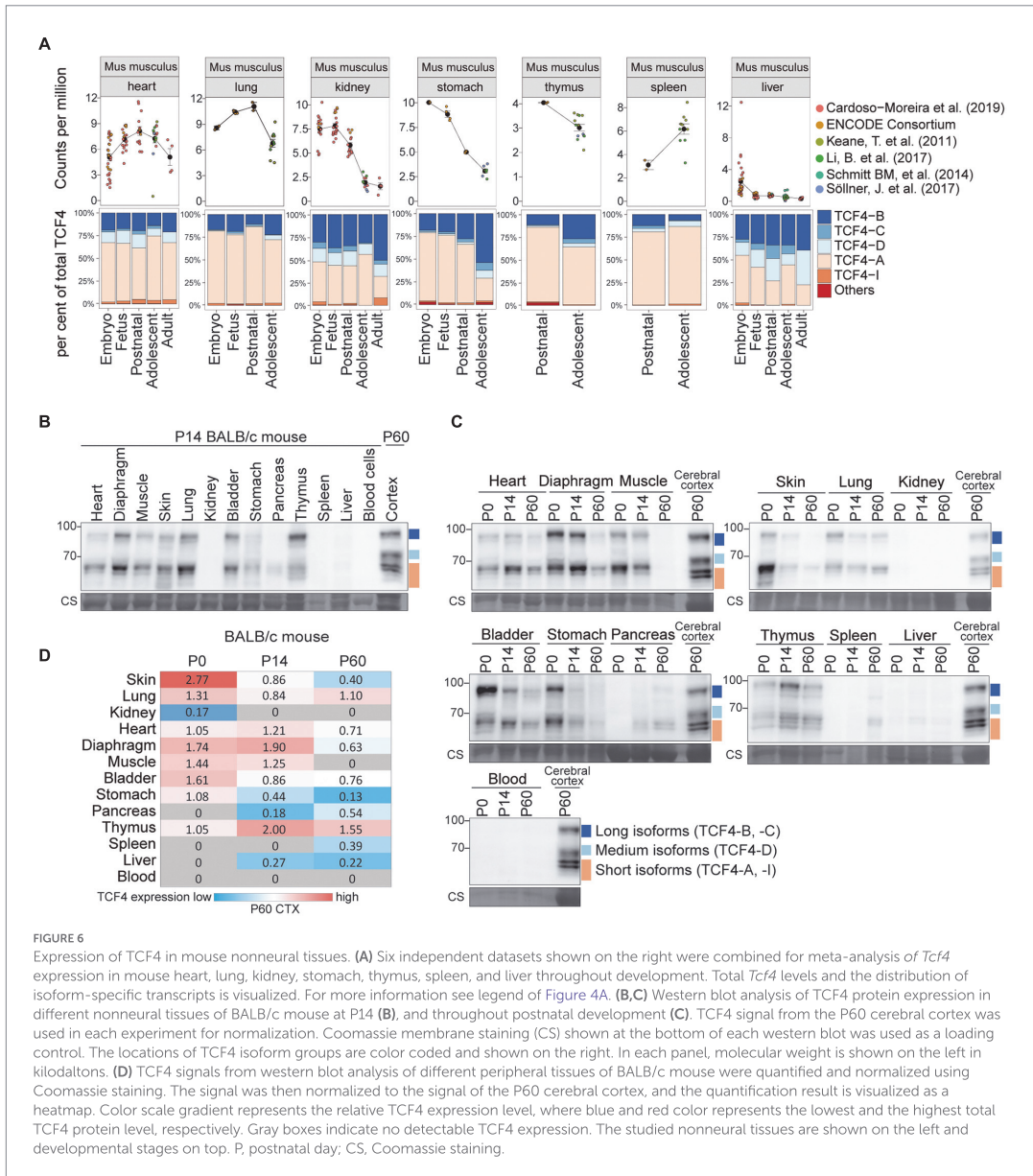


FIGURE 6

Expression of TCF4 in mouse nonneural tissues. (A) Six independent datasets shown on the right were combined for meta-analysis of *Tcf4* expression in mouse heart, lung, kidney, stomach, thymus, spleen, and liver throughout development. Total *Tcf4* levels and the distribution of isoform-specific transcripts is visualized. For more information see legend of Figure 4A. (B,C) Western blot analysis of TCF4 protein expression in different nonneural tissues of BALB/c mouse at P14 (B), and throughout postnatal development (C). TCF4 signal from the P60 cerebral cortex was used in each experiment for normalization. Coomassie membrane staining (CS) shown at the bottom of each western blot was used as a loading control. The locations of TCF4 isoform groups are color coded and shown on the right. In each panel, molecular weight is shown on the left in kilodaltons. (D) TCF4 signals from western blot analysis of different peripheral tissues of BALB/c mouse were quantified and normalized using Coomassie staining. The signal was then normalized to the signal of the P60 cerebral cortex, and the quantification result is visualized as a heatmap. Color scale gradient represents the relative TCF4 expression level, where blue and red color represents the lowest and the highest total TCF4 protein level, respectively. Gray boxes indicate no detectable TCF4 expression. The studied nonneural tissues are shown on the left and developmental stages on top. P, postnatal day; CS, Coomassie staining.

For both Cardoso-Moreira et al. and GTEx datasets, transcripts encoding TCF4-A made up around 50% of the total TCF4 mRNA levels in all the studied tissues, with the only exception being the testis (Figures 8A,B; Supplementary Figure S7). In the human testis, mRNA transcripts encoding TCF4-J accounted for the majority of total TCF4 levels beginning from adolescence, which coincides with the start of spermatogenesis (Figure 8A;

Supplementary Figure S6). The other major isoform-specific transcripts expressed in human tissues were TCF4-B, -C, and-D (Figures 8A,B; Supplementary Figure S6).

Next, we aimed to investigate TCF4 isoform composition in the adult human cerebral cortex and hippocampus. For this, we prepared protein lysates from these brain regions and human neuroblastoma cell line SH-SY5Y used for isoforms' mobility comparison. Western blot analysis revealed that TCF4 protein

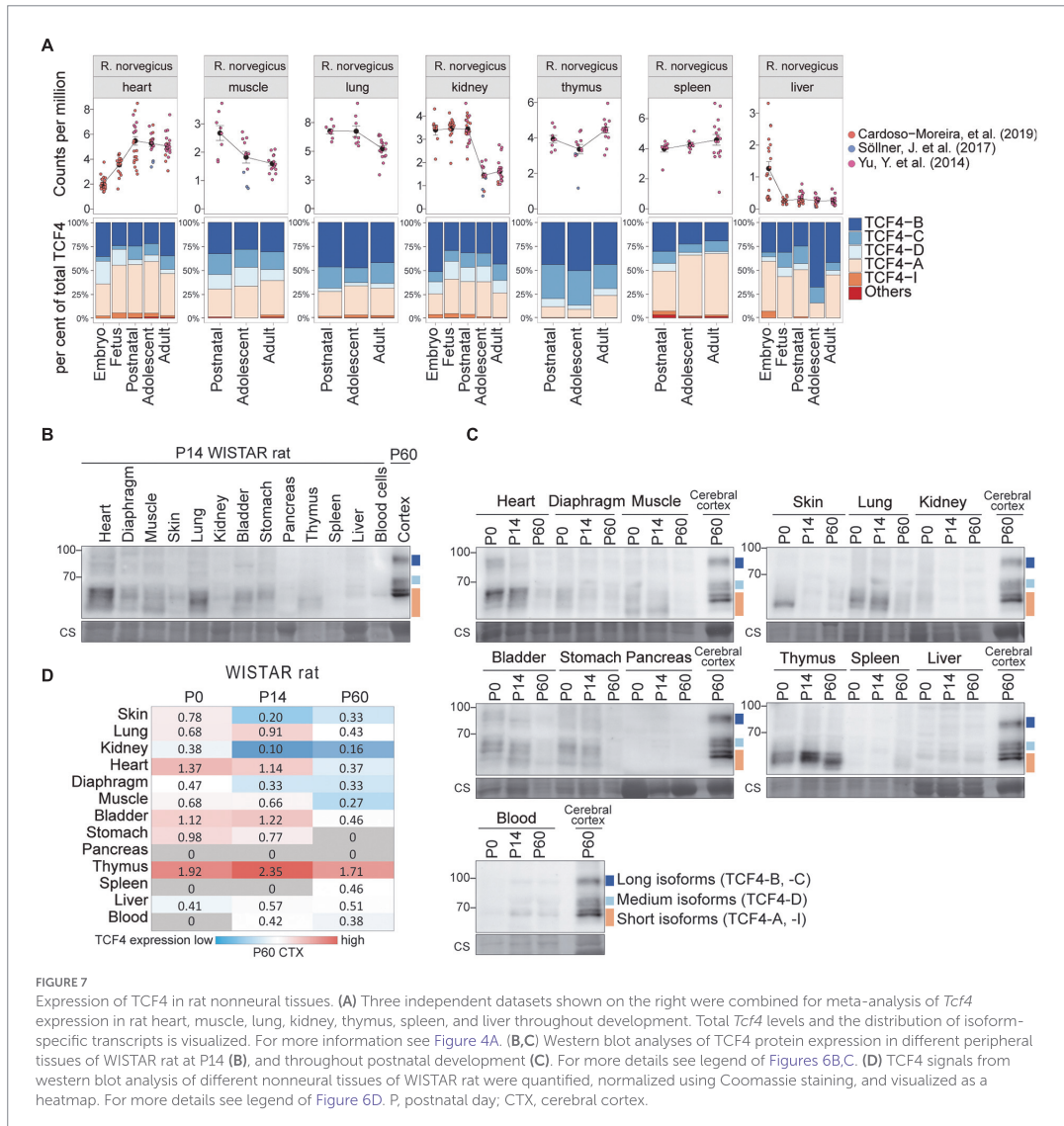


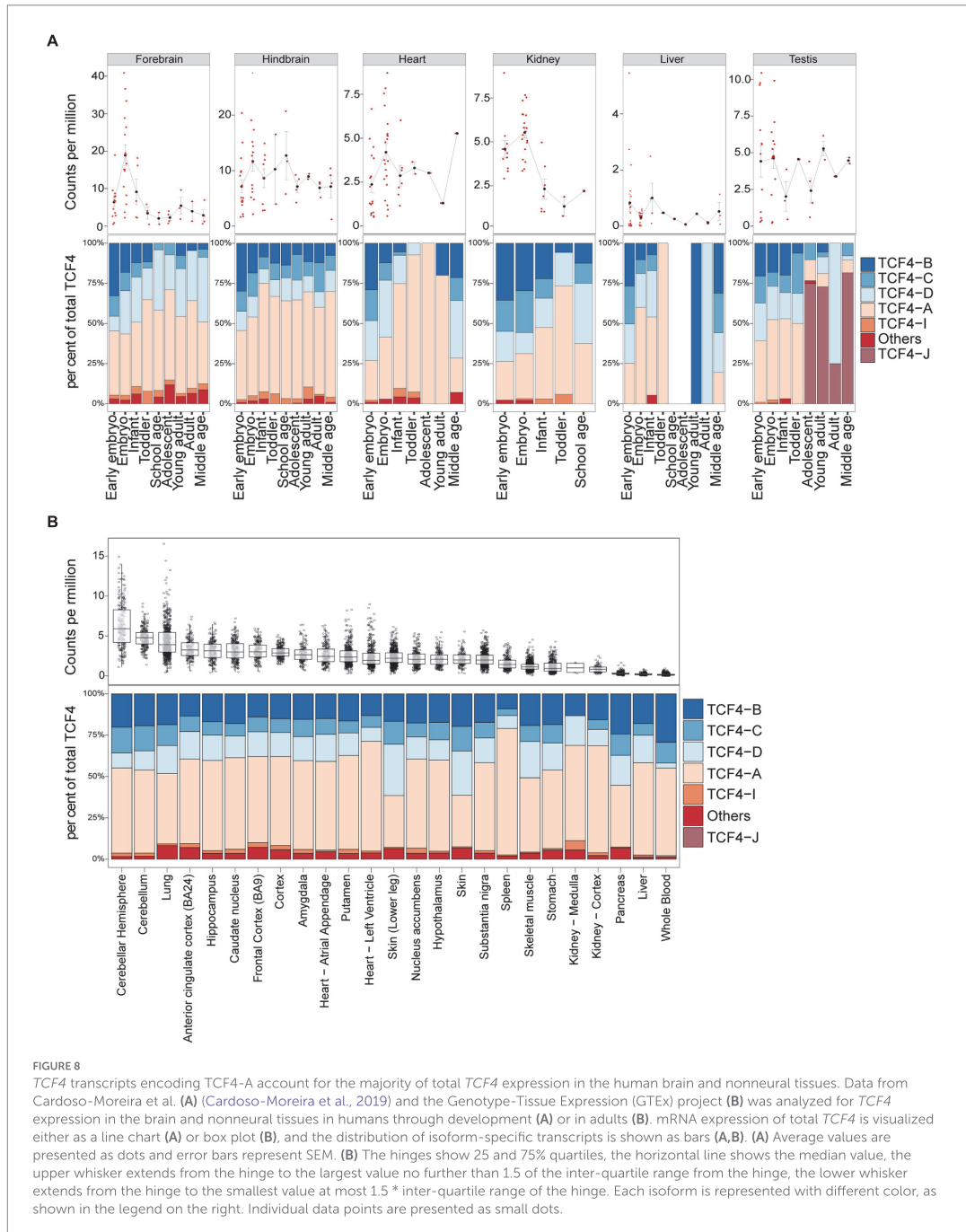
FIGURE 7

Expression of TCF4 in rat nonneural tissues. (A) Three independent datasets shown on the right were combined for meta-analysis of *Tcf4* expression in rat heart, muscle, lung, kidney, thymus, spleen, and liver throughout development. Total *Tcf4* levels and the distribution of isoform-specific transcripts is visualized. For more information see Figure 4A. (B,C) Western blot analyses of TCF4 protein expression in different peripheral tissues of Wistar rat at P14 (B), and throughout postnatal development (C). For more details see legend of Figures 6B,C. (D) TCF4 signals from western blot analysis of different nonneural tissues of Wistar rat were quantified, normalized using Coomassie staining, and visualized as a heatmap. For more details see legend of Figure 6D. P, postnatal day; CTX, cerebral cortex.

signal was detectable in both adult human brain and SH-SY5Y cell line (Figure 9). We also detected a possible non-specific signal located between the long and medium TCF4 isoforms in both SH-SY5Y and human brain lysates (Figure 9) since this signal was not detected using other TCF4 antibodies (data not shown) validated by us before (Nurm et al., 2021). Different to SH-SY5Y cell line, we detected all three TCF4 isoform groups in the adult human brain, however expression level of longer TCF4 isoforms was higher compared to the medium and short isoforms (Figure 9; Supplementary Figure S8). This result was on the contrary with protein isoform patterns seen in the rodent brain

and the results from our human RNA-seq data analysis, which could result from protein stability, post-mortem artifacts or signal masking by other similar-sized proteins. Nevertheless, the presence of long, medium and short TCF4 isoforms in adult human brain matched with TCF4 isoform pattern in rodents, however the species-and tissue-specific temporal expression dynamics of different TCF4 isoforms during the development cannot be emphasized more.

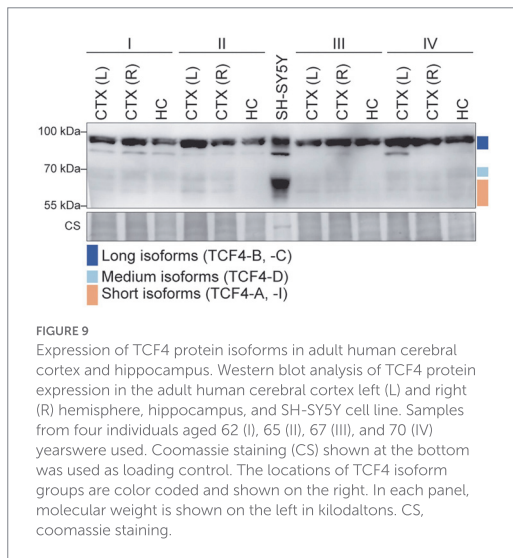
Altogether, our results show that *TCF4* mRNA is expressed at high levels in the human brain during development and the expression is retained in the adulthood. In most tissues transcripts



encoding for TCF4-A were the most prominent ones, while in the testis TCF4-J encoding transcripts were mostly expressed. In adult human brain, long, medium and short TCF4 protein isoforms are expressed.

Discussion

Transcription factor TCF4 has been extensively studied due to its linkage with neurocognitive disorders such as intellectual



disability, schizophrenia and Pitt-Hopkins syndrome (Stefansson et al., 2009; Kharbanda et al., 2016; Zollino et al., 2019). Knowledge of TCF4 expression across tissues and development would lay the foundation to understanding how these diseases develop and may help with the generation of gene therapy applications for the many TCF4 associated diseases. Transcripts from the mouse and human *TCF4* gene have been previously described in our lab using mRNA and expressed sequence tag data from public databases (Sepp et al., 2011; Nurm et al., 2021). Short read RNA-seq data can also be used to describe *Tcf4* transcripts. However, due to the structure of the *Tcf4* gene, it can be complicated to describe expression of transcripts encoding different isoforms based only on short read RNA-seq data as splicing features or 5' exons can be difficult to detect. Our direct long-read RNA-seq analysis of *Tcf4* transcripts in the rodent brain revealed that transcription from the *Tcf4* gene results in transcripts encoding 5N-terminally distinct TCF4 protein isoforms in the rodent brain – TCF4-B, -C, -D, -A, and -I. This result falls in line with previous observations by Nurm et al. (2021).

Expression of total *Tcf4* mRNA during development has been extensively studied mainly in the mouse cerebral cortex at the total mRNA level, with the highest expression reported around birth (E16-P6; Li et al., 2019; Phan et al., 2020). This is in accordance with our RNA-seq meta-analysis and applies for both mouse and rat. In addition, we show that the expression dynamics of *Tcf4* in the rodent brain and nonneural tissues are similar—highest *Tcf4* expression can be detected around birth, followed by a decline during postnatal development. However, studying only total *Tcf4* mRNA levels provides only partial information about *Tcf4* expression as transcription from the *Tcf4* gene results in numerous transcripts, and encoded protein isoforms have different functional protein domains and transactivation capability (Sepp

et al., 2011, 2017; Nurm et al., 2021). We have previously developed a method which quantifies the expression of different *Tcf4* protein isoform-encoding transcripts using short read RNA-seq data (Sirp et al., 2020). Here, we applied the same approach to describe the expression of different TCF4 isoforms throughout development using previously published RNA-seq data. When leaving aside the great increase in the expression of TCF4-J in the adolescent human testis (transcripts encoding TCF4-J are not present in rodents), no drastic change concerning switching from the expression of one TCF4 isoform to the other was detected during the rodent and human development. This suggests that the same TCF4 isoforms that are necessary during development may also be vital for the TCF4-mediated normal functioning of the adult organism.

The necessity of so many different TCF4 isoforms remains unknown. In humans, mutations in the 5' region of *TCF4* gene, which affect only the longer isoforms, lead to mild-moderate intellectual disability (Kharbanda et al., 2016). As the resulting disease is not as severe as the Pitt-Hopkins syndrome, it may mean that a slight decrease in overall TCF4 expression causes the phenotype. However, it is also possible that the longer TCF4 isoforms have specific functions which cannot be compensated by other TCF4 isoforms, and mutations affecting only a subset of TCF4 isoforms result in less severe effects than seen for mutations affecting all the isoforms. Recently, it has been shown that postnatal normalization of TCF4 expression to wild type levels can rescue the phenotype of TCF4 heterozygous knockout mice (Kim et al., 2022). In addition, studies of Daughterless, the orthologue of TCF4 in the fruit fly, have shown that it is possible to partially rescue the severe embryonic neuronal phenotype of Daughterless null mutation by overexpressing either human TCF4-A or TCF4-B (Tamberg et al., 2015). The generation of TCF4 isoform-specific mutant mice would help to identify whether TCF4 isoforms have distinct or similar functions. Such a model could be used to determine whether it is possible to rescue the negative phenotype resulting from a knock-out of a single TCF4 isoform by increasing the level of another TCF4 isoform. However, generating such a model comes with many challenges. To begin with, it can be complicated to silence all the TCF4 isoforms individually by causing just frameshift mutations as only some isoforms (e.g., TCF4-B and -A) have their translation start sites located in independent 5' exons. In addition, mutating one *TCF4* transcript can result in the upregulation of another *TCF4* transcript – an effect that we saw when silencing TCF4-A in Neuro2a cells that resulted in an increase in TCF4-I levels. We have also previously shown that Fuchs' Endothelial Dystrophy-related endogenous downregulation of transcripts encoding longer TCF4 isoforms results in the upregulation of shorter isoforms (Sirp et al., 2020).

To fully characterize TCF4 expression, it is important to consider all TCF4 protein isoforms. Previously, a large study on the expression of TCF4 protein during neurodevelopment has been performed by Matthias Jung and colleagues using immunohistochemical analysis with an antibody specific only for the long TCF4 protein isoforms (Jung et al., 2018). Another study

by Kim and colleagues used TCF4-GFP reporter mice to characterize total TCF4 expression in the mouse brain (Kim et al., 2020). A major limitation of these methods is that they cannot be used to describe the expression of different TCF4 protein isoforms. Overall, our results of total TCF4 protein expression levels during postnatal development of different mouse brain areas agree with the previously reported data. However, by using a TCF4 antibody specific for all the TCF4 isoforms in western blot analysis, we were able to distinguish TCF4 protein expression in three different groups – long (TCF4-B, TCF4-C), medium (TCF4-D) and short TCF4 isoforms (TCF4-A, TCF4-I). Isoform-specific silencing of TCF4 and *in vitro* translated TCF4 protein isoforms confirmed the locations of TCF4 isoforms in western blot. However, it should be noted that a similar pattern of TCF4 isoforms in western blot analysis between different tissues may not necessarily indicate the presence of exactly the same TCF4 isoforms. We acknowledge that *in vitro* and *in vivo* translated proteins can migrate differently in western blot analysis due to the differences in post-translational modifications of the proteins in various cell types.

The expression dynamics of TCF4 during the development varied in different brain regions. In contrast to the cerebral cortex where TCF4 expression levels decline after birth, in the cerebellum, hippocampus and olfactory bulb we saw a more prolonged high TCF4 protein expression. In the cerebellum TCF4 protein expression peaks about a week later than in any of the other brain regions. While the majority of the neurogenesis in the central nervous system happens during prenatal development, the granule cell precursors of the cerebellum and olfactory bulb, and the dentate gyrus of the hippocampus proliferate and differentiate after birth (Chen et al., 2017), where TCF4 was shown to be highly expressed (Jung et al., 2018; Kim et al., 2020), and regulate the maturation of the cerebellar granule cells (Kim et al., 2020). We propose that high TCF4 expression is necessary for the maturation of distinct brain regions, whereas fully developed brain areas display low and stable TCF4 expression necessary for normal function of the adult nervous system.

Expression of long and short TCF4 protein isoforms was seen in all brain regions and nonneural tissues where TCF4 was detectable. However, in rodents the medium TCF4 isoforms (TCF4-D) were only observed in the brain, specifically in the cerebral cortex, hippocampus, and olfactory bulb. Interestingly, in the whole rodent brain the expression of medium isoforms became apparent only in later stages of embryonic development. The only well-known functional protein domain located in the N-terminal region of longer TCF4 isoforms is activation domain 1. While the long TCF4 isoforms (TCF4-B and -C) contain this domain, TCF4-D lacks it. In addition, different from short TCF4 isoforms (TCF4-A and -I), TCF4-D contains a nuclear localisation signal. It remains to be studied what the function of TCF4-D in the development of the nervous system is and why this TCF4 isoform is missing in the cerebellum where TCF4 is otherwise highly expressed.

Based on the results of the present study we propose that a mixture consisting of TCF4-B, -C, -D, -A, and -I encoding constructs could be used in gene therapy approaches for

Pitt-Hopkins syndrome. However, it should be noted that TCF4 expression levels vary between brain regions and cell types during development (Jung et al., 2018; Kim et al., 2020), suggesting that the dosage of TCF4 isoforms needs to be highly regulated. The direct administration of a cocktail of TCF4 isoforms may allow easier control of each isoform compared to other gene therapy approaches such as activation of endogenous promoters and enhancers. As a next step of this study, a similar TCF4 protein expression analysis should be done for human brain regions with a focus on the hippocampus and cerebral cortex, as studies of structural brain anomalies in PTHS-patients and *Tcf4*-heterozygous mice have shown hypoplasia of these brain regions (Marangi and Zollino, 2015). In addition, the expression of TCF4 different transcripts and the protein isoforms they encode should be studied at the single cell level to better understand how the many TCF4 isoforms are regulated between cell types.

Data availability statement

The datasets analyzed and presented in this study can be found in online repositories and in the [Supplementary material](#). The names of the repository/repositories and accession number(s) can be found in the article.

Ethics statement

The studies involving human participants were reviewed and approved by Tallinn Committee for Medical Studies, National Institute for Health Development (Permit Number 402). The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Ministry of Agriculture of Estonia (Permit Number: 45).

Author contributions

ASi and JT designed research, performed research, analyzed data, and wrote the paper. ASh, LT, and CK performed research, analyzed data, and wrote the paper. LK performed research. TT designed research, wrote the paper, and acquired funding. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Estonian Research Council (grant PRG805), European Union through the European Regional Development Fund (project no. 2014-2020.4.01.15-0012) and H2020-MSCA-RISE-2016 (EU734791), and Pitt Hopkins Research Foundation (grant no. 21). The funding sources were not

involved in study design, analysis and interpretation of data, writing of the report and in the decision to submit the article for publication.

Acknowledgments

We thank Epp Väli for technical assistance, Enn Jõeeste for the dissection of human tissue samples and Mari Sepp for critical reading of the manuscript. We would like to acknowledge the ENCODE Consortium and the ENCODE production laboratories for generating the datasets used in this study. We 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.

Conflict of interest

JT and TT were employed by Protobios LLC.

References

- Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., et al. (2007). Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* 80, 988–993. doi: 10.1086/515582
- Badowska, D. M., Brzózka, M. M., Kannaiyan, N., Thomas, C., Dibaj, P., Chowdhury, A., et al. (2020). Modulation of cognition and neuronal plasticity in gain- and loss-of-function mouse models of the schizophrenia risk gene Tcf4. *Transl. Psychiatry* 10:343. doi: 10.1038/s41398-020-01026-7
- 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
- Cardoso-Moreira, M., Halbert, J., Vallotton, D., Velten, B., Chen, C., Shao, Y., et al. (2019). Gene expression across mammalian organ development. *Nature* 571, 505–509. doi: 10.1038/s41586-019-1338-5
- Champlin, A. K., Dorr, D. L., and Gates, A. H. (1973). Determining the stage of the estrous cycle in the mouse by the appearance of the vagina. *Biol. Reprod.* 8, 491–494. doi: 10.1093/biolreprod/8.4.491
- Chen, V. S., Morrison, J. P., Southwell, M. F., Foley, J. F., Bolon, B., and Elmore, S. A. (2017). Histology atlas of the developing prenatal and postnatal mouse central nervous system, with emphasis on prenatal days E7.5 to E18.5. *Toxicol. Pathol.* 45, 705–744. doi: 10.1177/0192623317728134
- Chiaromello, A., Neuman, K., Palm, K., Metsis, M., and Neuman, T. (1995). Helix-loop-helix transcription factors mediate activation and repression of the p75LNGFR gene. *Mol. Cell. Biol.* 15, 6036–6044. doi: 10.1128/MCB.15.11.6036
- Corneliusson, 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. doi: 10.1128/jvi.65.11.6084-6093.1991
- de Pontual, L., 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
- 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:eau4139. doi: 10.1126/sciadv.aau4139
- Einarson, M. B., and Chao, M. V. (1995). Regulation of Id1 and its association with basic helix-loop-helix proteins during nerve growth factor-induced differentiation of PC12 cells. *Mol. Cell. Biol.* 15, 4175–4183. doi: 10.1128/MCB.15.8.4175
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- 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 Dev.* 9:23. doi: 10.1186/1749-8104-9-23
- 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
- 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
- Jung, M., Häberle, B. M., Tschakowsky, 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
- Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289–294. doi: 10.1038/nature10413
- 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
- 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, H., Gao, E. B., Draper, A., Berens, N. C., Vihma, H., Zhang, X., et al. (2022). Rescue of behavioral and electrophysiological phenotypes in a Pitt-Hopkins syndrome mouse model by genetic restoration of Tcf4 expression. *elife* 11:e72290. doi: 10.7554/eLife.72290

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnmol.2022.1033224/full#supplementary-material>

- Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., et al. (2017). A comprehensive mouse Transcriptomic BodyMap across 17 tissues by RNA-seq. *Sci. Rep.* 7:4200. doi: 10.1038/s41598-017-04520-z
- 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* 24, 1235–1246. doi: 10.1038/s41380-019-0353-0
- Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., et al. (2020). New developments on the encyclopedia of DNA elements (ENCODE) data portal. *Nucleic Acids Res.* 48, D882–D889. doi: 10.1093/nar/gkz1062
- 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* 8:67. doi: 10.1038/s41398-018-0114-x
- Marangi, G., and Zollino, M. (2015). Pitt–Hopkins syndrome and differential diagnosis: a molecular and clinical challenge. *J. Pediatr. Genet.* 4, 168–176. doi: 10.1055/s-0035-1564570
- Massari, M. E., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* 20, 429–440. doi: 10.1128/MCB.20.2.429-440.2000
- 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
- Persson, P., Jögi, A., Grynfeld, A., Pählman, S., and Axelsson, H. (2000). HASH-1 and E2-2 are expressed in human neuroblastoma cells and form a functional complex. *Biochem. Biophys. Res. Commun.* 274, 22–31. doi: 10.1006/bbrc.2000.3090
- Phan, B. N., Bohlen, J. E., 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
- 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. doi: 10.1002/j.1460-2075.1996.tb01058.x
- Quednow, B. B., Ettinger, U., Mössner, R., Rujescu, D., Giegling, I., Collier, D. A., et al. (2011). The schizophrenia risk allele C of the TCF4 rs9960767 polymorphism disrupts sensorimotor gating in schizophrenia Spectrum and healthy volunteers. *J. Neurosci.* 31, 6684–6691. doi: 10.1523/JNEUROSCI.0526-11.2011
- 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., Neale, B. M., Corvin, A., Walters, J. T., Farh, K.-H., Holmans, P. A., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. doi: 10.1038/nature13595
- 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, 494–411. doi: 10.1038/s41398-021-01618-x
- Schmitt, B. M., Rudolph, K. L. M., Karagianni, P., Fonseca, N. A., White, R. J., Taliandis, I., et al. (2014). High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface. *Genome Res.* 24, 1797–1807. doi: 10.1101/gr.176784.114
- 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/ddl112
- 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
- Shafik, A. M., Zhang, F., Guo, Z., Dai, Q., Pajdzik, K., Li, Y., et al. (2021). N6-methyladenosine dynamics in neurodevelopment and aging, and its potential role in Alzheimer's disease. *Genome Biol.* 22:17. doi: 10.1186/s13059-020-02249-z
- Sirp, A., Leite, K., Tuvikene, J., Nurm, K., Sepp, M., and Timmusk, T. (2020). The Fuchs corneal dystrophy-associated CTG repeat expansion in the TCF4 gene affects transcription from its alternative promoters. *Sci. Rep.* 10:18424. doi: 10.1038/s41598-020-75437-3
- 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
- Söllner, J. F., Lepar, G., Hildebrandt, T., Klein, H., Thomas, L., Stupka, E., et al. (2017). An RNA-Seq atlas of gene expression in mouse and rat normal tissues. *Sci. Data* 4:170185. doi: 10.1038/sdata.2017.185
- Soosaar, A., Chiamello, 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
- 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
- Tamberg, L., Jaago, M., Säälk, K., Sirp, A., Tuvikene, J., Shubina, A., et al. (2020). Daughterless, the drosophila orthologue of TCF4, is required for associative learning and maintenance of the synaptic proteome. *Dis. Model. Mech.* 13:dmm042747. doi: 10.1242/dmm.042747
- 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
- 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
- Vied, C., Ray, S., Badger, C.-D., Bundy, J. L., Arbeitman, M. N., and Nowakowski, R. S. (2016). Transcriptomic analysis of the hippocampus from six inbred strains of mice suggests a basis for sex-specific susceptibility and severity of neurological disorders. *J. Comp. Neurol.* 524, 2696–2710. doi: 10.1002/cne.23989
- 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, 668–681. doi: 10.1038/s41588-018-0090-3
- Yu, Y., Fuscoe, J. C., Zhao, C., Guo, C., Jia, M., Qing, T., et al. (2014). A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat. Commun.* 5:3230. doi: 10.1038/ncomms4230
- 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. doi: 10.1128/MCB.16.6.2898
- 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., 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

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



OPEN

The Fuchs corneal dystrophy-associated CTG repeat expansion in the *TCF4* gene affects transcription from its alternative promoters

Alex Sirp^{1,5}, Kristian Leite^{1,3,5}, Jürgen Tuvikene^{1,2,5}, Kaja Nurm¹, Mari Sepp^{1,4} & Tõnis Timmusk^{1,2}✉

The CTG trinucleotide repeat (TNR) expansion in *Transcription factor 4* (*TCF4*) intron 3 is the main cause of Fuchs' endothelial corneal dystrophy (FECD) and may confer an increased risk of developing bipolar disorder (BD). Usage of alternative 5' exons for transcribing the human *TCF4* gene results in numerous *TCF4* transcripts which encode for at least 18 N-terminally different protein isoforms that vary in their function and transactivation capability. Here we studied the *TCF4* region containing the CTG TNR and characterized the transcription initiation sites of the nearby downstream 5' exons 4a, 4b and 4c. We demonstrate that these exons are linked to alternative promoters and show that the CTG TNR expansion decreases the activity of the nearby downstream *TCF4* promoters in primary cultured neurons. We confirm this finding using two RNA sequencing (RNA-seq) datasets of corneal endothelium from FECD patients with expanded CTG TNR in the *TCF4* gene. Furthermore, we report an increase in the expression of various other *TCF4* transcripts in FECD, possibly indicating a compensatory mechanism. We conclude that the CTG TNR affects *TCF4* expression in a transcript-specific manner both in neurons and in the cornea.

Transcription factor 4 (*TCF4*) is a basic helix-loop-helix transcription factor that plays a vital role in the development of the nervous and immune system^{1–3}. *TCF4* is expressed in almost every tissue type in human^{4,5}. In the brain *TCF4* expression peaks during late embryonic development and continues at a relatively high level during postnatal development^{6,7}. Notably, transcription from *TCF4* gene can begin at multiple mutually exclusive 5' exons leading to transcripts with varying composition of functional protein domains which modulate the ability of *TCF4* to regulate transcription^{4,8}.

TCF4 gene is implicated in susceptibility to schizophrenia, and mutations in *TCF4* cause Pitt-Hopkins syndrome, a rare developmental disorder characterized by severe motor and mental retardation^{6,9–11}. In addition, mutations in the gene regions included only in the longer isoforms of *TCF4* have been associated with intellectual disability¹². Alterations in *TCF4* expression levels have been described in patients with depression¹³. An expansion of a CTG TNR, located in an alternative promoter region between *TCF4* exons 3 and 4 (known as CTG 18.1), upstream of *TCF4* 5' exons 4a, 4b and 4c causes Fuchs' endothelial corneal dystrophy (FECD) and has been linked to Bipolar disorder (BD)^{14,15}.

FECD is an ocular disorder associated with corneal edema and vision disruption having a varying prevalence between different populations affecting about 4% of people over 40 in the United States¹⁶. Out of the many genetic mutations associated with FECD the CTG TNR expansion in *TCF4* is thought to be one of the major factors in the development of the disease, as the presence of a *TCF4* allele with 50 or more CTG TNR-s confers an increased risk of developing the disease¹⁷. BD is a psychiatric disorder that affects up to 1% of the global population, causing severe mood alterations in affected patients¹⁸. It has been shown that a *TCF4* CTG TNR expansion of over 40

¹Department of Chemistry and Biotechnology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia. ²Protobios LLC, 12618 Tallinn, Estonia. ³Present address: Department of Neurology, University Medicine Göttingen, Waldweg 33, 37073 Göttingen, Germany. ⁴Present address: Center for Molecular Biology of Heidelberg University (ZMBH), 69120 Heidelberg, Germany. ⁵These authors contributed equally: Alex Sirp, Kristian Leite and Jürgen Tuvikene. ✉email: tonis.timmusk@taltech.ee

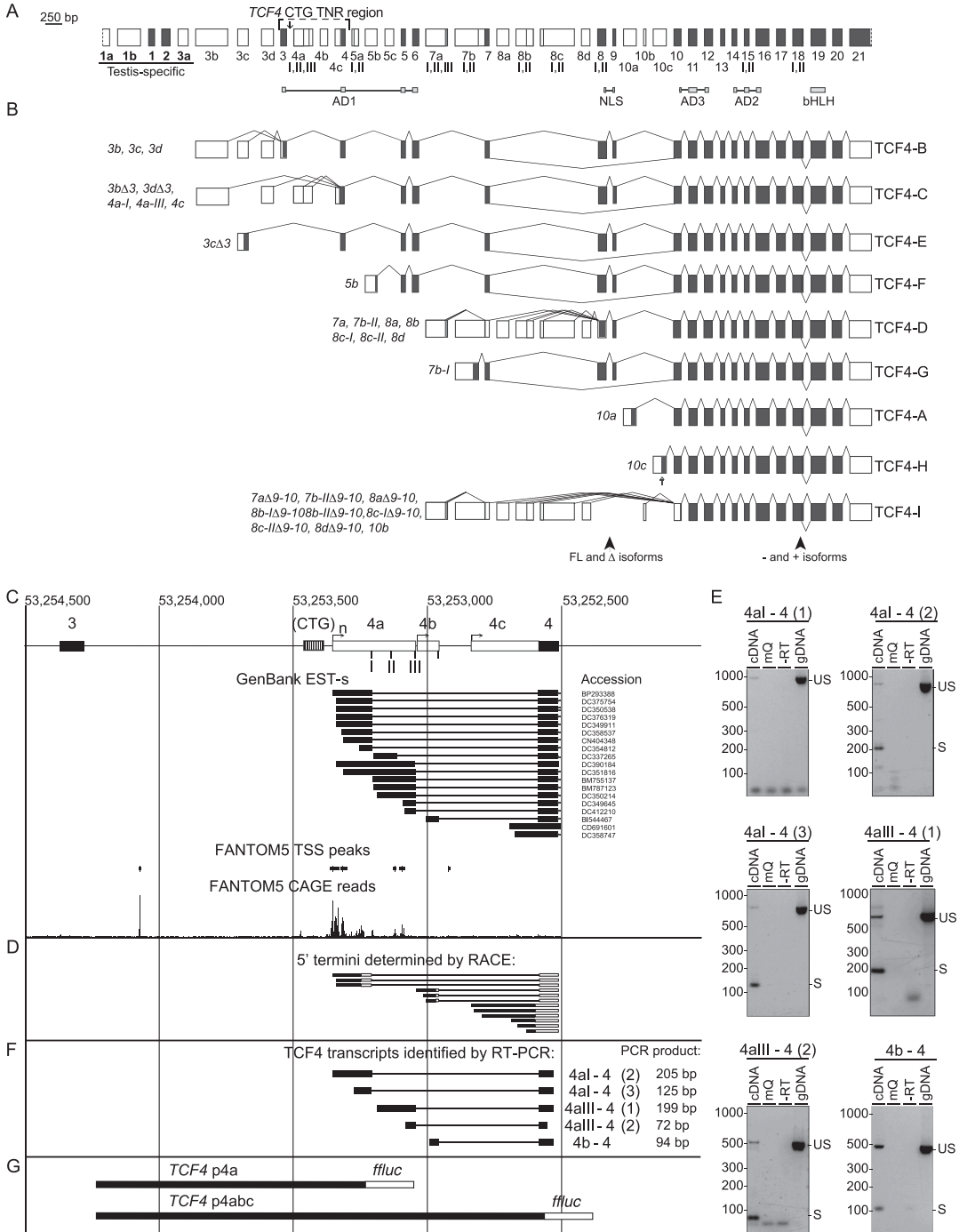


Figure 1. Transcription start sites, cloned insert sequences and analysis of splice sites in the *TCF4* CTG repeat-containing region in the adult human cerebellum. **(A)** A schematic of the *TCF4* gene including alternative 5' exons marked as white boxes and internal exons marked as black boxes drawn in scale. Functional protein domains have been marked under the exons and the *TCF4* CTG TNR region has been marked on top with an arrow displaying the repeat location in the intron between exons 3 and 4a. Alternative splice sites have been marked as roman numerals. Testis specific exons are underlined and marked bold. **(B)** Major alternative transcripts of the human *TCF4* gene. Coding regions are represented as black boxes and untranslated regions are shown as white boxes. Transcripts are named after the 5' exon and the splice site. The names of protein isoforms encoded by the transcripts are shown on the right. Locations of alternative splicing that generates full-length (FL), Δ , - and + isoforms are shown at the bottom. A and B adapted from Sepp et al.⁴. **(C)** A schematic drawing of the *TCF4* gene region proximal to the CTG TNR. Genomic coordinates are based on the GRCh37/hg19 genome build. Black boxes indicate internal exons, white boxes 5' exons and striped open box marks the location of the CTG TNR. EST-s from GenBank are shown with accession numbers on the right. TSS peaks (FANTOM5 DPI peak, robust set) and CAGE reads (total counts of CAGE for reverse strand encoding for *TCF4*) from FANTOM5 project are visualized. **(D)** Transcripts identified by 5' RACE are indicated—black boxes show sequenced components, white boxes unsequenced components. Transcription start regions of exon 4a, exon 4b and exon 4c are indicated by arrows. **(E)** RT-PCR analysis of the splicing in the region. “US” indicates RT-PCR fragments amplified from gDNA or unspliced pre-mRNA. “S” denotes RT-PCR fragments from spliced mRNA. The numbers in brackets indicate different primer pairs used for PCR. Three different primer pairs were used to identify splice sites between exons 4a-I and 4, whereas two were used to identify splice sites between exons 4a-III and 4. **(F)** The structure of *TCF4* transcripts identified by RT-PCR in D. **(G)** The cloned promoter regions are indicated with filled boxes and the firefly luciferase (*flLuc*) reporter gene are indicated with an open box. *AD* activation domain, *NLS* nuclear localization domain, *bHLH* basic helix-loop-helix, *FL* full length, Δ lack of nuclear localization domain, *EST* expressed sequence tag, *TSS* transcription start site, *FANTOM5* functional annotation of the mammalian genome, *DPI* decomposition-based peak identification, *CAGE* cap analysis of gene expression.

CTG TNR-s is frequent in a subset of patients with a severe type of BD and that the repeat expansion in *TCF4* may increase vulnerability to BD¹⁵.

Studies on the connection between the *TCF4* CTG TNR expansion and the mRNA levels of *TCF4* transcripts spanning the *TCF4* CTG TNR region have produced contradictory results. A study by Foja et al.¹⁹ reported that *TCF4* CTG TNR expansion is connected with a reduction in the levels of *TCF4* transcripts beginning in the proximity of the CTG TNR, whereas a study by Okumura et al.²⁰ has reported that *TCF4* CTG TNR expansion is connected with an increase in overall *TCF4* levels and in the levels of *TCF4* transcripts beginning in proximity of the CTG TNR. Two other studies^{21,22} have reported no effect of the CTG TNR expansion on total *TCF4* mRNA levels.

It is currently unknown whether the *TCF4* CTG TNR expansion affects the levels of *TCF4* transcripts and total *TCF4* levels. Here, we hypothesized that the *TCF4* CTG TNR region contains functional promoters that regulate the transcription of nearby 5' exons and that the activity of these promoters is altered by the length of *TCF4* CTG TNR expansion. For this, we first characterized the *TCF4* alternative promoter region containing the CTG TNR by identifying transcription start sites (TSS) and describing the splicing of nearby 5' exons 4a, 4b, and 4c. We then used luciferase reporter assay to investigate whether the CTG TNR expansion could influence *TCF4* expression in primary neurons by affecting the ability of surrounding regulatory regions to promote transcription. Furthermore, RNA-seq data from corneal tissue of FECD patients with an expanded *TCF4* CTG TNR was analyzed to determine the expression levels of *TCF4* transcripts beginning both proximal and distal to the CTG TNR. Collectively, our results demonstrate that the CTG TNR expansion differentially modulates the activity of *TCF4* promoters.

Results

Transcription start and splice donor site usage of *TCF4* 5' exons in vicinity of the CTG TNR. The CTG TNR immediately precedes *TCF4* 5' exons 4a, 4b and 4c, which are located between internal exons 3 and 4 (Fig. 1A). The major *TCF4* transcripts transcribed from the alternative 5' exons of *TCF4* in proximity of the CTG TNR are transcripts encoding for protein isoforms TCF4-B and TCF4-C (Fig. 1B). To characterize the transcription start sites (TSS-s) in this region we performed bioinformatical and 5' RACE analysis. Analysis of GenBank data revealed that a total of 19 expressed sequence tags (EST-s) with 17 different TSS-s can be found between *TCF4* internal exons 3 and 4 with none beginning downstream of exon 3 and upstream of the CTG TNR (Fig. 1C). In addition, analysis of TSS peak data from the FANTOM5 (Functional Annotation of the Mammalian Genome) project revealed 6 TSS peaks near the CTG TNR: 1 TSS peak before the CTG TNR and 5 TSS peaks (of which 3 TSS peaks match with TSS-s from GenBank EST-s) downstream of the repeat (Fig. 1C). To validate the potential transcripts and TSS-s from our bioinformatical analysis, we performed 5' RACE from adult human cerebellar RNA, as it exhibits high levels of *TCF4* expression²³. Our 5' RACE analysis revealed twelve TSS-s—three for exon 4a, three for exon 4b and six for exon 4c, distributed across a ~250 bp region (Fig. 1D). Out of the 12 TSS-s detected by 5' RACE only 4 matched with the TSS-s from our bioinformatical analysis. Importantly, the CTG TNR was never present in the 5' UTR of exon 4a, 4b and 4c transcripts since EST-s from GenBank and our 5' RACE revealed no TSS directly upstream of the *TCF4* CTG TNR region (Fig. 1D, Supplementary Fig. S1). When considering previously published data about the TSS-s of *TCF4* exon 4a, 4b and 4c and data obtained from our 5' RACE analysis, the promoters in the *TCF4* CTG TNR region appear to be dispersed promoters,

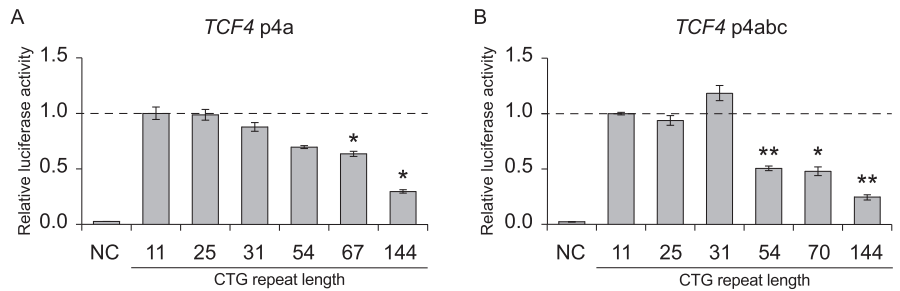


Figure 2. The *TCF4* CTG trinucleotide repeat region modulates the activity of proximal downstream *TCF4* promoters in a length dependent manner in neurons. (A,B) Promoter activities of *TCF4* p4a (A) and *TCF4* p4abc (B) constructs with different CTG repeat sizes (11, 25, 31, 54, 67 or 144). Promoter activity was measured in rat cultured cortical neurons transfected with luciferase reporter constructs. Three independent experiments were performed for each promoter construct. A negative control (NC, promoterless pGL4.15 vector) is also shown. Statistical analysis was performed using one-way repeated-measures analysis of variance (ANOVA) with Greenhouse–Geisser correction followed by Dunnett’s post hoc test. Statistical significance is shown compared to the construct with 11 CTG repeats (* $p < 0.05$, ** $p < 0.01$). Luciferase activity is shown relative to the luciferase signal obtained with the respective promoter construct with 11 CTG repeats.

which are defined as a type of promoter where transcription start sites are spread across a region of around 100 nucleotides^{24–26}.

To study the usage of splice donor sites at the 5' exons located near the CTG TNR, we amplified the fragments encompassing the splice sites from adult human cerebellar RNA using RT-PCR. We identified all previously described *TCF4* exon 4a and 4b splice sites⁴ in adult human cerebellum but could not detect mRNAs starting with exon 4a-II. Similar results were obtained in our previous study⁴, although one respective sequence is present in GenBank, suggesting that the levels of these *TCF4* transcripts are very low. In addition, we found that during splicing, donor site closest to the TSS is used. For instance, *TCF4* mRNAs initiated upstream of 4a-I splice site used donor site 4a-I exclusively and not downstream splice sites 4a-III or 4b (Fig. 1D–F). The RT-PCR analysis confirmed the absence of TSS-s directly upstream of CTG TNR (Fig. 1E). Collectively these results revealed that *TCF4* CTG TNR is not included in the 5' UTR of exon 4a, 4b and 4c transcripts and instead locates in a dispersed promoter^{24–26} region characterized by spread TSS distribution.

Activity of *TCF4* promoters immediately downstream of the CTG TNR decreases with increasing repeat length. We next determined whether the region surrounding the CTG TNR in *TCF4* intron 3 upstream of alternative 5' exons contains functional promoters (Fig. 1C). For that, we analyzed two DNA fragments—a shorter sequence (*TCF4* p4a) spanning the *TCF4* CTG TNR region from just downstream of exon 3 into 5' exon 4a and a longer sequence (*TCF4* p4abc) spanning the entire *TCF4* CTG TNR from just downstream of exon 3 to inside exon 4 (Fig. 1G). We cloned these fragments into pGL4.15[luc2P/Hygro] luciferase reporter vectors and transfected the vectors into rat primary cortical neurons. The expression of luciferase was increased by > 30-fold using reporter constructs containing either p4a or p4abc sequences when compared to a negative control construct without a promoter (Fig. 2A,B). These results indicate that p4a and p4abc sequences contain functional promoters.

To assess the effect of the CTG TNR length on the activity of *TCF4* p4a and p4abc promoter regions, we generated twelve luciferase reporter constructs where each construct contained either the *TCF4* p4a or p4abc promoter sequence combined with six different CTG TNR lengths (11, 25, 31, 54, 67/70, 144). The luciferase reporter assay revealed that an extended CTG TNR with a length of 54, 67/70 or 144 repeats significantly reduced the activity of the promoter for both *TCF4* p4a and p4abc (Fig. 2A,B). The presence of 144 repeats reduced the activity of p4a and p4abc by 70% ($p = 0.0192$) and 75% ($p = 0.0095$), respectively. These results demonstrate that the CTG TNR expansion interferes with transcription from the *TCF4* p4abc promoter region.

The CTG TNR expansion in *TCF4* gene affects the transcription of *TCF4* alternative 5' exons in FECD patients. To describe whether an increased CTG TNR affects the expression of different *TCF4* transcripts, we performed a comprehensive analysis of two previously published RNA-seq datasets from the corneal endothelium of FECD patients with an expanded *TCF4* CTG TNR and control groups without the repeat expansion^{27,28}. The 2019 dataset generated by Nikitina et al. includes 6 controls and 8 FECD patients with an expanded *TCF4* CTG TNR²⁷, and the 2020 dataset by Chu et al. includes 9 controls and 6 FECD patients with an expanded *TCF4* CTG TNR²⁸. First, we evaluated how the levels of transcripts beginning from the CTG TNR region change in FECD. The expression levels of exons 4aI and 4aIII showed a strong decrease in FECD patients, which is in agreement with our luciferase reporter assays in neurons (Fig. 3). In contrast, the levels of transcripts containing 5' exon 4c were increased in FECD patients (Fig. 3).

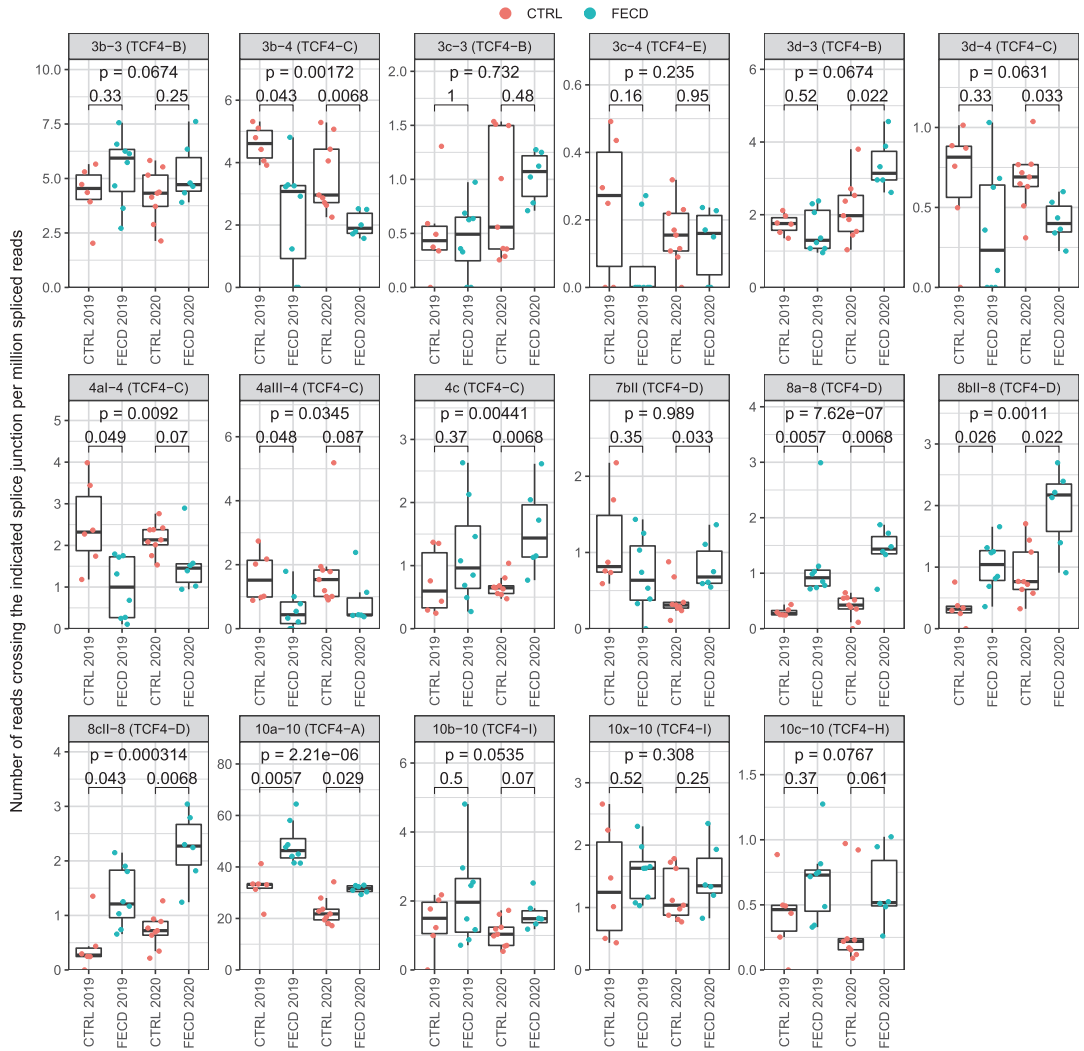


Figure 3. The expression of *TCF4* alternative transcripts containing 5' exons spliced to exon 4 is decreased in the cornea of FECD patients. Two independent FECD RNA-seq datasets—by Nikitina et al.²⁷ (named 2019 in the figure) and Chu et al.²⁸ (named 2020 in the figure)—were used to analyze the expression levels of *TCF4* alternative 5' exons in corneal endothelium. The expression levels of different *TCF4* splicing events were quantified using the number of reads crossing the splice junction (shown above the graphs) normalized with the total number of spliced reads and multiplied by million. The *TCF4* protein isoform encoded by the transcripts containing the indicated splicing event is shown in parentheses above the graphs. The data is visualized as box plots—the hinges show 25% and 75% quartiles, the horizontal line shows the median value, the upper whisker extends from the hinge to the largest value no further than 1.5 * inter-quartile range from the hinge, the lower whisker extends from the hinge to the smallest value at most 1.5 * inter-quartile range of the hinge. All data points are shown with dots. The 2019 dataset includes 6 controls and 8 FECD patients with an expanded *TCF4* CTG TNR, and the 2020 dataset includes 9 controls and 6 FECD patients with an expanded *TCF4* CTG TNR. 10x-10 is a previously undescribed splicing event. Within-experiment statistical analysis between the CTRL and FECD patients (indicated groups) was done using Mann-Whitney U test, p-values were corrected for multiple testing using false discovery rate. Generalized linear model, followed by Wald test was used to determine statistical significance of the disease state combining data from both 2019 and 2020 experiments, p-values were corrected using FDR (p-value in the upper part of the panel).

Next, we investigated whether the CTG TNR affects the levels of *TCF4* transcripts starting from far upstream of the CTG TNR (e.g. exons 3b, 3c, etc.). Our analysis revealed that FECD patients with an expanded CTG TNR displayed reduced levels of transcripts containing *TCF4* alternative 5' exons 3b and 3d spliced directly to internal exon 4, just downstream of the repeat region, thus skipping internal exon 3 (Fig. 3). In contrast, the levels of transcripts containing these exons spliced to the internal exon 3 were either not changed (exons 3c and 3d) or were upregulated (exon 3b) in FECD. These results suggest that the CTG TNR affects both promoter activity and alternative splicing in transcripts starting from upstream of the repeat region (Fig. 3). Notably, we also found that FECD patients had increased levels of transcripts containing 5' exons 8a, 8bII, 8cII and 10a, which are all located far downstream of the CTG TNR (Fig. 3).

Different *TCF4* transcripts encode for various *TCF4* protein isoforms (Fig. 1B) that vary in their function and transactivation capability^{4,8}. The major *TCF4* transcripts comprising of 5' exons 3b, 3c and 3d encode for isoform TCF4-B when spliced to internal exon 3; transcripts with exons 3b and 3d encode for isoform TCF4-C when spliced to exon 4; transcripts with exons 8a, 8bII and 8cII encode for isoform TCF4-D when spliced to exon 8; transcripts with exons 10a, 10b and 10c spliced to exon 10 encode for isoforms TCF4-A, TCF4-I and TCF4-H, respectively (Fig. 1B). Data from *TCF4* transcripts which encode the same protein isoform were combined to determine whether the levels of different *TCF4* transcripts encoding specific *TCF4* protein isoforms are changed in FECD. We found that the expression of transcripts encoding isoform TCF4-C decreased while the levels of isoforms TCF4-A, TCF4-B, TCF4-D and TCF4-H increased in FECD (Fig. 4). These observations were confirmed by analyzing *TCF4* transcripts by the expression levels of internal exons. Decreases in transcripts comprising of internal exons 6–8 were observed in FECD patients, which can be explained by reduced expression of isoform TCF4-C encoding mRNAs (Figs. 4, 5). FECD patients and the control group exhibited equal amounts of transcripts containing exons 8 and 9 (Fig. 4). The sudden elevation of transcripts comprising of exons 8 and 9 when compared to exons 4–8 in FECD patients accounts for the increase in the expression of isoform TCF4-D encoding mRNAs (Figs. 4, 5). An increase in the levels of transcripts containing exons 10–16 present in all *TCF4* transcripts is caused by the increase in the expression of TCF4-A, TCF4-B and TCF4-D mRNAs in FECD patients (Fig. 4). As a contradictory result we saw that transcripts containing exons 3 and 4 which should reflect the levels of isoform TCF4-B did not increase in FECD patients even though the levels of transcripts containing 5' exons included in isoform TCF4-B did increase (Fig. 3).

Only the major transcripts/splicing events are reported in the Figs. 3, 4 and 5. The results of all studied *TCF4* exons/splicing events and isoforms can be found in Supplementary Fig. S2. In conclusion, the expression levels of *TCF4* transcripts were altered in FECD patients—the repeat expansion caused a reduction in transcripts starting immediately downstream of CTG TNR and transcripts containing 5' exons spliced directly to exon 4, and an increase in transcripts encoded by distal 5' exons located hundreds of kbp downstream of the repeat. The results of RNA-seq experiments have been summarized in Fig. 5.

Discussion

Previous studies have indicated that the CTG TNR expansion in intron 3 of *TCF4* strongly increases the risk of developing FECD and also vulnerability and severity of BD^{15,29}. The pathogenic mechanism of the *TCF4* CTG TNR and other TNR-s in general is still a major question. We investigated the hypothesis that the CTG TNR impacts the transcription of *TCF4* mRNAs initiated from nearby 5' exons, leading to an imbalance of the levels of alternative *TCF4* protein isoforms. We show for the first time that the expansion of the CTG TNR directly reduces the activity of the nearby downstream *TCF4* promoters in a length dependent manner—longer, more expanded repeats reduce the activity of proximal downstream promoters. The lengths of the extended repeats that were studied fit into the pathogenic range for both bipolar disorder (>40) and Fuchs' dystrophy (>50)^{14,15}. Soliman et al. has shown that the severity of FECD correlates with the length of the CTG TNR in *TCF4* as patients with a CTG TNR expansion exhibited a more severe form of FECD, but the mechanism underlying this phenotype remains unknown³⁰. It is important to note that our determination of TSS-s by 5' RACE and reporter experiments were done using human cerebellar RNA and rat cultured cortical neurons, respectively. Therefore, it would be interesting to conduct similar experiments in human corneal endothelial cells. This would help to translate our findings between different cell types and further validate the effect of the CTG TNR expansion on transcription also in FECD patients.

Detailed analysis of previously published FECD RNA-seq datasets^{27,28} revealed that the levels of *TCF4* transcripts containing alternative 5' exons 4aI and 4aIII were reduced in the corneal endothelial cells of FECD patients with an expanded CTG TNR. These results support our findings that the *TCF4* CTG TNR expansion reduces the activity of proximal downstream promoters linked to these 5' exons also in human corneal endothelium. Interestingly, an increase in the levels of *TCF4* transcripts encoded by downstream alternative 5' exons distal to the CTG TNR was also noted, which may indicate a compensatory mechanism to rescue the levels of *TCF4* protein arising from the deficit of transcripts encoding TCF4-C. This compensation phenomenon needs to be considered when studying *TCF4* expression levels in FECD and other diseases connected with the *TCF4* intronic CTG TNR and could explain why different research groups have published contradictory results concerning changes in *TCF4* levels when studying FECD^{19–22}. Our results indicate that the levels of *TCF4* transcripts change bidirectionally in response to an expanded CTG TNR—transcripts beginning near the repeat region decline just as Foja et al. reported¹⁹ while certain transcripts beginning downstream of the repeat region increase as reported here, and mask the decrease of long *TCF4* transcripts. As the expression of different *TCF4* transcripts decline and rise simultaneously the overall *TCF4* levels may not change significantly in FECD as has been reported by Mootha et al.²¹ and Oldak et al.²². In contrast, Okumura et al.²⁰ found an increase in total *TCF4* expression levels which is also evident in our RNA-seq analysis as we saw a slight rise in *TCF4* expression when measuring the expression of internal exons present in all *TCF4* transcripts (exons 10–21). Our results illustrate the importance

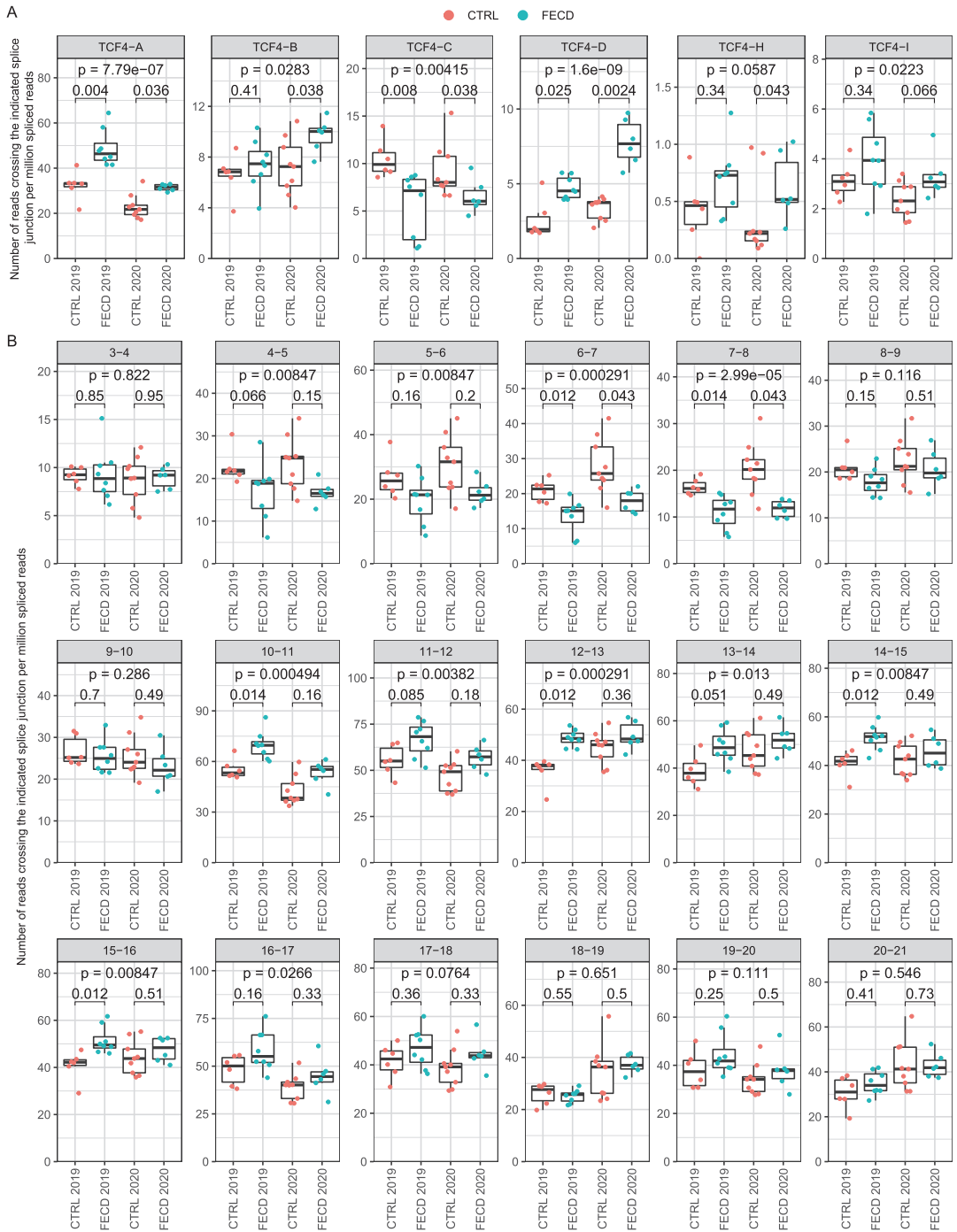


Figure 4. Expression of transcripts encoding TCF4-C isoform are downregulated, whereas transcripts encoding other TCF4 isoforms are upregulated in the cornea of FECD patients. Two independent FECD RNA-seq datasets—by Nikitina et al.²⁷ (named 2019 in the figure) and Chu et al.²⁸ (named 2020 in the figure)—were used to analyze the expression levels of *TCF4* exons in corneal endothelium. The expression levels of (A) TCF4 isoforms and (B) internal exons were quantified using the number of reads (A) from transcripts encoding for the respective isoform (shown above the graphs) and (B) crossing the indicated splice junction (shown above the graphs). Data was normalized with the total number of spliced reads and multiplied by million. For more details, see legend of Fig. 3.

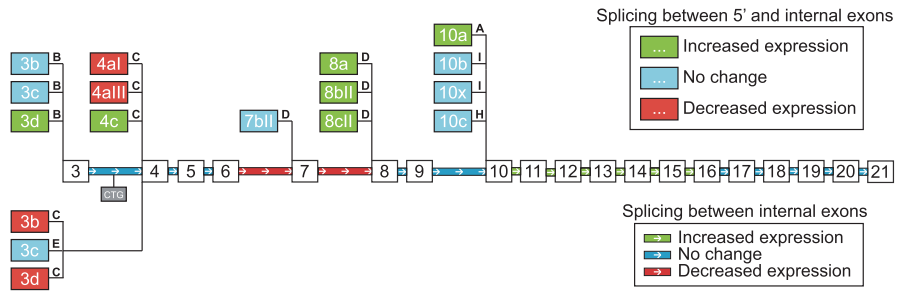


Figure 5. Summary of the expression levels of *TCF4* alternative 5' and internal exons in FECD. Schematic depiction of the human *TCF4* gene structure and transcripts which displayed high expression levels in RNA-seq data analysis (see Figs. 4, 5). Internal exons are depicted as boxes (3–21) and 5' exons are shown as columns on top or under the internal exons. Name of the respective *TCF4* exon is shown inside the box. Splicing of 5' exons is shown by black lines which lead to the respective internal exon. Isoform encoded by the transcript is indicated as a single character (A, B, C, D, E, H or I) marked bold on the line next to the 5' exon. 5' exons marked green or red display significantly increased or decreased expression, respectively, in FECD, when spliced to the respective internal exon. Internal exons are connected by green, blue or red lines. A green or a red line between internal exons indicates a significant increase or decrease, respectively, in the expression level of transcripts containing splicing between these exons in FECD. A blue line or box indicates no change in the levels of transcripts containing splicing between these exons. The CTG TNR is marked between internal exons 3 and 4 as a grey box (CTG).

of the exact transcript measured when studying *TCF4* expression levels. The original RNA-seq study by Chu and others concluded that the *TCF4* CTG TNR expansion increases the stability and thus the amount of expanded CTG repeat-including intronic RNAs in the corneal endothelium and causes comprehensive changes in splicing. No alterations in the overall expression of mature *TCF4* mRNA was noted²⁸. The study by Nikitina et al. was a data article and no conclusions were made²⁷.

Interestingly, we also detected an increase in the expression of 5' exons spliced to exon 3 encoding for TCF4-B in patients with FECD, showing that almost all the *TCF4* promoters far upstream from the CTG TNR had increased activity due to the repeat expansion. However, the increase in upstream promoter activity did not reflect in the levels of transcripts containing exons 3 and 4. It is plausible that the CTG repeat expansion could regulate transcriptional elongation of RNA polymerase by slowing down the polymerase in the CTG TNR region³¹. This can cause an accumulation of RNA polymerases in the CTG TNR and dissociation of the polymerase, leading to a decrease of full-length *TCF4* transcripts beginning from upstream of the CTG repeat. An increase in the expression of 5' exons spliced to exon 3 may be due to preferential splicing of transcripts to the exon before the CTG TNR (exon 3), as changes in splicing have been described before in diseases associated with repeat expansion³¹.

We have previously shown that *TCF4* protein isoforms can be divided into longer and shorter isoforms which vary in their function and transactivation capability^{4,8}. Currently FECD research focuses mainly on the CTG TNR and missplicing of longer *TCF4* transcripts in FECD¹⁷, and little research has been done to analyze the expression of all the *TCF4* transcripts in FECD. Our detailed analysis provides new insight into FECD as we show that the CTG TNR directly modulates the expression of *TCF4* which may be among the underlying causes for the development of the disease. Since *TCF4* mRNAs detected in the present study are expressed virtually in all tissues, with high levels in the fetal and postnatal brain⁴, there may also be a similar correlation between the CTG TNR length and the expression levels of *TCF4* transcripts in vivo in the brain which could predispose development of BD. However, it should be noted that the link between the CTG TNR expansion in *TCF4* and BD has only been shown once and has not been reported by newer studies.

Strong evidence has also been provided in support of a mechanism in which the toxic (CUG)_n TNR containing *TCF4* mRNAs are the cause of Fuchs' corneal dystrophy^{16,32,33}. According to this mechanism, the TNR-carrying RNAs cluster RNA binding proteins, interfering with the splicing of various mRNAs. Of note, antisense therapy using Fuchs' dystrophy ex vivo cell models leads to inhibition of RNA foci and mis-splicing in Fuchs' dystrophy^{34,35}. Since the CTG TNR is located in the intron between exons 3 and 4, this repeat is not included in the fully mature *TCF4* mRNA⁴, but the CTG TNR is still included in the pre-mRNA of transcripts initiated at the upstream promoters (exon 1a, 1b, 3a, 3b, 3c, 3d promoters).

Repeat expansions have been associated with more than 40 diseases²⁹ and unstable TNR-s may occur in both coding and noncoding regions, including promoters, introns and untranslated regions (UTR) of genes³⁶. Among noncoding TNR-s, one of the most studied is the TNR repeat (CGG) located in the 5' UTR of Fragile X Mental Retardation (FMR1) gene. This TNR causes hypermethylation and silencing or increases in the expression level of the gene, depending on TNR length³⁷. TNR diseases with TNR in the promoter region of the affected gene have been less studied. Recently, an intronic polymorphic CGG repeat in a conserved alternative promoter of the *AFF3* gene, an autosomal homolog of the X-linked *AFF2/FMR2* gene, was shown to lead to hypermethylation of the promoter and transcriptional silencing of *AFF3* expression in the brain³⁸. However, the effect of TNR on promoter activity using transient expression analysis of promoters linked to TNR was not studied. Research on

Friedreich ataxia, which is caused by an expansion of the intronic TNR (GAA) in the FXN gene, has revealed reduced expression of the gene in patient derived cell lines³⁹. A hexamer repeat expansion (GGGGCC) located in the 5' regulatory region of the C9ORF72 gene, causing hereditary amyotrophic lateral sclerosis, has been shown to reduce the ability of the surrounding region to promote the expression of a reporter protein in human kidney and neuroblastoma cell lines⁴⁰. Overall, these results indicate that expansion of TNR can alter the expression of the nearby genes. This is in agreement with our results showing that the expression levels of different *TCF4* transcripts are altered in FECD due to the CTG TNR expansion.

Taken together, our results help to explain why previous research on the levels of *TCF4* transcripts in FECD has displayed varying results. Analyzing only total *TCF4* levels or levels of certain *TCF4* transcripts can produce misleading results due to the complexity of the *TCF4* gene and its regulation. The current study shows that the *TCF4* CTG trinucleotide repeat expansion modulates the activity of nearby *TCF4* promoters in a length dependent manner—an expanded CTG TNR causes reduction in promoter activity. Analysis of RNA-seq datasets revealed that the expression levels of the many *TCF4* transcripts are increased or decreased simultaneously in the cornea of FECD patients. Further work is needed to elucidate the exact mechanism how this repeat region affects *TCF4* transcription and whether the changed *TCF4* levels contribute to the development of FECD and BD.

Methods

Generation of DNA constructs. Human postmortem tissues were used to obtain DNA and RNA samples. All protocols using human tissue samples were approved by Tallinn Committee for Medical Studies, National Institute for Health Development (Permit Number 402). All experiments were performed in accordance with relevant guidelines and regulations.

TCF4 gene fragments were screened from human DNA samples for the *TCF4* CTG TNR length and fragments with the desired CTG TNR length were amplified by PCR from 20 ng of genomic DNA in a 20 μ l mixture using 0.4 units of Phusion Hot Start II (Thermo Scientific) and primer p4a_p4abc_F paired with the p4abc_R or p4a_R primer (Supplementary Table S1) with a final concentration of 0.25 μ M to amplify the longer (*TCF4* p4abc) and the shorter (*TCF4* p4a) sequence of the *TCF4* gene (Fig. 2). Following amplification, the PCR mixtures were incubated for 15 min at 72 °C with 1 unit of FirePol DNA polymerase (Solis BioDyne) for the synthesis of adenine overhangs for cloning. The PCR products were first inserted into the pSTBlue-1 acceptor vector (Merck Millipore) and then to the pGL4.15[luc2P/Hygro] luciferase reporter vector (#E6701, Promega).

Promoter regions encompassing CTG TNR-s with five different lengths (11, 25, 31, 54 and 67 or 70 repeats) were acquired from human genomic DNA by PCR. A sixth synthetic DNA segment with 144 CTG repeats was ordered from GenScript. All the generated constructs were verified by sequencing as the length of TNR tended to be unstable in bacteria when producing plasmids (Supplementary Table S1).

Luciferase reporter assay and neuron cultures. The protocols involving animals were approved by the ethics committee of animal experiments at Ministry of Agriculture of Estonia (Permit Number: 45). All experiments were performed in accordance with the relevant guidelines and regulations.

Prenatal rat cortical neurons were cultured as described previously⁴¹. Neurons grown 6 days in vitro were transfected with 180 ng firefly reporter construct and 20 ng pGL4.83[hRlucP/PGK1/Puro] as described previously⁴ for 4 h on a plate shaker using Lipofectamine 2000 (#11668019, Thermo Fisher Scientific) with a reagent to DNA ratio 3:1. Two days after transfection neurons were lysed in 50 μ l Passive Lysis Buffer (Promega) and luciferase reporter assay was performed using the Dual-Glo Luciferase Assay System (Promega) according to manufacturer's protocol. Luciferase signals were measured using the GENios Pro microtiter plate reader (Tecan). For analysis, the signals were first normalized to the signal of the Renilla luciferase and then normalized to the respective ratio in cells transfected with the 11 repeat CTG construct. One-way repeated-measures analysis of variance (ANOVA) with Greenhouse–Geisser correction followed by Dunnett's post hoc test was used to determine the statistical significance compared to the luciferase signals from the 11 repeat CTG construct group.

5' rapid amplification of DNA ends (5' RACE) analysis and reverse transcription polymerase chain reaction (RT-PCR). Total RNA from post-mortem adult human cerebellum was treated with Turbo DNase (Thermo Fisher Scientific) according to the supplier's protocol. 5' RACE analysis was carried out on human cerebellar RNA using the GeneRacer Kit (Thermo Fisher Scientific) according to manufacturer's protocol with primers outlined in Supplementary Tables S1 and S2.

For RT-PCR, cDNA was synthesized from human cerebellar RNA using 100 units of SuperScript III reverse transcriptase (Thermo Fisher Scientific) with oligo(dT)₂₀ and a random hexamer primer mixture (1:1 ratio, Microsynth) according to the manufacturer's protocol. A negative control (– RT) was also included where SuperScript III reverse transcriptase was not added. After cDNA synthesis, PCR was performed in 20 μ l using 3 units of Hot FirePol (Solis BioDyne) and primers listed in Supplementary Table S1 with a final concentration of 0.25 μ M. All the sense primers used for RT-PCR were combined with the antisense primer hTCF4_exon4_as2 except for sense primer hTCF4_4aIII_s (2) which was used together with the antisense primer hTCF4_exon4_as1.

Bioinformatic analysis. Cap Analysis of Gene Expression (CAGE) data from the Functional Annotation of the Mammalian Genome project phase 5 (FANTOM5)⁴² was used to locate potential *TCF4* TSS-s. Both predicted TSS-s (FANTOM5 DPI, robust set) and total counts of CAGE reads for the reverse strand (encoding for *TCF4*) were visualized in UCSC Genome Browser together with the EST-s from GenBank (accessed at 10.07.2020) in the area surrounding the *TCF4* CTG TNR region (chr18:53,254,500–53,252,500, human GRCh37/hg19 assembly). The FANTOM5 data can be accessed at <https://fantom.gsc.riken.jp/5/datahub/hg19/reads/ctsTotalCounts.rev.bw>.

Raw RNA-seq data from corneal endothelium of FECD patients and controls (see Supplementary Table S3 for sample information) were obtained from Sequence Read Archive database (accession numbers PRJNA524323²⁷ and SRP238609²⁸) using prefetch tool (version 2.10.0) from the SRA toolkit. Reads in fastq format were extracted using fasterq-dump. Adapter and quality trimming was done using BBDuk (part of bbmap version 38.79) using the following parameters: ktrim = r k = 23 mink = 11 hdist = 1 tbo qtrim = lr trimq = 10 minlen = 100 (minlen = 85 for data from PRJNA524323). Reads were mapped to hg19 genome (primary assembly and annotation obtained from GENCODE, release 34, GRCh37) using STAR aligner (version 2.7.3a) with default parameters. To increase sensitivity for unannotated splice junctions, splice junctions obtained from the 1st pass were combined (per dataset) and filtered as follows: junctions on mitochondrial DNA and non-canonical intron motifs were removed; only junctions supported by at least 6 reads in the whole dataset were kept. The filtered junctions were added to the 2nd pass mapping using STAR. Intron-spanning reads were quantified using FeatureCounts (version 2.0.0) with the following parameters: -p -B -C -s 2 -J. To count reads from TCF4 extended exons (exons 4c and 7bII), reads crossing a region 2 bp 5' from the internal exon (exon 4 and 7, respectively) were quantified using FeatureCounts and a custom-made saf file. Splice junctions in the TCF4 region showing less than 4 reads for the whole dataset were discarded, the rest of the splice junctions associated with TCF4 were manually curated and annotated according to Sepp et al.⁴. A custom R script was used to quantify the expression of different TCF4 splice variants. Reads crossing the indicated splice junctions were normalized using the number of all splice-junction crossing reads in the respective samples. Then, data summed by the Exon column (see Supplementary Table S4) to obtain expression levels of splice junctions for TCF4 5' exons. Next, data was aggregated by the Isoform column (see Supplementary Table S4) to obtain expression levels of spliced reads of TCF4 internal exons and transcripts encoding different TCF4 protein isoforms. The annotated splice junction table for quantifying different TCF4 splice sites and transcripts encoding different isoforms is shown in Supplementary Table S4. The results were visualized using ggplot2 package (version 3.3.1) in R (version 4.0.1). Statistical analysis of the RNA-seq data was carried out in R as follows. To determine statistical significance between control and FECD patients within an experiment, non-parametric Mann–Whitney U-test was performed, p-values were corrected for multiple comparisons within experiment (per figure) using false discovery rate (FDR). To determine general statistical significance of the disease state for combined data of the two experiments, normalized data was transformed by adding 0.01, followed by fitting generalized linear model with Gamma distribution using Experiment + Disease + Experiment:Disease as the model. p-value for the disease state was obtained using Wald test and corrected for multiple comparisons using FDR (per figure).

Received: 17 August 2020; Accepted: 14 October 2020

Published online: 28 October 2020

References

- Zhuang, Y., Cheng, P. & Weintraub, H. 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 (1996).
- Guillemot, F. Spatial and temporal specification of neural fates by transcription factor codes. *Development* **134**, 3771–3780 (2007).
- Zweier, C. et al. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am. J. Hum. Genet.* **80**, 994–1001 (2007).
- Sepp, M., Kannike, K., Eesmaa, A., Urb, M. & Timmusk, T. Functional diversity of human basic helix-loop-helix transcription factor TCF4 isoforms generated by alternative 5' exon usage and splicing. *PLoS ONE* **6**, e21238 (2011).
- Fagerberg, L. et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell Proteomics* **13**, 397–406 (2014).
- Rannals, M. D. & Maher, B. J. Molecular mechanisms of transcription factor 4 in Pitt Hopkins syndrome. *Curr. Genet. Med. Rep.* **5**, 1–7 (2017).
- Ma, C., Gu, C., Huo, Y., Li, X. & Luo, X.-J. The integrated landscape of causal genes and pathways in schizophrenia. *Transl. Psychiatry* <https://doi.org/10.1038/s41398-018-0114-x> (2018).
- Sepp, M. et al. The intellectual disability and schizophrenia associated transcription factor TCF4 is regulated by neuronal activity and protein kinase A. *J. Neurosci.* **37**, 10516–10527 (2017).
- Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
- Stefansson, H. et al. Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
- Dooستparast Torshizi, A. et al. Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. *Sci. Adv.* **5**, 4139 (2019).
- Kharbanda, M. et al. 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 (2016).
- Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668 (2018).
- Wieben, E. D. et al. A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2–2) gene predicts Fuchs corneal dystrophy. *PLoS ONE* **7**, e49083 (2012).
- Del-Favero, J. et al. European combined analysis of the CTG18.1 and the ERDA1 CAG/CTG repeats in bipolar disorder. *Eur. J. Hum. Genet.* **10**, 276–280 (2002).
- Fautsch, M. P. et al. TCF4-mediated Fuchs endothelial corneal dystrophy: Insights into a common trinucleotide repeat-associated disease. *Prog. Retinal Eye Res.* <https://doi.org/10.1016/j.preteyeres.2020.100883> (2020).
- Ong Tone, S. et al. Fuchs endothelial corneal dystrophy: The vicious cycle of Fuchs pathogenesis. *Prog. Retinal Eye Res.* <https://doi.org/10.1016/j.preteyeres.2020.100863> (2020).
- Vieta, E. et al. Bipolar disorders. *Nat. Rev. Dis. Primers* **4**, 1–16 (2018).
- Foja, S., Luther, M., Hoffmann, K., Rupprecht, A. & Gruenauer-Kloevckorn, C. CTG181 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 (2017).
- Okumura, N. et al. Effect of trinucleotide repeat expansion on the expression of TCF4 mRNA in Fuchs' endothelial corneal dystrophy. *Investig. Ophthalmol. Vis. Sci.* **60**, 779–786 (2019).

21. Mootha, V. V. *et al.* TCF4 triplet repeat expansion and nuclear RNA foci in Fuchs' endothelial corneal dystrophy. *Investig. Ophthalmol. Vis. Sci.* **56**, 2003–2011 (2015).
22. Oldak, M. *et al.* Fuchs endothelial corneal dystrophy: Strong association with rs613872 not paralleled by changes in corneal endothelial TCF4 mRNA level. *Biomed. Res. Int.* **2015**, 640234 (2015).
23. Ward, M. C. & Gilad, Y. Human genomics: Cracking the regulatory code. *Nature* **550**, 190–191 (2017).
24. Juven-Gershon, T., Hsu, J.-Y., Theisen, J. W. M. & Kadonaga, J. T. The RNA polymerase II core promoter—The gateway to transcription. *Curr. Opin. Cell Biol.* **20**, 253–259 (2008).
25. Brandenberger, R. *et al.* Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat. Biotechnol.* **22**, 707–716 (2004).
26. Suzuki, Y. *et al.* Large-scale collection and characterization of promoters of human and mouse genes. *In Silico Biol.* **4**, 429–444 (2004).
27. Nikitina, A. S. *et al.* Dataset on transcriptome profiling of corneal endothelium from patients with Fuchs endothelial corneal dystrophy. *Data Brief* **25**, 104047 (2019).
28. Chu, Y. *et al.* Analyzing pre-symptomatic tissue to gain insights into the molecular and mechanistic origins of late-onset degenerative trinucleotide repeat disease. *Nucleic Acids Res.* **48**, 6740–6758 (2020).
29. Paulson, H. Repeat expansion diseases. *Handb. Clin. Neurol.* **147**, 105–123 (2018).
30. Soliman, A. Z., Xing, C., Radwan, S. H., Gong, X. & Mootha, V. V. Correlation of severity of Fuchs endothelial corneal dystrophy with triplet repeat expansion in TCF4. *JAMA Ophthalmol.* **133**, 1386–1391 (2015).
31. Rohilla, K. J. & Gagnon, K. T. RNA biology of disease-associated microsatellite repeat expansions. *Acta Neuropathol. Commun.* <https://doi.org/10.1186/s40478-017-0468-y> (2017).
32. Du, J. *et al.* RNA toxicity and missplicing in the common eye disease Fuchs endothelial corneal dystrophy. *J. Biol. Chem.* **290**, 5979–5990 (2015).
33. Rong, Z., Hu, J., Corey, D. R. & Mootha, V. V. Quantitative studies of muscleblind proteins and their interaction with TCF4 RNA foci support involvement in the mechanism of Fuchs' dystrophy. *Investig. Ophthalmol. Vis. Sci.* **60**, 3980–3991 (2019).
34. Hu, J. *et al.* Oligonucleotides targeting TCF4 triplet repeat expansion inhibit RNA foci and mis-splicing in Fuchs' dystrophy. *Hum. Mol. Genet.* <https://doi.org/10.1093/hmg/ddy018> (2018).
35. Zarouchlioti, C. *et al.* Antisense therapy for a common corneal dystrophy ameliorates TCF4 repeat expansion-mediated toxicity. *Am. J. Hum. Genet.* <https://doi.org/10.1016/j.ajhg.2018.02.010> (2018).
36. Nelson, D. L., Orr, H. T. & Warren, S. T. The unstable repeats—Three evolving faces of neurological disease. *Neuron* **77**, 825–843 (2013).
37. Salcedo-Arellano, M. J., Dufour, B., McLennan, Y., Martinez-Cerdeno, V. & Hagerman, R. Fragile X syndrome and associated disorders: Clinical aspects and pathology. *Neurobiol. Dis.* **136**, 104740 (2020).
38. Metsu, S. *et al.* FRA2A is a CGG repeat expansion associated with silencing of AFF3. *PLoS Genet.* **10**, e1004242 (2014).
39. Chutake, Y. K., Lam, C., Costello, W. N., Anderson, M. & Bidichandani, S. I. Epigenetic promoter silencing in Friedreich ataxia is dependent on repeat length. *Ann. Neurol.* **76**, 522–528 (2014).
40. Gijssels, I. *et al.* The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Mol. Psychiatry* **21**, 1112–1124 (2016).
41. Esvold, E.-E. *et al.* CREB family transcription factors are major mediators of BDNF transcriptional autoregulation in cortical neurons. *J. Neurosci.* **40**, 1405–1426 (2020).
42. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 22 (2015).

Acknowledgements

We 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. The authors would also like to thank Laura Tamberg, Anastassia Šubina and Mari Maria Palgi for critical reading of the manuscript and Epp Väli for technical assistance.

Author contributions

A.S.: conceptualization, writing, experimentation, bioinformatics; K.L.: conceptualization, writing, experimentation; J.T.: conceptualization, writing, bioinformatics; K.N.: conceptualization, writing, supervision; M.S.: conceptualization, writing supervision, funding acquisition; T.T.: conceptualization, writing, supervision, funding acquisition.

Funding

This project was supported by Estonian Research Council (institutional research funding IUT19-18 and grant PRG805), European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012), H2020-MSCA-RISE-2016 (Grant EU734791), Pitt Hopkins Research Foundation (Grants No. 8 and No. 21) and 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). The funding sources were not involved in study design, analysis and interpretation of data, writing of the report and in the decision to submit the article for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-75437-3>.

Correspondence and requests for materials should be addressed to T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

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.



Functional consequences of TCF4 missense substitutions associated with Pitt-Hopkins syndrome, mild intellectual disability, and schizophrenia

Received for publication, August 6, 2021, and in revised form, October 28, 2021. Published, Papers in Press, November 6, 2021.
<https://doi.org/10.1016/j.jbc.2021.101381>

Alex Sirp^{1,†}, Kaisa Roots^{1,†}, Kaja Nurm¹, Jürgen Tuvikene^{1,2}, Mari Sepp^{1,5,*}, and Tõnis Timmusk^{1,2,5,*}

From the ¹Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia; ²Protobios LLC, Tallinn, Estonia

Edited by Roger Colbran

Transcription factor 4 (TCF4) is a basic helix-loop-helix transcription factor essential for neurocognitive development. The aberrations in *TCF4* are associated with neurodevelopmental disorders including schizophrenia, intellectual disability, and Pitt-Hopkins syndrome, an autism-spectrum disorder characterized by developmental delay. Several disease-associated missense mutations in *TCF4* have been shown to interfere with TCF4 function, but for many mutations, the impact remains undefined. Here, we tested the effects of 12 functionally uncharacterized disease-associated missense mutations and variations in *TCF4* using transient expression in mammalian cells, confocal imaging, *in vitro* DNA-binding assays, and reporter assays. We show that Pitt-Hopkins syndrome-associated missense mutations within the basic helix-loop-helix domain of TCF4 and a Rett-like syndrome-associated mutation in a transcription activation domain result in altered DNA-binding and transcriptional activity of the protein. Some of the missense variations found in schizophrenia patients slightly increase TCF4 transcriptional activity, whereas no effects were detected for missense mutations linked to mild intellectual disability. We in addition find that the outcomes of several disease-related mutations are affected by cell type, TCF4 isoform, and dimerization partner, suggesting that the effects of TCF4 mutations are context-dependent. Together with previous work, this study provides a basis for the interpretation of the functional consequences of *TCF4* missense variants.

Transcription factor 4 (TCF4) is vital for normal development of the central nervous system, and the mutations within the *TCF4* gene have been linked to several neurodevelopmental diseases such as Pitt-Hopkins syndrome (PTHS), mild-to-moderate intellectual disability (MMID), and schizophrenia (SCZ) (1). A single missense mutation in the basic helix-loop-helix (bHLH) region, which mediates

dimerization and DNA binding, is enough to cause PTHS: a rare syndromic encephalopathy characterized by severe intellectual disability, autistic-like behaviour, distinct facial features, breathing abnormalities, absent language, motor deficits, and epilepsy (2–4). In addition, a variety of *TCF4 de novo* translocations, deletions, insertions, nonsense, frameshift, and splice-site mutations which affect overall *TCF4* expression or the functionality of the bHLH domain have been identified in PTHS patients (5–7). *De novo* partial deletions, translocations, missense, and truncating mutations in *TCF4* have also been identified in MMID patients without the typical characteristics of PTHS (8–13) and one missense mutation in TCF4 has been linked to a Rett-like syndrome (RTT-like) (14). *TCF4* is one of the susceptibility genes for SCZ with several intronic SNPs (15–19) and the missense variations (20, 21) in *TCF4* found to be associated with the disease. In addition to neurodevelopmental diseases, *TCF4* is one of the most frequently mutated genes in adult sonic-hedgehog-driven medulloblastoma, a tumor of the cerebellum (22, 23).

As a class I bHLH transcription factor, TCF4 forms homo or heterodimers with other bHLH transcription factor family members and binds to the Ephrussi box consensus sequence (CANNTG) (24–26). The overall regulation of TCF4 activity is complex and relies on the expression pattern of its activating and inhibiting interaction partners throughout development (27, 28). The differential expression of interaction partners allows TCF4 to exert various functions necessary for normal brain development such as apoptosis, proliferation, signaling, and migration (29–31). For example, TCF4 heterodimerizes with achaete-scute homolog 1 (ASCL1) (32) and regulates the differentiation of neural stem cells (33).

Over 18 different protein isoforms are encoded by the human *TCF4* gene (34). The most studied TCF4 isoforms to date are TCF4-B and TCF4-A, which represent a long and a short TCF4 isoform, respectively (Fig. 1A). All TCF4 isoforms exhibit the evolutionally conserved bHLH domain and activation domains 2 and 3 (AD2 and AD3), but only long isoforms have an additional activation domain 1 (AD1) (34, 35). As a result, the longer isoform TCF4-B, that exhibits full-length AD1, has a higher transcriptional activity than the shorter isoform TCF4-A, that lacks AD1 (34, 36). The TCF4

[†] These authors contributed equally to this work as first authors.

[‡] These authors contributed equally to this work as senior authors.

* For correspondence: Mari Sepp, m.sepp@zmbh.uni-heidelberg.de; Tõnis Timmusk, tonis.timmusk@taltech.ee.

Present address for Mari Sepp: Center for Molecular Biology of Heidelberg University (ZMBH), Heidelberg, 69120, Germany.

Functional consequences of TCF4 missense substitutions

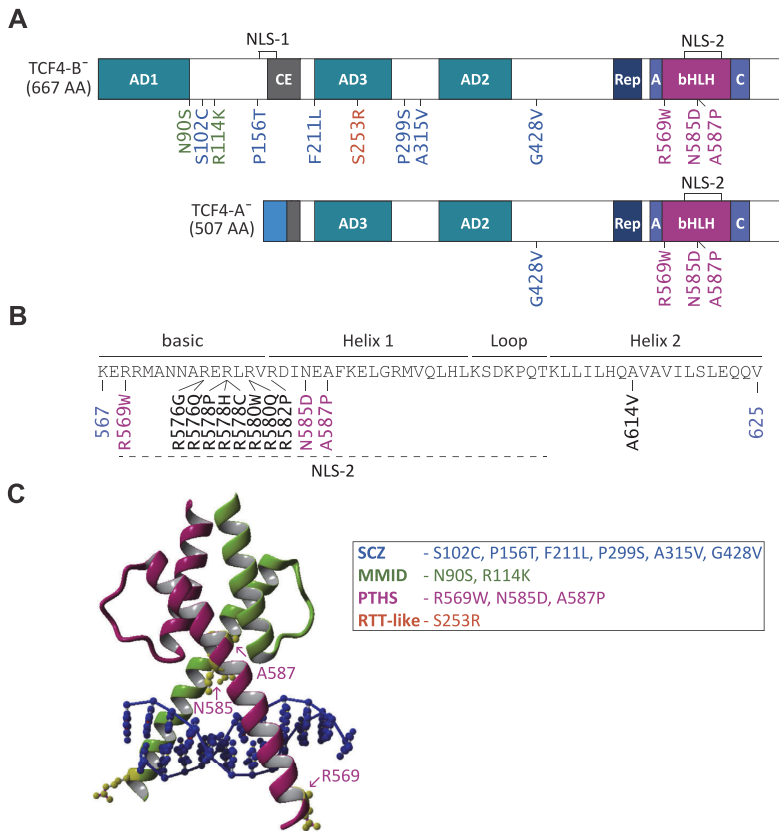


Figure 1. Localization of disease-related missense variations and mutations in TCF4. *A*, representation of TCF4 isoforms B and A. The colored boxes show conserved protein domains. The dark-grey region in TCF4-A represents a nonfunctional partial CE domain (according to Herbst and Kolligs (37)) and the blue region in TCF4-A shows unique amino acids encoded by exon 10a of *TCF4* (34). Nuclear localization signals are shown on top. Disease-related missense substitutions analyzed in the present article are shown below the protein isoforms and marked according to the color code below. Missense mutations and variations have been associated with mild-to-moderate intellectual disability (MMID; N90S and R114K, marked green); schizophrenia (SCZ; S102C, P156T, F211L, P299S, A315V, and G428V, marked blue), Rett-like syndrome (RTT-like; S253R, marked orange), and Pitt-Hopkins syndrome (PTHS; R569W, N585D, and A587P, marked purple). *B*, the position of missense mutations in TCF4 bHLH identified in PTHS patients to date. Novel PTHS-associated missense mutations are shown in purple. The protein regions are indicated with lines on top, and NLS-2 is marked with a dashed underline. The bHLH beginning (567) and end (625) coordinates are in the context of TCF4-B⁻. *C*, ribbon drawing of TCF4 structure model (5) showing bHLH homodimer bound to DNA (dark blue). Pink and green structures indicate the 'nonspecific' and 'specific' TCF4 subunit, respectively. The novel missense substitutions identified in PTHS patients are colored yellow and indicated with arrows. A, motif A; bHLH, basic helix-loop-helix; C, motif C; CE, conserved element; AD1, AD2, and AD3, transcription activation domain; NLS, nuclear localization signal; Rep, repression domain; TCF4, transcription factor 4.

activation domains are regulated by the so-called conserved element (CE), which can repress AD1-dependent transcription (37), and the transcriptional repression domain which can repress both AD1 and AD2 (38). Motifs A and C on either side of the bHLH domain regulate homo and heterodimerization (39, 40). The nuclear localization of TCF4 is mediated by two nuclear localization signals (NLS): a bipartite NLS-1 coded by exons 8 and 9 (34), which can also function as nucleolar localization signal, and a recently identified NLS-2 inside the bHLH domain (41).

TCF4 is broadly expressed, but its mRNA levels vary in different tissues with high expression detected in different brain regions (34, 42–45). Homozygous deletion of *Tcf4* is lethal in mice (46). *Tcf4* haplo-insufficient mice display

impaired learning and motor control, memory deficits, social isolation (47) and microcephaly, hyperactivity, reduced anxiety, and enhanced long-term potentiation (48, 49). *In utero* TCF4 gain-of-function studies in rats lead to enhanced spontaneous activity of prefrontal neurons and disruption of the formation of prefrontal cortical minicolumns (44). *In utero* suppression of *Tcf4* in rats leads to decreased excitability of prefrontal neurons (50). Overexpression of TCF4 in the mouse brain causes impairments in cognition and sensorimotor gating (51) and enhanced long-term depression (48). In addition, the overexpression of Daughterless, the *Drosophila melanogaster* orthologue of TCF4, is lethal in adult flies, whereas the downregulation of Daughterless impairs memory and learning (52, 53).

Functional consequences of TCF4 missense substitutions

It is known that PTHS is caused by loss-of-function mutations leading to *TCF4* haploinsufficiency (7). It is less known how TCF4 may be involved in the development of MMID and SCZ, however, evidence suggests that it may be because of changes in TCF4 dosage. MMID is caused by mutations located in the 5' region of the *TCF4* gene thus only affecting the expression of long TCF4 isoforms (12). In the case of SCZ, the studies suggest that *TCF4* expression levels are elevated ((54) medRxiv, (55)). Functional impact of many of the PTHS-associated missense mutations have been studied before (4, 5, 45, 56). According to these studies, mutant TCF4 proteins display changes in DNA binding, dimerization, transcription activation, and intranuclear localization. The effects range from hypomorphic to dominant negative, and the missense mutations located in the bHLH domain have the most severe functional impact. In addition, the variations associated with SCZ (located outside the bHLH region) have been shown to increase the activity of TCF4 (36).

Several novel disease-related missense variations and the mutations associated with SCZ, MMID, RTT-like syndrome, and PTHS have been identified in TCF4, however, the functional impact of these substitutions has not been studied. Here, we investigated the effects of previously uncharacterized disease-related missense mutations in TCF4 on its dimerization, DNA binding, and transactivation ability and on its subcellular distribution. Our results extend current knowledge on how amino acid substitutions in TCF4 can affect the functionality of the protein.

Results

Mapping of disease-related missense mutations in TCF4

Several missense variations and mutations across the *TCF4* coding sequence have been identified in MMID, SCZ, RTT-like syndrome, and PTHS patients. For the present study, we selected 12 functionally undescribed missense variations and mutations, numbered them in the context of the full-length isoform TCF4-B⁺, and mapped them onto isoforms TCF4-B and TCF4-A (Fig. 1A). Schizophrenia-associated variations have been described in two separate studies (20, 21). All but one of the schizophrenia-associated missense variations have been identified in a single patient and no controls, however, the variation A315V has been found in 14 patients and 10 controls. The variations S102C, F211L, P299S, A315V, and G428V are located in exons 6, 10, 12, 13, and 16, respectively. These variations do not map to any known functionally important region of TCF4 protein. The variation P156T is in exon 8 in front of NLS-1 coding sequence. MMID mutations N90S and R114K (Pitt-Hopkins Research Foundation, Audrey Davidow Lapidus, personal communication) are in *TCF4* exons 6 and 7, respectively. N90S is located at the end of AD1, whereas R114K is not located to any functionally important domain of TCF4. The RTT-like syndrome mutation S253R (14) is in exon 11 in AD3 coding sequence (Fig. 1A).

Missense mutations in PTHS patients are predominantly, but not always, located in the bHLH domain of TCF4. Previous

Table 1

List of missense mutations in TCF4 basic helix-loop-helix domain identified in Pitt-Hopkins syndrome patients to date

No	TCF4 mutations	AA changes	References
1	1705 C>T	R569W	Whalen <i>et al.</i> , 2012 (6)
2	1726 C>G	R576G	de Pontual <i>et al.</i> , 2009 (45)
3	1727 G>A	R576Q	de Pontual <i>et al.</i> , 2009 (45)
4	1732 C>T	R578C	Marangi <i>et al.</i> , 2012 (57)
5	1733 G>C	R578P	Zweier <i>et al.</i> , 2008 (67)
6	1733 G>A	R578H	Zweier <i>et al.</i> , 2008 (67)
7	1738 C>T	R580W	Amiel <i>et al.</i> , 2007 (2); Zweier <i>et al.</i> , 2007 (4)
8	1739 G>A	R580Q	Amiel <i>et al.</i> , 2007 (2)
9	1745 G>C	R582P	Takano <i>et al.</i> , 2010 (68)
10	1753 A>G	N585D	Marangi <i>et al.</i> , 2012 (57)
11	1759 G>C	A587P	Whalen <i>et al.</i> , 2012 (6)
12	1841 C>T	A614V	de Pontual <i>et al.</i> , 2009 (45)

studies have identified 12 amino acid substitutions in eight positions of the TCF4 bHLH region in 25 PTHS patients (Table 1, Fig. 1B). Most of the amino acid substitutions affect four arginines (R569, R576, R578, and R580) in the basic region of the bHLH domain. Three missense mutations (R582P, N585D, and A587P) are in helix 1 and one missense mutation (A614V) in helix 2. Of these, we have previously functionally characterized mutations at R576, R578, and R580 in the basic region, R582 in helix 1, and A614 in helix 2. These mutations showed varied severity ranging from hypomorphic effects to complete loss-of-function (5). Here, we addressed the functional impact of three additional amino acid substitutions R569W, N585D, and A587P (6, 57). We began by estimating possible effects of these missense mutations based on previously generated bHLH structure model (5). R569 resides in the beginning of the basic region and according to the model, contacts DNA at an E-box flanking region. N585 and A587 are both located in helix 1 of the bHLH domain. In the TCF4 model, the amino acid N585 forms hydrogen bonds with the DNA backbone whereas A587 packs against residue L611 in helix 2 of the dimerization partner (Fig. 1C).

Only one SCZ (P156T), MMID (N90S), and RTT-like syndrome (S253R) related missense mutation is in a functionally important domain of TCF4. However, all PTHS-associated missense mutations in the bHLH domain affect conserved amino acids (5, 52), suggesting a high importance of those residues in modulating dimerization, DNA binding, and/or nuclear localization.

Intranuclear localization of TCF4-B is altered by PTHS-associated missense mutations

To dissect the functional impact of the novel disease-related TCF4 missense mutations, we introduced point mutations into the TCF4-B expression vector *via* site-directed mutagenesis and transfected WT and mutant constructs into HEK293 cells. Western blot analysis revealed that both the untagged and E2-tagged mutant proteins were overexpressed at comparable levels in HEK293 cells, although A587P mutant proteins displayed lower expression levels than the WT protein (Fig. 2A).

Functional consequences of TCF4 missense substitutions

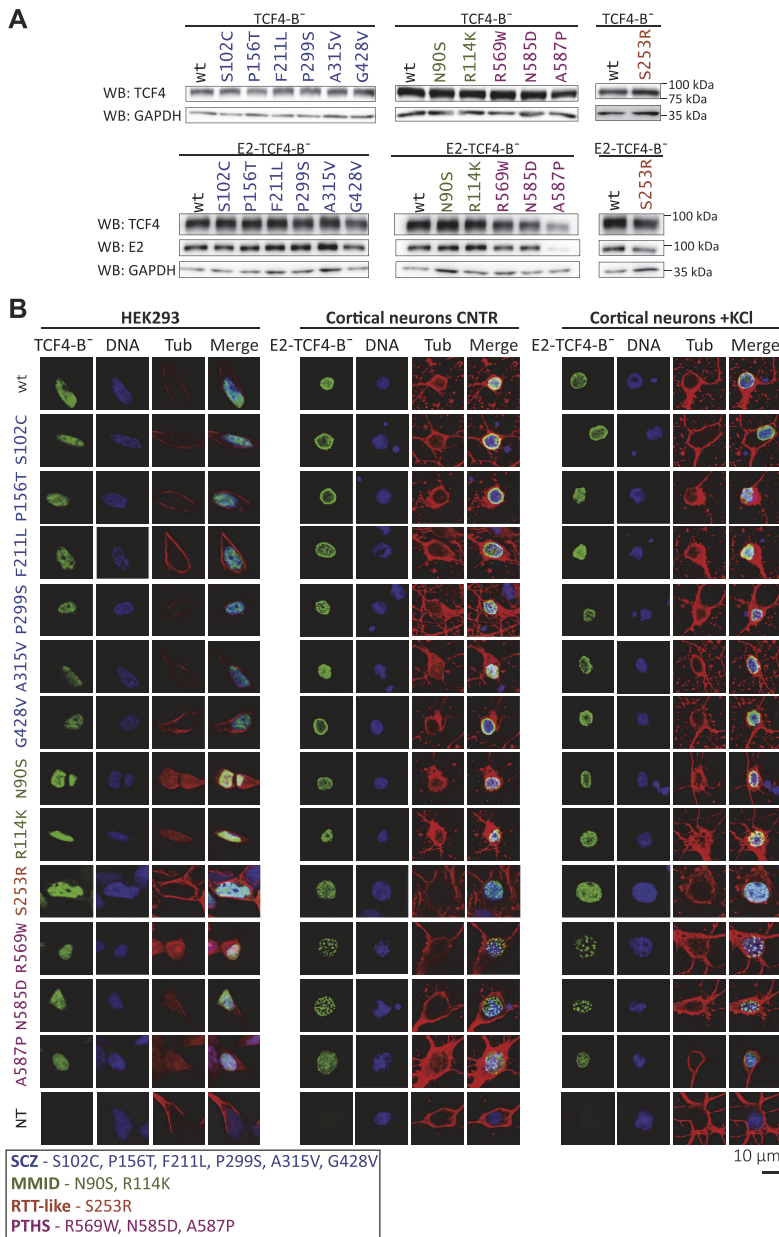


Figure 2. Intracellular localization of WT and mutant TCF4-B in cultured cells. A, Western blot analysis of HEK293 cells transfected with untagged or E2-tagged WT or mutant TCF4-B. TCF4 and GAPDH (loading control) signals were detected with specific antibodies. Molecular mass markers are indicated on the right. B, immunocytochemical analysis of WT or mutant TCF4-B overexpressed in HEK293 cells and of WT or mutant E2-TCF4-B overexpressed in rat hippocampal and cortical primary neurons under basal (CNTR) and depolarized (+KCl) conditions. TCF4, nuclei, and cytoskeleton were visualized with anti-TCF4 or anti-E2 antibody (green, indicated as TCF4-B), Hoechst 33342 (blue, indicated as DNA), and anti-tubulin- β antibody (red, indicated as Tub), respectively. The lack of signal in cells not transfected with TCF4 constructs (NT) demonstrates the specificity of anti-TCF4 (HEK293 cells) and anti-E2 (cortical neurons) antibodies. The representative confocal microscopy images are shown. HEK, human embryonic kidney; MMID, Mild-to-moderate intellectual disability; PTHS, Pitt-Hopkins syndrome; RTT-like, Rett-like syndrome; SCZ, Schizophrenia; TCF4, transcription factor 4.

Previous studies have shown that the long TCF4-B isoform contains two NLSs located at the amino acids R157-K175 (NLS-1) and R569-K607 (NLS-2) (34, 41).

We hypothesized that schizophrenia-associated missense variation P156T could disrupt the N-terminal bipartite NLS-1, whereas PTHS-associated mutations could have an

Functional consequences of TCF4 missense substitutions

impact on the function of the NLS-2 located in the bHLH region.

We studied the localization of TCF4 proteins in HEK293 cells and in cultured rat cortical and hippocampal primary neurons transfected with WT or mutant TCF4-B expression constructs. Because TCF4 is highly expressed in the brain (42), we used E2-tagged TCF4-B coding constructs in experiments with primary neurons. The use of E2-tagged proteins avoided the simultaneous detection of overexpressed and endogenous TCF4 in primary neurons, where endogenous TCF4 levels are much higher than in HEK293 cells. As TCF4 is an activity-regulated transcription factor (36), we in addition studied TCF4-B localization in neurons treated with 25 mM KCl for 8 h to induce neuronal activity. Our results show that all mutant TCF4-B proteins localize to the cell nucleus both in HEK293 cells and in primary neurons, similarly to WT TCF4-B (Fig. 2B). However, in contrast to the homogenous intranuclear distribution of the WT protein, R569W and N585D mutants were detected in the nuclei of primary neurons as puncta, which were universally present in all transfected cells and remained unchanged in response to KCl treatment (Fig. 2B). We conclude that the studied disease-related missense substitutions in TCF4-B do not affect nuclear import of the protein, but the PTHS-associated mutations R569W and N585D modify its distribution inside the nuclei of neurons.

Heterodimerization capacity of TCF4-A is not impaired by mutations in the C-terminal region of the protein

We next investigated the possibility that disease-related missense mutations in TCF4 could interfere with the ability of TCF4 to form heterodimers with its well-known dimerization partners such as the class II bHLH transcription factors ASCL1 and NEUROD2, which are involved in the regulation of

cell-type-specific transcription (32, 51). To study the heterodimerization capacity of mutant TCF4 proteins, we took advantage of the knowledge that the overexpressed TCF4-A is located both in the cytoplasm and nucleus, and localization strictly to the nucleus is accomplished *via* heterodimerization with NLS-bearing dimerization partners (5). In these experiments, we focused on the four C-terminal mutations located within or close to the bHLH domain. Schizophrenia-related G428V and PTHS-associated R569W, N585D and A587P were introduced into TCF4-A vector (Fig. 1A) and expressed alone or together with ASCL1 or NEUROD2-E2 in HEK293 cells. Immunocytochemical analysis demonstrated that similarly to WT TCF4-A, all mutant proteins localized both in the cytoplasm and in the nucleus, possibly because of heterodimerization with endogenously expressed dimerization partners or NLS-2 (Fig. 3). The coexpression with ASCL1 or NEUROD2-E2 caused the mutant TCF4-A proteins to localize strictly into the cell nucleus, and no signal was detected in the cytoplasm (Fig. 3). These results show that the four studied disease-related missense substitutions do not disrupt TCF4 heterodimerization.

PTHS and RTT-like syndrome-associated mutations in TCF4 decrease its DNA-binding activity in a dimerization context-dependent manner

Subsequently, we evaluated the ability of mutant TCF4-B to bind DNA using EMSA. For this, *in vitro* translated WT and mutant TCF4-B proteins were incubated with μ E5 (CACCTG) containing oligonucleotides and separated on a gel. *In vitro* translated TCF4-B mutants were first visualized by Western blot analysis (Fig. 4A) to confirm their translation. The binding of TCF4-B to the μ E5 oligonucleotides was specific (Figs. 4, B–D and S1), as also shown before (5).

First, we studied the effects of disease-related mutations on the ability of TCF4-B homodimers to bind DNA (Figs. 4B and

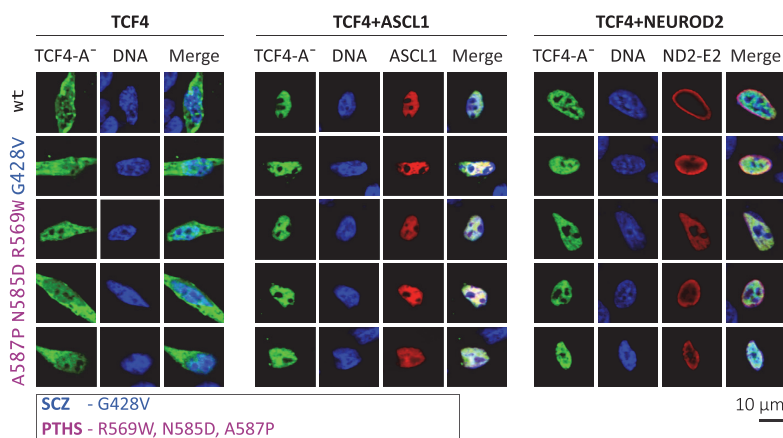


Figure 3. Heterodimerization of WT or mutant TCF4-A with ASCL1 and NEUROD2 in HEK293 cells. Nuclear redirection assay with WT or mutant TCF4-A proteins overexpressed in HEK293 cells alone, together with ASCL1 or together with NEUROD2-E2 (ND2-E2). Immunocytochemical staining was carried out with anti-TCF4 (green, indicated as TCF4-A), anti-MASH1 (red, indicated as ASCL1), and anti-E2 (red, indicated as ND2-E2) antibodies, and the nuclei were visualized with Hoechst 33342 (blue, indicated as DNA). The images were taken by confocal microscopy. HEK, human embryonic kidney; PTHS, Pitt-Hopkins syndrome; SCZ, Schizophrenia; TCF4, transcription factor 4.

Functional consequences of TCF4 missense substitutions

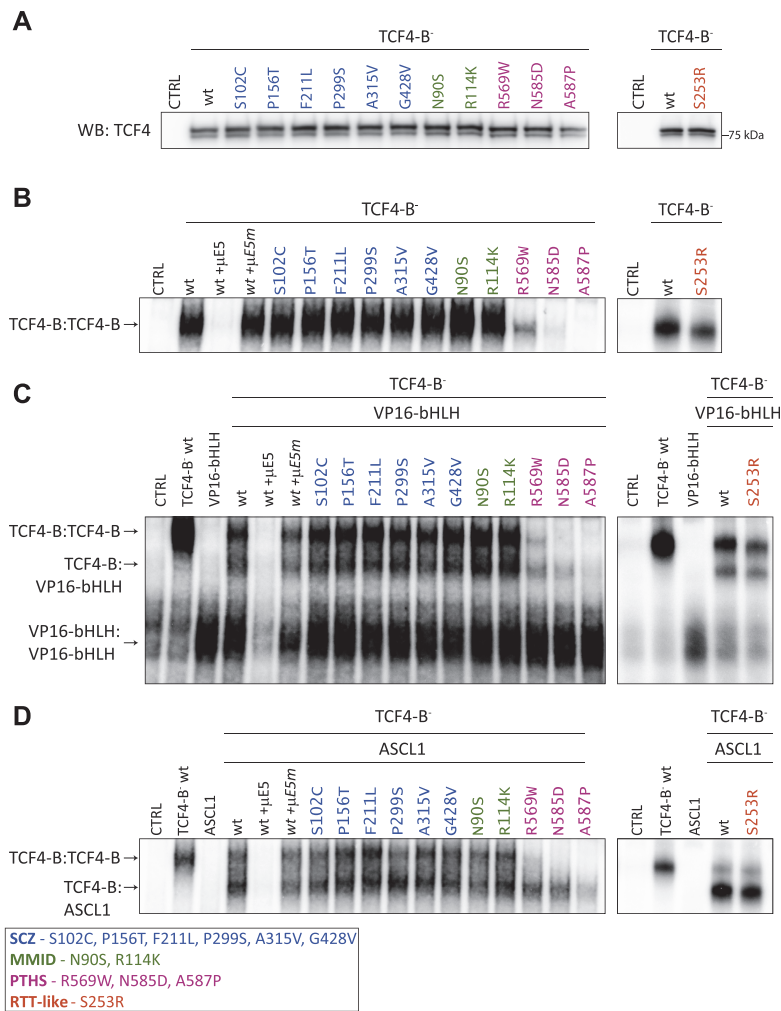


Figure 4. DNA binding of TCF4 homo and heterodimers is impaired by mutations associated with PTHS. *A*, Western blot analysis of *in vitro*-translated WT or disease-associated missense variations or mutations containing TCF4-B proteins. Molecular mass marker is shown on the *right*. *B–D*, EMSA to study the binding of *in vitro*-translated WT or mutant TCF4-B proteins to the 32^{P} -labeled μ E5 E-box (CACTG) containing oligonucleotide as (*B*) homodimers (TCF4-B:TCF4-B), (*C*) intra-TCF4 heterodimers consisting of one WT or mutant TCF4-B and one WT VP16-bHLH subunit (TCF4-B:VP16-bHLH), and (*D*) heterodimers with ASCL1 (TCF4-B:ASCL1). The unlabeled WT (μ E5) or mutated (μ E5m) E-box oligonucleotides were added to the binding mixture for competition where indicated in *italics*. The uncropped EMSA images can be found in [Figure S1](#). bHLH, basic helix-loop-helix; kDa, kilodalton; MMID, Mild-to-moderate intellectual disability; PTHS, Pitt-Hopkins syndrome; RTT-like, Rett-like syndrome; SCZ, Schizophrenia; TCF4, transcription factor 4; WB, Western Blot.

S1). The substitutions associated with SCZ and MMID did not change the ability of TCF4-B homodimers to bind DNA (Figs. 4B and S1). Of the PTHS-associated mutants, R569W showed lower DNA-binding activity than WT TCF4-B, whereas the mutants N585D and A587P completely abrogated the binding of TCF4-B homodimers to the μ E5 E-boxes (Figs. 4B and S1). RTT-like syndrome associated mutant S253R homodimers displayed slightly reduced DNA binding (Figs. 4B and S1).

Next, we studied DNA binding activity in the context of intra-TCF4 heterodimers where one of the dimerization

subunits contains a WT bHLH domain (Figs. 4C and S1). We cotranslated WT or mutant TCF4-B together with the previously described WT bHLH containing VP16-bHLH protein (5) and assayed the binding of the formed dimers to the μ E5 E-boxes. No differences from WT TCF4-B were seen in DNA-binding activity of intra-TCF4 heterodimers for mutants associated with SCZ, MMID, or RTT-like syndrome (Figs. 4C and S1). In case of the PTHS mutants, the DNA-binding ability of mutant N585D was partially rescued in the context of TCF4 intra-heterodimers compared with the mutant homodimers, whereas the mutants R569W and A587P displayed no change

Functional consequences of TCF4 missense substitutions

in DNA-binding activity in response to dimerization with a WT bHLH subunit, showing very low or no DNA binding, respectively (Figs. 4C and S1).

Finally, we studied the effect of the disease-associated mutations on the DNA-binding ability of TCF4-B heterodimers with its interaction partner ASCL1. When translated alone, ASCL1 was not able to bind to the μ E5 E-box sequence, but when cotranslated with TCF4-B, we detected faster moving protein complexes than TCF4-B homodimers (Figs. 4D and S1). All TCF4-B mutants were able to dimerize with ASCL1 and bind DNA as heterodimers (Figs. 4D and S1). Notably, dimerization of the PTHS mutants R569W and N585D with ASCL1 alleviated the negative effect of these mutations on the DNA-binding ability of TCF4-B, and some DNA binding activity was seen for ASCL1 heterodimers with the A587P mutant as well (Figs. 4D and S1).

To summarize, our results show that the RTT-like syndrome and PTHS-associated mutations studied here reduce TCF4 DNA-binding activity to a varying extent. The binding ability is least affected in the case of S253R, more in the case of R569W and N585D, and the most severe effects were seen for mutant A587P. The other studied mutations did not affect DNA binding of TCF4 *in vitro*.

PTHS- and RTT-like syndrome-associated mutations in TCF4 modulate its ability to initiate transcription in HEK293 cells

We assessed the ability of mutant TCF4 proteins to initiate reporter gene transcription using a previously described luciferase reporter assay system (5). Briefly, HEK293 cells were transfected with vectors encoding for WT or mutant TCF4-B, firefly luciferase construct with 12 μ E5 E-boxes (CACCTG) in front of a minimal promoter and *Renilla* luciferase construct with phosphoglycerine kinase (PGK) promoter for normalization.

SCZ- and MMID-associated variations had no effect on the transactivation by TCF4-B (Fig. 5A). Unexpectedly, a 4.1-fold increase ($p < 0.0001$, $n = 3$) in transcription activation was detected for PTHS-associated mutant R569W. The two other PTHS mutants N585D and A587P showed drastically decreased reporter activity, by 8.2-fold and 118-fold, respectively, ($p < 0.0001$, $p < 0.0001$, $n = 3$) (Fig. 5A). The RTT-like syndrome mutant S253R showed a 2.4-fold ($p < 0.0001$, $n = 3$) reduction in transactivation (Fig. 5B).

All the studied TCF4 missense mutants bound DNA *in vitro* when heterodimerized with ASCL1 according to our EMSA results. Therefore, we next studied whether TCF4-B mutants are functional in driving reporter gene expression when coexpressed with ASCL1 in HEK293 cells. Indeed, coexpression of ASCL1 and TCF4-B increased the transcriptional activity of both the WT and mutant TCF4-B proteins. The SCZ- and MMID-associated mutants showed no aberrations in transcription activation when coexpressed with ASCL1 (Fig. 5B). The heterodimers of mutants S253R and A587P displayed a 1.4-fold ($p < 0.0001$, $n = 3$) and 4.8-fold ($p < 0.0001$, $n = 3$) reduction in luciferase signals, respectively (Fig. 5A). The mutants R569W and N585D increased the

transactivation of TCF4-B heterodimers with ASCL1 by 2.2- and 1.6-fold ($p = 0.0005$, 0.0806 , $n = 3$), respectively (Fig. 5A).

To assess the synergistic effects between TCF4 and ASCL1, we calculated cooperation indexes based on the data in Figure 5A for each WT and mutant TCF4-B protein. The SCZ and MMID mutants and the PTHS-mutant R569W had similar synergistic effects with ASCL1 as WT TCF4-B (Fig. 5B). The mutant N585D displayed an increase, whereas mutant S253R showed a slight decrease in the cooperation index compared with the WT TCF4-B. A587P largely differed from all the other studied mutants as it displayed an antagonistic effect on the dimerization of TCF4 with ASCL1 (Fig. 5B).

Taken together, these results show that the studied disease-related mutations outside the bHLH domain (SCZ and MMID) have no effect on the transactivation of TCF4-B, except for the RTT-like syndrome associated mutation S253R. However, mutations within the bHLH region can both increase and reduce the transactivation capability of TCF4-B in HEK293 cells.

Pitt-Hopkins syndrome-associated missense mutations reduce the transactivation ability of TCF4 in primary cortical neurons

TCF4 is an activity regulated transcription factor in neurons (5, 36). To study whether the disease-associated substitutions in TCF4-B cause aberrations in transcription activation in the cultured primary neurons, we transfected rat cortical and hippocampal primary neurons with plasmids encoding for WT or mutant TCF4-B or TCF4-A, ASCL1, firefly luciferase construct carrying 12 μ E5 E-boxes (CACCTG) in front of thymidine kinase (TK) promoter and *Renilla* luciferase construct with PGK promoter for normalization. To study the effect of depolarization on transcription activation, the neurons were treated with 25 mM KCl for 8 h or left untreated.

We have previously shown that two of the SCZ-linked missense variations (P299S and G428V) have a slight effect on the ability of TCF4-B to activate transcription in primary neurons. P299S increased reporter activity in basal conditions and G428V in both basal and depolarized conditions (36). Here, we performed a meta-analysis of our previous data ($n = 5$) and additional experiments carried out in this study ($n = 3$), confirming our results on the variants P299S and G428V, and in addition revealing a mild increase in the transcriptional activity of A315V variant compared with the WT TCF4-B in basal conditions (Fig. S2A).

The RTT-like syndrome and MMID-associated TCF4-B mutants had similar transactivation ability as WT TCF4-B in both basal and depolarized neurons (Fig. 6A). Significant changes in the transcription activation were observed for PTHS mutants (Fig. 6A). In basal conditions, the mutants R569W, N585D and A587P displayed decreased luciferase signals by 1.9-fold, 7.3-fold, and 20-fold, respectively, when compared with the WT TCF4-B ($p = 0.276$, $p < 0.0001$, $p < 0.0001$, $n = 3$) (Fig. 6A). In depolarized conditions, the transcriptional activity of all the studied PTHS mutants remained lower than WT TCF4-B in the same conditions (Fig. 6A).

Functional consequences of TCF4 missense substitutions

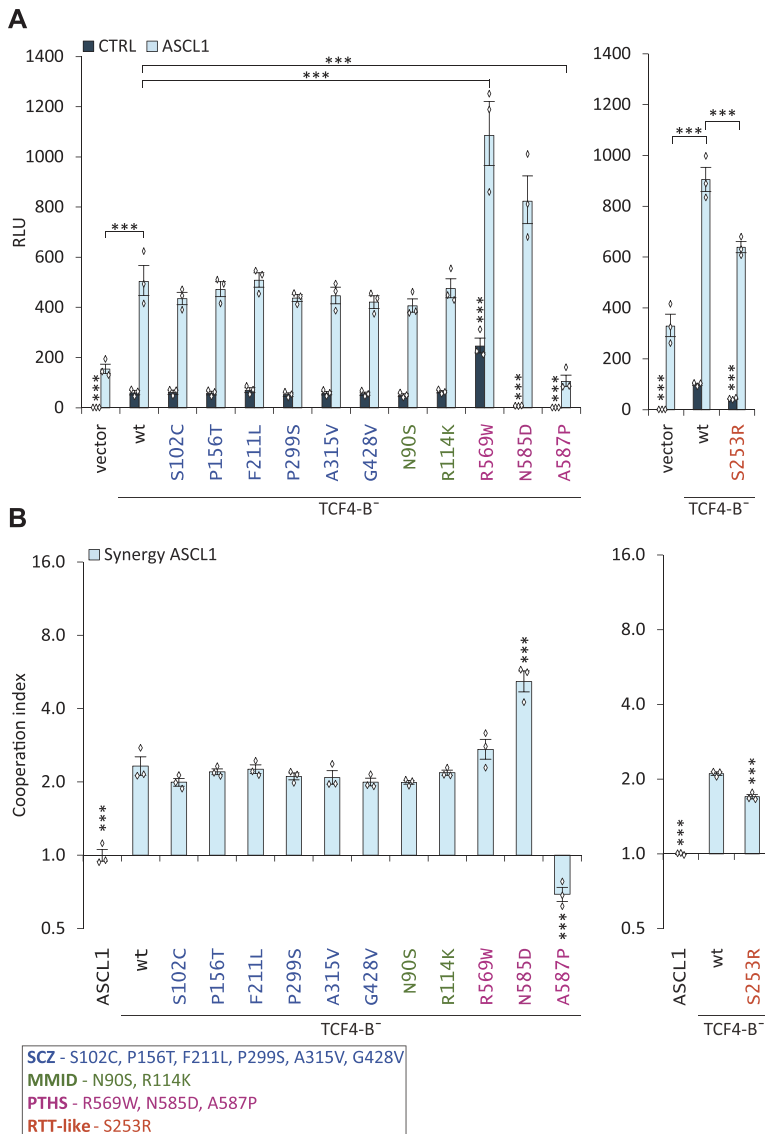


Figure 5. Missense mutations associated with PTHS alter the ability of TCF4 to activate transcription in HEK293 cells. *A*, luciferase reporter assay with WT or mutant TCF4-B. The cells were cotransfected with WT or mutant TCF4-B vectors alone or together with ASCL1, firefly luciferase reporter construct carrying 12 μ E5 E-box regulatory sequences (CACCTG) in front of the minimal promoter and *Renilla* luciferase construct with PGK promoter for normalization. *B*, index of cooperation between TCF4-B (WT or mutant) and ASCL1 calculated from data in (*A*). Four (SCZ-associated variations) or three (MMID-, PTHS, and RTT-like syndrome-associated mutations) independent experiments were performed in duplicates. The luciferase data is presented as fold-induced levels above the signals measured from empty vector-transfected (vector) untreated cells. The error bars indicate SEM. For statistical analysis, one-way ANOVA (SCZ, MMID, and PTHS mutants $F(25, 50) = 258.2$, $p < 0.0001$; RTT-like syndrome mutant $F(1.529, 4.586) = 821.5$, $p < 0.0001$) followed by Holm-Sidak's multiple comparisons test (*A*) or one-way ANOVA (SCZ, MMID, and PTHS mutants $F(12, 24) = 59.52$, $p < 0.0001$; RTT-like syndrome mutant $F(5, 10) = 1166$, $p < 0.0001$) followed by Dunnett's multiple comparisons test (*B*) was used. The individual data points are shown as white diamonds. Statistical significance is shown with asterisks and is relative to the cells overexpressing WT TCF4-B or between the bars connected with lines; ***, $p < 0.001$. MMID, Mild-to-moderate intellectual disability; PTHS, Pitt-Hopkins syndrome; RLU, relative luciferase units; RTT-like, Rett-like syndrome; SCZ, Schizophrenia; TCF4, transcription factor 4.

Next, we asked whether the disease-associated mutations in TCF4 could affect the cooperation of TCF4 with ASCL1 in neurons, as we found to be the case in HEK293 cells. For this,

we first studied the effect of ASCL1 coexpression on the transcriptional activity of WT isoforms TCF4-B and TCF4-A in neurons. We detected no cooperation between ASCL1 and

Functional consequences of TCF4 missense substitutions

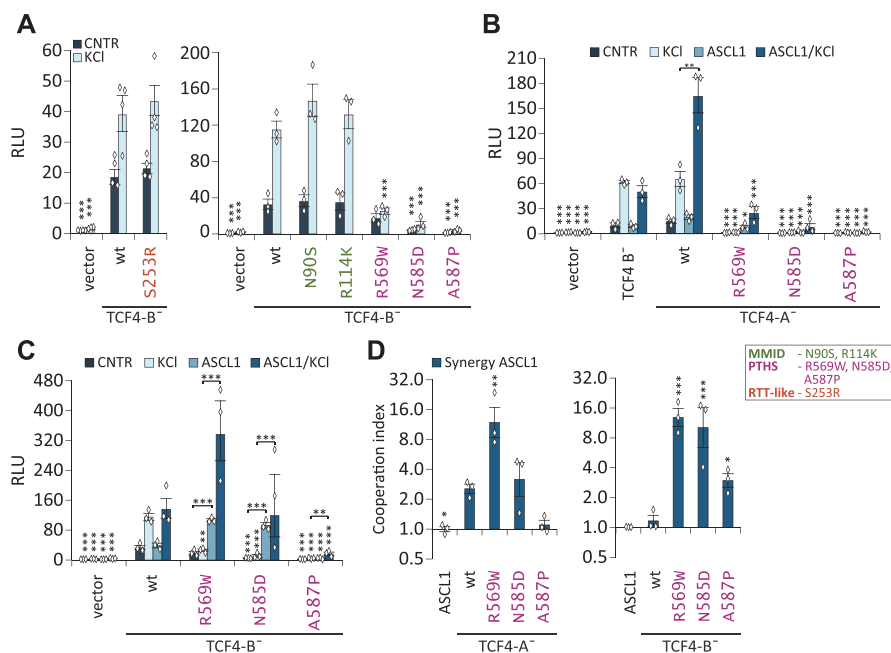


Figure 6. PTHS-associated missense mutations alter the transcriptional activity of TCF4 in rat cortical and hippocampal primary neurons. A–C, luciferase reporter assay with WT and mutant TCF4 in rat cortical and hippocampal primary neurons. The cells were cotransfected with WT or mutant TCF4 vectors alone or with ASCL1 vector, firefly luciferase reporter construct carrying 12 μ E5 E-box regulatory sequences (CACCTG) in front of TK promoter and *Renilla* luciferase construct with PGK promoter for normalization. The transfected neurons were left untreated (CNTR) or treated with 25 mM KCl for 8 h (KCl) to induce membrane depolarization. Luciferase assays were performed with TCF4-B carrying RTT-like syndrome, MMID, or PTHS-associated mutations (A), PTHS mutations containing TCF4-A (B) or TCF4-B (C). D, index of cooperation between ASCL1 and WT or mutant TCF4 calculated from data in (B) or (C). Three independent experiments were performed in duplicates. The luciferase data is presented as fold-induced levels above the signals measured from empty vector-transfected (vector) untreated cells. The error bars indicate SEM. For statistical analysis, one-way ANOVA (A: RTT-like syndrome mutant F (5, 10) = 88.62 p < 0.0001; MMID and SCZ mutants F (13, 26) = 70.52 p < 0.0001) (B: F (27, 54) = 79.12 p < 0.0001) (C: F (19, 38) = 48.36 p < 0.0001) followed by Holm-Sidak's multiple comparisons test (A–C) or one-way ANOVA (TCF4-A F (5, 10) = 22.05 p < 0.0001; TCF4-B F (6, 12) = 31.49 p < 0.0001) Dunnett's multiple comparisons test (D) was used. The individual data points are shown as *white diamonds*. Statistical significance shown with *asterisks* is relative to cells overexpressing WT TCF4-B (A and C), WT TCF4-A (B and C) or between the bars connected with lines; * p < 0.05; ** p < 0.01; *** p < 0.001. MMID, Mild-to-moderate intellectual disability; PTHS, Pitt-Hopkins syndrome; RLU, relative luciferase units; RTT-like, Rett-like syndrome; SCZ, Schizophrenia; TCF4, transcription factor 4.

WT TCF4-B in neurons as the presence of ASCL1 did not change the transcriptional activity of TCF4-B in the conditions studied (Fig. 6, B and C). Contrary to the WT TCF4-B, the transcriptional activity of WT TCF4-A was increased 2.5-fold in depolarized neurons when ASCL1 was coexpressed (p = 0.0196, n = 3) (Fig. 6B), indicating synergistic interaction between the short TCF4 isoform and ASCL1 in neurons (Fig. 6C). We detected no cooperation between ASCL1 and WT TCF4-B in neurons as the presence of ASCL1 did not change the transcriptional activity of TCF4-B in the conditions studied (Fig. 6, B and C).

Based on these results, we decided to study the cooperation of ASCL1 and PTHS mutant proteins in the context of TCF4-A isoform. When PTHS mutant TCF4-A proteins were overexpressed alone in neurons, they showed almost no transcriptional activity in basal or depolarized conditions (Fig. 6B). The coexpression of ASCL1 with PTHS mutants resulted in a slight increase of reporter activity, which remained lower from the activity of the WT TCF4-A protein (Fig. 6B).

We hypothesised that similarly to TCF4-A PTHS mutants, TCF4-B PTHS mutants may also activate transcription differently from the WT protein when coexpressed with ASCL1. Indeed, we observed that in basal conditions, when the TCF4-B mutants R569W and N585D were coexpressed with ASCL1, the reporter signals were increased 2.8- and 2.4-fold, respectively (p < 0.0001, p < 0.0001, n = 3), when compared with the WT TCF4-B. ASCL1 heterodimers with TCF4-B mutant A587P displayed almost no reporter activity in basal conditions (Fig. 6D). In depolarized neurons, the transcriptional activity of ASCL1 heterodimers with TCF4-B was higher in case of R569W, lower in case of A587P, and kept at a similar level to WT in the case of N585D mutation (Fig. 6D). For mutations associated with SCZ, MMID, and RTT-like syndrome, no significant changes in transactivation capability of TCF4-B when coexpressed with ASCL1 were detected in any of the studied conditions (n = 3; Fig. S2, B and C).

The calculated cooperation indexes indicated that R569W mutation increases and A587P abolishes the synergistic effect of ASCL1 with TCF4-A isoform in depolarized neurons

Functional consequences of TCF4 missense substitutions

(Fig. 6D). Differentially from the WT TCF4-B isoform, all studied TCF4-B PTHS mutants cooperate with ASCL1 in neurons (Fig. 6D).

Collectively, these results show that the PTHS-associated mutations modify the transcriptional activity of TCF4 in an isoform- and dimerisation partner-dependent manner. The TCF4-A PTHS mutants displayed more severe deficiencies in transcription activation compared with the TCF4-B PTHS mutants. The coexpression with ASCL1 caused little changes in transcription activation of the TCF4-A PTHS mutants, whereas the effect on TCF4-B PTHS mutants was notable. Interestingly, A587P was the only mutation that caused severely reduced reporter signals in all the studied conditions.

Discussion

Previous studies have shown that PTHS-causing missense mutations within *TCF4* alter the function of the protein by regulating transcription activation, dimerization, intracellular localization, and DNA binding (4, 5, 45, 56). In the present study, we analysed 12 novel missense variations and mutations associated with different diseases which are located within and outside the bHLH domain of TCF4. The functional characterization of the missense mutations in TCF4 from this and

previous studies is summarized in Figure 7. Altogether, the data reveal that disease-associated mutations in TCF4 mainly modulate DNA binding and transactivation of the protein, whereas transport to the cell nucleus is not affected. The most severe effects are seen for PTHS-associated mutations located in the bHLH region of TCF4. Other disease-related mutations in TCF4 have little or no effect on the functionality of TCF4 in the studied conditions (Fig. 7).

TCF4 has been claimed to be one of the master regulators of SCZ, but the exact mechanisms of how changes in *TCF4* contribute to the development of the disease remain unknown (18, 58). We have previously shown that two (P299S and G428V) out of the six *TCF4* missense variations found in SCZ patients (20, 21) increase the transcriptional activity of TCF4-B in primary cultured neurons (36). Here, we confirm these effects and further show a mild increase in the transcriptional activity of A315V variant. We in addition investigated whether the SCZ-associated missense variations in TCF4-B alter any other functions of TCF4 such as intracellular location, formation of active heterodimers, and binding to DNA but detected no changes. This can partly be explained by the fact that the studied variations were not in any known functionally important regions of the TCF4 protein (except for P156T). We hypothesize that even slight changes in TCF4 functions could

Substitution	Related condition	Cellular localization		DNA Binding		Transcription activation			Reference	
		Cell lines	Primary neurons	Homodimers	Heterodimers	Cell lines	Primary neurons	Heterodimers		
N90S	MMID	=	=	=	=	=	=	NA	=	This work
S102C	SCZ	=	=	=	=	=	=	=	=	This work
S102C	SCZ	NA	NA	NA	NA	NA	NA	=	NA	Sepp et al. 2017
R114K	MMID	=	=	=	=	=	=	=	=	This work
P156T	SCZ	=	=	=	=	=	=	=	=	This work
P156T	SCZ	NA	NA	NA	NA	NA	NA	=	NA	Sepp et al. 2017
F211L	SCZ	=	=	=	=	=	=	=	=	This work
F211L	SCZ	NA	NA	NA	NA	NA	NA	=	NA	Sepp et al. 2017
S253R	RTT-like	=	=	↓	=	↓	↓	=	=	This work
P299S	SCZ	=	=	=	=	=	=	↑	=	This work
P299S	SCZ	NA	NA	NA	NA	NA	NA	↑	NA	Sepp et al. 2017
A315V	SCZ	=	=	=	=	=	=	↑	=	This work
A315V	SCZ	NA	NA	NA	NA	NA	NA	=	NA	Sepp et al. 2017
G358V	PTHS	=	=	=	=	=	=	=	NA	Sepp et al. 2012
G358V	PTHS	=	NA	NA	NA	=	↓	NA	NA	Forrest et al. 2012
G428V	SCZ	=	=	=	=	=	=	↑	=	This work
G428V	SCZ	NA	NA	NA	NA	NA	NA	↑	NA	Sepp et al. 2017
D535G	PTHS	=	=	=	=	=	=	↑	NA	Sepp et al. 2012
D535G	PTHS	=	NA	NA	NA	↑	↓	NA	NA	Forrest et al. 2012
R569W	PTHS	=	NP	↓	↓	↑	↑	↓	↓ ^B ↑ ^A	This work
R576Q	PTHS	=	=	↓	↓	↓	↓	↓	NA	Sepp et al. 2012
R576G	PTHS	NA	NA	NA	NA	NA	↓	NA	NA	de Pontual et al. 2009
R578P	PTHS	NP	NA	NA	NA	↓	↓	NA	NA	Forrest et al. 2012
R578H	PTHS	=	NA	↓	↓	↓	↓	↓	NA	Sepp et al. 2012
R800W	PTHS	=	=	↓	↓	↓	↓	↓	NA	Sepp et al. 2012
R800V	PTHS	NP	NA	NA	NA	↓	↓	NA	NA	Forrest et al. 2012
R800Q	PTHS	NA	NA	NA	NA	NA	↓	NA	NA	de Pontual et al. 2009
R582P	PTHS	=	=	↓	↓	↓	↓	↓	NA	Sepp et al., 2012
N585D	PTHS	=	NP	↓	↓	↓	↓	↓	↓ ^B ↑ ^A	This work
A587P	PTHS	=	=	↓	↓	↓	↓	↓	↓	This work
A614V	PTHS	=	=	=	=	↓	=	↓	NA	Sepp et al. 2012
A614V	PTHS	NP	NA	NA	NA	=	↓	NA	NA	Forrest et al. 2012

Figure 7. Summary of the effects of amino acid substitutions on the functionality of TCF4. The effects of TCF4 amino acid substitutions on cellular localization, DNA binding, and transcription activation alone or together with ASCL1 are shown. *Upwards* and *downwards* arrows denote an increase or decrease in protein function, and "=" denotes no change. *Two downwards arrows* mark dominant negative effects. A or B next to arrows indicate a change specific for TCF4-A or TCF4-B. The positions of amino acids affected by disease-related missense mutations and the variations in TCF4-B are shown on the left. For more details, see the legend of Figure 1A. MMID, Mild-to-moderate intellectual disability; NA, not analysed; NP, nuclear punctae; PTHS, Pitt-Hopkins syndrome; RTT-like, Rett-like syndrome; SCZ, Schizophrenia; TCF4, transcription factor 4.

Functional consequences of TCF4 missense substitutions

be part of a complex network of changes, which contribute to the development of this polygenic disorder (17, 59). Alternatively, the effects of SCZ-associated missense variations in *TCF4* could be more pronounced in conditions, cell types, and/or developmental stages not studied here.

Deletions in the 5' coding region of the *TCF4* gene have been described in patients with mild nonsyndromic intellectual disability (8, 12). These mutations lead to reduced dosage of the long-TCF4 isoforms. The mild phenotype of MMID compared with PTHS may be explained by the fact that deletions in the 5' region of the *TCF4* gene do not affect all *TCF4* transcripts, as *TCF4* is transcribed using many alternative 5' exons located throughout the gene (34). Here, we studied two MMID-associated missense mutations in *TCF4* (N90S and R114K), but neither of these had an effect on the functionality of TCF4-B, even though N90S is located in AD1. It may be that mutations in the 5' coding region of *TCF4* cause context-specific effects yet to be discovered and/or are not the only factors underlying the development of MMID in the patients carrying these mutations.

A single mutation (S253R) in the *TCF4* gene has been described in a male patient with RTT-like syndrome. The patient exhibited severe intellectual disability and facial dysmorphisms similar to the phenotype of PTHS (7, 14). In our reporter experiments, mutation S253R was the only TCF4-B mutation outside of the bHLH region that caused significantly reduced reporter activation in HEK293 cells. In addition, EMSA revealed reduced DNA binding of S253R TCF4-B. S253R is located in AD3 of the *TCF4* in a position that interacts directly with the primary core-promoter recognition factor transcription factor TFIID complex (35). Broadly, TFIID can modulate transcriptional activity of RNA polymerase II, but more specifically, it may also stabilize E-proteins through direct interaction, which may be necessary for the binding of coactivators or repressors to TCF4 (35, 60). It is possible that transcription initiation mediated by the mutant S253R is decreased because of reduced DNA binding or impaired binding of coactivators or increased binding of repressors. This however seems to be a cell-type specific effect as we saw no changes in the transcriptional activity of S253R TCF4-B in cultured neurons.

We confirm that missense mutations in the bHLH region of *TCF4* have severe effects on the functioning of the protein, especially mutations affecting the arginine residues. PTHS-associated mutants R569W and N585D in TCF4-B displayed aberrant intranuclear localization and were detected as dots in primary neuron cultures. Previous experiments from our workgroup (5) suggest that the protein aggregates observed as nuclear dots could refer to protein destabilization and misfolding. Therefore, it is possible that the mutations R569W and N585D affect the folding stability of TCF4. Another explanation for the development of intranuclear puncta may be related to the localization of R569W and N585D in the recently described NLS-2 (41). R569W and N585D could cause NLS-2 to dysfunction leaving NLS-1 the only functional NLS. This suggests that NLS-1 is necessary for transport of TCF4 to the nucleus, whereas NLS-2 may be

involved in intranuclear localization as NLS-2 present in TCF4-A is not sufficient to cause strict nuclear localization as seen for TCF4-B. However, the detected aberrant localization of R569W and N585D in the nucleus is cell type-dependent because we detected changes in the localization of these mutants in neurons, and not in HEK293 cells. Whether the localization of TCF4-B mutants as puncta may be due to the lack of a functional NLS-2 remains to be clarified.

The reporter experiments revealed that R569W increased the transcriptional activity of TCF4-B in HEK293 cells, whereas in neurons, the transcriptional activity of R569W TCF4-B was reduced. Interestingly, the coexpression of R569W TCF4-B with ASCL1 resulted in the increased activity relative to WT TCF4-B. This observation is supported by experiments by Forrest *et al.* who showed that PTHS-associated mutations present variable effects depending on the context (56). The detected cell-type specific effects on transcription activation could be due to regulatory partners that are present in HEK293 cells but are absent in rat primary neuron cultures or vice versa. Similar experiments in different cell types, including neurons, derived from human iPSCs would be instrumental in elucidating the cell-type specific activity of TCF4 in the human context.

N585D TCF4-B displayed impaired transcription activation, whereas its heterodimers with ASCL1 were transcriptionally more active than those of the WT TCF4-B. This implicates that N585D disrupts DNA binding of homodimers, whereas it enhances the formation of heterodimers or heterodimer binding of TCF4-B. This could be caused by structural changes caused by the mutation as N585 forms hydrogen bonds with DNA backbone in TCF4 homodimers. N585D could in part disrupt the α -helical structure important for homodimer DNA-binding, and the other bHLH dimerization partners rescue the damaging effects. This is also in accordance with our EMSA results because we saw that the dimerization of N585D with ASCL1 alleviates the negative effect of the mutation on DNA binding to some extent.

We have previously shown that a mutation outside of the basic DNA-binding region but within the bHLH domain (A614V) almost completely abrogates DNA binding and transcription activation capability of TCF4. However, the negative effect of that mutation was rescued by dimerization with ASCL1 (5). Here, we show that mutation A587P, which is in the first helix of the bHLH region, completely abrogates the transcriptional activity of TCF4-B and TCF4-A. The negative effect of A587P on transcription activation was only partly rescued by interaction with ASCL1 in HEK293 cells, but not in primary neuron cultures. Our analysis shows that mutation A587P does not interfere with nuclear localization but rather impairs DNA binding of mutant A587P dimers. A587 packs against residue L611 in helix 2 of the opposite monomer, therefore, the mutation in this region can impair the formation of functionally active dimers. Combined, our results indicate that A587P may act as a dominant-negative mutation, which results in the formation of inactive dimers that cannot bind to DNA.

Functional consequences of TCF4 missense substitutions

Different TCF4 isoforms present varying transactivation capabilities between cell types (34, 36). Here, we show that amino acid substitutions in TCF4 lead to variable effects depending on the protein isoform. Contrary to the respective TCF4-B mutants, R569W and N585D TCF4-A mutants displayed drastically lower transcription activation in cultured primary neurons than the WT protein. This may arise from the differences in the presence of protein domains. TCF4-B carries functional protein regions AD1, CE, and NLS-1, which are not present in the short isoform TCF4-A. It would be of interest to elucidate whether deleting one or more of these functional regions in TCF4-B mutant proteins results in similar effects on transcription activation as seen for TCF4-A mutants.

Our results indicate that the heterodimerization of TCF4 mutants may both decrease and increase the transcriptional activity of TCF4. As TCF4 dimerization partners are differentially expressed during development of the central nervous system (27), it is possible that the dysregulation of TCF4's activity in response to the studied missense mutations is much more complex and cannot be modelled in cell cultures. There may exist a regulatory mechanism that defines which dimerization partners can interact with TCF4 during development that is impaired in response to missense mutations in TCF4. As TCF4 expression levels in the cerebral cortex of rodents and humans are highest around birth (42, 61) it would be interesting to study the effects of TCF4 missense mutations on the functionality of TCF4 *in vivo* at perinatal stages.

The effects of SCZ, MMID, and RTT-like syndrome associated mutations and variations on the functioning of TCF4 were only studied in the context of TCF4-B which may be the reason why we saw mild or no effects of SCZ, MMID, and RTT-like syndrome associated mutations and variations on the functioning of TCF4. However, 18 N-terminally distinct protein isoforms with different transcriptional activities are encoded by the human *TCF4* gene, of which the most studied long-TCF4 isoform is isoform TCF4-B (34, 36). The mutations N90S, R114K and P156T are present in the major TCF4 protein isoforms TCF4-B, -C, -E, -F, -D, and -G, expressed in the central nervous system (34). All the other studied mutations and variations are in the C-terminal end of TCF4 and are present in all the TCF4 protein isoforms. Because our experiments focused on the effect of mutations and variations on TCF4-B and TCF4-A, it remains unknown whether the studied aberrations could present varying effects depending on different TCF4 protein isoforms. In addition, Sepp *et al.* (5, 36) and Forrest *et al.* (56) have studied the effects of TCF4 missense mutations on DNA binding and transcription activation using the CACCTG E-box sequence. A ChIP-seq study in SH-SY5Y cells has shown that TCF4-A and TCF4-B display enrichment for the CATCTG E-box sequence (62). Still, the binding specificity of TCF4 to various E-box combinations is regulated by dimerization partners. For example, the preferred E-box motif of ASCL1 is CAGCTG (63). It remains to be elucidated whether changes in E-box sequences can affect the activity of TCF4. Our experiments were carried out using cell cultures, but the use of different model systems such as

differentiated stem cells or organoids in addition to *in vivo* studies could provide more insight on how different mutations affect TCF4 and whether the E-box binding specificity is affected by amino acid substitutions in TCF4.

To conclude, the results of this and previous studies show that missense mutations in TCF4 affect transcription activation and DNA binding of the protein, which can be partially rescued by the formation of heterodimers with dimerization partners such as ASCL1. Further studies are needed to understand how or whether SCZ and MMID associated missense variations and mutations in TCF4 may contribute to disease pathogenesis. The most severe effects on protein function, including complete loss-of-function, were seen for PTHS-associated mutations in the bHLH domain of TCF4. Together with previous work, our study gives an overview of the functional consequences of disease-related missense mutations in *TCF4* providing a basis for the interpretation of *TCF4* missense variants and their potential pathogenicity.

Experimental procedures

Constructs

Schizophrenia-associated missense variations (S102C, P156T, F211L, P299S, A315V, and G428V), RTT-like syndrome associated missense mutation (S253R), mild-to-moderate intellectual disability associated missense mutations (N90S and R114K) and Pitt-Hopkins syndrome associated missense mutations (R569W, N585D, and A587P) were introduced into pcDNA3.1/TCF4-B vectors (34) *via* site-directed mutagenesis with complementary oligonucleotides (Table S1). For all constructs, the initial CMV promoter was substituted with elongation factor 1 α promoter, as described before (36). All the mutations were mapped and numbered according to the TCF4-B⁺ NCBI reference sequence NP_001077431.1.

For pcDNA3.1/EF1a/E2-TCF4-B vectors, E2-tag was cloned from pQM/CMV/E2_N (Icosagen) and inserted in front of *TCF4* coding sequence. pcDNA3.1/EF1a/TCF4-A has been described (36). pcDNA3.1/EF1a/TCF4-A vectors with missense mutations G428V, R569W, N585D, and A587P were generated from corresponding pcDNA3.1/EF1a/TCF4-B vectors by replacing the fragment between the KpnI and MluI restriction sites (Thermo Scientific). All the vectors were verified by sequencing.

pcDNA3.1/ASCL1, pcDNA3.1/NEUROD2-E2, pACT-bHLH, and luciferase reporter assay constructs pGL4.29 [luc2P/12 μ E5/min/Hygro] with 12 E-boxes in front of minimal promoter, pGL4.29 [luc2P/12 μ E5/TK/Hygro] with 12 E-boxes in front of thymidine kinase promoter, and pGL4.83 [hRlucP/PKG1/Puro] with phosphoglycerate kinase 1 promoter have been previously described (5, 34, 36).

Protein structure visualization

TCF4 bHLH model was previously generated using E-protein E47 (TCF3 isoform) crystal structure coordinates (5). Protein Data Bank compatible files and PyMOL Molecular Graphics System session files were used for the visualization of

Functional consequences of TCF4 missense substitutions

the protein structure. The analysis of protein-DNA and protein-protein interactions was made in PyMOL Molecular Graphics System (version 1.7.4.5 trial for educational use). The ribbon drawing of TCF4-B* (NP_001077431.1) bHLH homodimer structure was generated using YASARA View (version 17.4.17, YASARA Biosciences GmbH) (64).

Cell culture and transfection

Human embryonic kidney HEK293 cells were grown in Minimal Essential Medium with Earle's Salts (PAA Laboratories) supplemented with 10% Fetal Bovine Serum (SeraPlus and PAN Biotech), 100 U/ml penicillin, and 0.1 mg/ml streptomycin (Gibco) at 37 °C in 5% CO₂. Rat primary cortical and hippocampal mixed neuronal cultures were plated from embryonic day 20 to 21 Sprague-Dawley rat fetuses and maintained, as described previously (65). The protocols involving animals were approved by the ethics committee of animal experiments at the Ministry of Agriculture of Estonia (Permit Number: 45). All the experiments were performed in accordance with the relevant guidelines and regulations.

The HEK293 cells were transfected using polyethylenimine (Sigma-Aldrich) with polyethylenimine to DNA ratio 2:1. For Western blot analysis, the cells grown on 6-well plates were transfected with 1.8 µg of TCF4-B encoding vector and with 0.2 µg of pEGFP vector to evaluate transfection efficiency. For luciferase reporter assays, the cells grown on 48-well plates were transfected with 0.1875 µg of effector protein construct(s) (TCF4-B alone or TCF4-B and ASCL1 together), 0.1875 µg of firefly luciferase construct pGL4.29[luc2P/12µE5/min/Hygro], and 0.01 µg of *Renilla* luciferase construct pGL4.83[hRlucP/PGK1/Puro]. In case of cotransfection of effector protein constructs, pcDNA3.1/EF1a/TCF4-B and pcDNA3.1/ASCL1 were added in a ratio of 2:1. For nuclear redirection assay, 0.2 µg of pcDNA3.1/EF1a/TCF4-A and 0.1 µg pcDNA3.1/ASCL1 or pcDNA3.1/NEUROD2-E2 were transfected to the HEK293 cells grown on 48-well plates.

The neurons plated on 48-well plates were transfected on 6 days *in vitro* using Lipofectamine 2000 (Invitrogen) with a reagent to DNA ratio 2:1. For luciferase reporter assays, 120 ng of effector protein construct(s) (TCF4-B alone or TCF4-B and ASCL1 together), 60 ng of firefly luciferase construct pGL4.29[luc2P/12µE5/TK/Hygro] and 20 ng of *Renilla* luciferase pGL4.83[hRlucP/PGK1/Puro] were used. In case of cotransfection of effector protein constructs, pcDNA3.1/EF1a/TCF4-B and pcDNA3.1/ASCL1 were added in a ratio of 2:1. 2 days posttransfection, and the neuronal cultures were treated with 25 mM KCl for 8 h where indicated. For protein localization assays, 0.2 µg of pcDNA3.1/EF1a/TCF4 (HEK293 cells) or pcDNA3.1/EF1a/E2-TCF4 (neuronal cultures) plasmids were transfected to the cells grown on 48-well plates using Lipofectamine 2000 (Invitrogen), with a reagent to DNA ratio 2:1.

Cell extracts and Western blotting

The cells were lysed 48 h posttransfection in radio-immunoprecipitation assay buffer (150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.2% sodium dodecyl sulfate,

50 mM Tris-HCl, 1× Protease inhibitor cocktail Complete (Roche), and 1 mM DTT). The protein concentrations were measured with Pierce BCA Protein Assay Kit (Thermo Scientific).

Equal amounts of protein were separated in 8 to 10% SDS-polyacrylamide gel and transferred to Immobilon-P polyvinylidene fluoride membrane (Millipore). The membrane was blocked with 5% skimmed milk in PBS-0.1% Tween 20 (PBST) (Sigma-Aldrich) at room temperature and incubated with primary and secondary antibody solutions in 2% skimmed milk in PBST. The antibodies were used in the following dilutions: rabbit polyclonal anti-TCF4 (CeMines) 1:1000, mouse monoclonal anti-E2 (5E11, Icosagen,) 1:5000, mouse monoclonal anti-GAPDH (MAB374, Millipore) 1:4000, horseradish peroxidase-conjugated goat anti-mouse/rabbit IgG (Thermo Scientific) 1:5000. For chemiluminescent reaction, SuperSignal West Femto Maximum Sensitivity Substrate Kit (Thermo Scientific) was used, and the reaction was visualized with ImageQuant LAS4000 camera system (GE Healthcare).

Immunocytochemistry

For protein localization and nuclear redirection assays, the cells were grown on poly-L-lysine coated cover slips. The cells were fixed 48 h posttransfection in 4% paraformaldehyde (Applichem) in PBS for 15 min, then treated with 50 mM ammonium chloride (Scharlau) in PBS, and permeabilized with 0.5% Triton X-100 (Amresco) solution in PBS. After each reaction, cover slips were washed with PBS. The cells were blocked with 2% bovine serum albumin (BSA, Naxo) in PBS for 1 h at room temperature and incubated first with primary and then with secondary antibodies in 0.2% BSA-PBS solution. The antibodies were diluted as follows: rabbit polyclonal anti-TCF4 (CeMines) 1:200, mouse monoclonal anti-E2 (5E11, Icosagen,) 1:1000, mouse monoclonal anti-MASH1 (ASCL1) (24B72D11.1, BD Pharmingen) 1:200, mouse monoclonal anti-tubulin-β (E7, DSHB) 1:1200, rabbit polyclonal anti-tubulin-β III (T2200, Sigma-Aldrich) 1:400, and Alexa Fluor 488- or Alexa Fluor 568-conjugated F(ab')₂ fragment of goat anti-mouse/rabbit IgG (Invitrogen) 1:2000. 1 µg/ml of Hoechst 33342 was included in secondary antibody solution to visualize nuclei. The cover slips were washed with PBST after both reactions. Finally, the cover slips were washed with water and mounted with Mowiol 4-88 mounting medium (Polysciences, Inc). The whole area of the samples was analysed by confocal microscopy (LSM 510 Duo Zeiss) using lasers Argon/2 (for Alexa Fluor 488, 488 nm), DPSS 561-10 (for Alexa Fluor 568, 568 nm) and Diode 405-50 (for Hoechst 33342, 405 nm), and objective lenses PlanApo 63× oil or PlanApo 100× oil (NA 1.4). Picture panels were made using Imaris (version 6.4.2, Bitplane), Photoshop, and Illustrator (Adobe) software. All the experiments were performed twice.

In vitro translation and EMSA

TnT T7 Quick Coupled Transcription/Translation System (Promega) was used according to manufacturer's instructions, using unlabeled methionine, to produce *in vitro*-translated

Functional consequences of TCF4 missense substitutions

proteins. For cotranslations, construct ratio of 2:1 was used in favor for the longer product-pcDNA3.1/EF1a/TCF4-B:pcDNA3.1/ASCL1 2:1 and pcDNA3.1/EF1a/TCF4-B:pACT-bHLH 2:1. Previously described μ E5 and mutated μ E5 oligonucleotides (5) were 32 P labeled with T4-polymerase kinase (Thermo Scientific). The sense and antisense oligonucleotides were annealed in annealing buffer (62.5 mM NaCl and 1.25 mM EDTA) in a 95 °C water bath, which was left to cool to room temperature overnight. The binding reaction was done in a 15 μ l reaction buffer (20 mM Hepes-KOH, pH 7.9, 50 mM KCl, 5% glycerol, 1 mM EDTA, 1 mM DTT, 13.3 ng/ μ l poly(dI-dC), and 0.1 mg/ml BSA), which included 1 μ l of *in vitro*-translated protein mixture and 0.1 pmol of 32 P-labeled μ E5 oligonucleotides. Specificity of binding to μ E5 boxes was tested by adding 1 pmol of unlabeled- μ E5 oligonucleotides to the reaction mixture. The DNA-protein complexes were resolved in a 5% nondenaturing polyacrylamide gel containing 0.25 \times TBE, 0.01% NP-40, and 2.5% glycerol and visualized by autoradiography.

Luciferase reporter assay

For luciferase reporter assays, the cells were lysed 48 h posttransfection in Passive Lysis Buffer (Promega) and kept at -80 °C overnight. The reporter luminescence signals were obtained in duplicates with Dual-Glo Luciferase assay kit (Promega) according to manufacturer's protocol and measured with GENios pro (Tecan) plate reader. All the experiments were repeated 3 times or more where indicated. In addition, the data from Sepp *et al.* 2017 (n = 5) (36) and this work (n = 3) was combined for more statistical power (n = 8) to better describe the effects of SCZ-related missense mutations on the functionality of TCF4. For data analysis, the background signals were subtracted, and firefly luciferase signals were normalized to *Renilla* luciferase signals. The data were log-transformed and auto-scaled. GraphPad Prism 7 (version 7.00) was used for one-way or two-way ANOVA statistical analysis followed by Dunnett's or Holm-Sidak's multiple comparisons test. The data were then back transformed into the original scale for graphical representation. Statistical significance was calculated for comparisons with WT TCF4, unless indicated otherwise. The cooperation index was calculated, as described previously (5, 66). An index value of higher or lower than 1 implies synergism or antagonism, respectively. For statistical analysis, one-way ANOVA followed by Dunnett's post hoc test was used.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supporting information—This article contains supporting information.

Acknowledgments—We thank Epp Väli for technical assistance and Laura Tamberg and Anastassia Šubina for discussions and critical reading of the article. We 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. This study was supported by Estonian Research Council (institutional research funding IUT19-18 and grant PRG805), European Union through the European Regional Development Fund (Project no. 2014-2020.4.01.15-0012) and H2020-MSCA-RISE-2016 (EU734791), Pitt Hopkins Research Foundation (Grants no. 8 and 21) and Million Dollar Bike Ride Pilot Grant Program for Rare Disease Research at the University of Pennsylvania Orphan Disease Center (grants MDBR-16-122-PHP and MDBR-17-127-Pitt Hopkins).

Author contributions—A. S., K. R., K. N., M. S., and T. T. conceptualization; A. S., K. R., and J. T. visualization; K. N., M. S., and T. T. supervision; M. S. and T. T. project administration; M. S. and T. T. funding acquisition; J. T. formal analysis; A. S. and K. R. investigation; A. S., K. R., M. S., and T. T. writing—original draft.

Conflict of interest—The authors declare that they have no conflict of interest with the contents of the article.

Abbreviations—The abbreviations used are: AD, activation domain; ASCL1, achaete-scute homolog 1; bHLH, basic helix-loop-helix; BSA, bovine serum albumin; CE, conserved element; MMID, Mild-to-moderate intellectual disability; NLS, Nuclear localization signal; PBST, PBS with Tween 20; PGK, phosphoglycerine kinase; PTHS, Pitt-Hopkins syndrome; RTT-like, Rett-like syndrome; SCZ, Schizophrenia; TCF4, Transcription factor 4; TK, thymidine kinase.

References

1. Quednow, B. B., Ettinger, U., Mössner, R., Rujescu, D., Giegling, I., Collier, D. A., Schmechtig, A., Kühn, K.-U., Möller, H.-J., Maier, W., Wagner, M., and Kumari, V. (2011) The schizophrenia risk allele C of the TCF4 rs9960767 polymorphism disrupts sensorimotor gating in schizophrenia spectrum and healthy volunteers. *J. Neurosci.* **31**, 6684–6691.
2. Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N. P., Lyonnet, S., Munnich, A., and Colleaux, L. (2007) Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* **80**, 988–993.
3. Brockschmidt, A., Todt, U., Ryu, S., Hoischen, A., Landwehr, C., Birnbaum, S., Frenck, W., Radlwimmer, B., Lichter, P., Engels, H., Driever, W., Kubisch, C., and Weber, R. G. (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.
4. Zweier, C., Peippo, M. M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., Reardon, W., Saraiva, J., Cabral, A., Göhring, I., Devriendt, K., de Ravel, T., Bijlsma, E. K., Hennekam, R. C. M., Orrico, A., *et al.* (2007) Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt–Hopkins syndrome). *Am. J. Hum. Genet.* **80**, 994–1001.
5. 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.
6. Whalen, S., Héron, D., Gaillon, T., Moldovan, O., Rossi, M., Devillard, F., Giuliano, F., Soares, G., Mathieu-Dramard, M., Afenjar, A., Charles, P., Mignot, C., Burglen, L., Van Maldergem, L., Piard, J., *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.

Functional consequences of TCF4 missense substitutions

- Zollino, M., Zweier, C., Van Balkom, I. D., Sweetser, D. A., Alaimo, J., Bijlsma, E. K., Cody, J., Elsea, S. H., Giurgea, I., Macchiaiolo, M., Smigiel, R., Thibert, R. L., Benoist, L., Clayton-Smith, J., De Winter, C. F., et al. (2019) Diagnosis and management in Pitt-Hopkins syndrome: First international consensus statement. *Clin. Genet.* **95**, 462–478
- Kalscheuer, V. M., Feenstra, I., Van Ravenswaaij-Arts, C. M. A., Smeets, D. F. C. M., Menzel, C., Ullmann, R., Musante, L., and Ropers, H.-H. (2008) Disruption of the TCF4 gene in child with mental retardation but without the classical Pitt-Hopkins syndrome. *Am. J. Med. Genet. A* **146A**, 2053–2059
- Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Ende, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Donato, N. D., Dufke, A., Cremer, K., Hempel, M., Horn, D., Hoyer, J., et al. (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *Lancet* **380**, 1674–1682
- Hamdan, F. F., Daoud, H., Patry, L., Dionne-Laporte, A., Spiegelman, D., Dobrzyńska, S., Rouleau, G. A., and Michaud, J. L. (2013) Parent-child exome sequencing identifies a *de novo* truncating mutation in TCF4 in non-syndromic intellectual disability. *Clin. Genet.* **83**, 198–200
- Hamdan, F. F., Srour, M., Capo-Chichi, J.-M., Daoud, H., Nassif, C., Patry, L., Massicotte, C., Ambalavanan, A., Spiegelman, D., Diallo, O., Henrion, E., Dionne-Laporte, A., Fougerat, A., Pshezhetsky, A. V., Venkateswaran, S., et al. (2014) *De novo* mutations in moderate or severe intellectual disability. *PLoS Genet.* **10**, e1004772
- 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
- Maduro, V., Pusey, B. N., Cherukuri, P. F., Atkins, P., du Souich, C., Rupp, R., Limbos, M., Adams, D. R., Bhatt, S. S., Eydoux, P., Links, A. E., Lehman, A., Malicdan, M. C., Mason, C. E., Morimoto, M., et al. (2016) Complex translocation disrupting TCF4 and altering TCF4 isoform expression segregates as mild autosomal dominant intellectual disability. *Orphanet J. Rare Dis.* **11**, 62
- Srivastava, S., Desai, S., Cohen, J., Smith-Hicks, C., Barañano, K., Fatemi, A., and Naidu, S. (2018) Monogenic disorders that mimic the phenotype of Rett syndrome. *Neurogenetics* **19**, 41–47
- Lotan, A., Fenckova, M., Bralten, J., Alttou, A., Dixon, L., Williams, R. W., and van der Voet, M. (2014) Neuroinformatic analyses of common and distinct genetic components associated with major neuropsychiatric disorders. *Front. Neurosci.* **8**, 331
- Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., Lin, D. Y., Duan, J., Ophoff, R. A., Andreassen, O. A., Scolnick, E., Cichon, S., Clair, D. S., Corvin, A., Gurling, H., et al. (2011) Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43**, 969–978
- Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K.-H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., Collier, D. A., Huang, H., Pers, T. H., Agart, I., Agerbo, E., Albus, M., Alexander, M., et al. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427
- Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O. P. H., Mors, O., Mortensen, P. B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., et al. (2009) Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747
- Steinberg, S., de Jong, S., Andreassen, O. A., Werge, T., Børglum, A. D., Mors, O., Mortensen, P. B., Gustafsson, O., Costas, J., Pietiläinen, O. P. H., Demontis, D., Papiol, S., Huttenlocher, J., Mattheisen, M., Breuer, R., et al. (2011) Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum. Mol. Genet.* **20**, 4076–4081
- Basmanav, F. B., Forstner, A. J., Fier, H., Herms, S., Meier, S., Degenhardt, F., Hoffmann, P., Barth, S., Fricker, N., Strohmaier, J., Witt, S. H., Ludwig, M., Schmael, C., Moebus, S., Maier, W., et al. (2015) Investigation of the role of TCF4 rare sequence variants in schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **168**, 354–362
- Hu, X., Zhang, B., Liu, W., Paciga, S., He, W., Lanz, T. A., Kleiman, R., Dougherty, B., Hall, S. K., McIntosh, A. M., Lawrie, S. M., Power, A., John, S. L., Blackwood, D., St Clair, D., et al. (2014) A survey of rare coding variants in candidate genes in schizophrenia by deep sequencing. *Mol. Psychiatry* **19**, 858–859
- Hellwig, M., Lauffer, M. C., Bockmayr, M., Spohn, M., Merk, D. J., Harrison, L., Ahlfeld, J., Kitowski, A., Neumann, J. E., Ohli, J., Holdhof, D., Niesen, J., Schoof, M., Kool, M., Kraus, C., et al. (2019) TCF4 (E2-2) harbors tumor suppressive functions in SHH medulloblastoma. *Acta Neuropathol.* **137**, 657–673
- Kool, M., Jones, D. T. W., Jäger, N., Northcott, P. A., Pugh, T. J., Hovestadt, V., Piro, R. M., Esparza, L. A., Markant, S. L., Remke, M., Milde, T., Bourdeaut, F., Ryzhova, M., Sturm, D., Pfaff, E., et al. (2014) Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25**, 393–405
- Ephrussi, A., Church, G. M., Tonegawa, S., and Gilbert, W. (1985) B lineage-specific interactions of an immunoglobulin enhancer with cellular factors *in vivo*. *Science* **227**, 134–140
- Massari, M. E., and Murre, C. (2000) Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* **20**, 429–440
- Murre, C. (2019) Helix-loop-helix proteins and the advent of cellular diversity: 30 years of discovery. *Genes Dev.* **33**, 6–25
- 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
- Ruzinova, M. B., and Benezra, R. (2003) Id proteins in development, cell cycle and cancer. *Trends Cell Biol.* **13**, 410–418
- 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
- Schoof, M., Hellwig, M., Harrison, L., Holdhof, D., Lauffer, M. C., Niesen, J., Virdi, S., Indenbirken, D., and Schüller, U. (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
- Sobrado, V. R., Moreno-Bueno, G., Cubillo, E., Holt, L. J., Nieto, M. A., Portillo, F., and Cano, A. (2009) The class I bHLH factors E2-2A and E2-2B regulate EMT. *J. Cell Sci.* **122**, 1014–1024
- Persson, P., Jögi, A., Grynfeld, A., Pählman, S., and Axelson, H. (2000) HASH-1 and E2-2 are expressed in human neuroblastoma cells and form a functional complex. *Biochem. Biophys. Res. Commun.* **274**, 22–31
- 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 Dev.* **9**, 23
- 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
- Chen, W.-Y., Zhang, J., Geng, H., Du, Z., Nakada, T., and Roeder, R. G. (2013) A TAF4 coactivator function for E proteins that involves enhanced TFIID binding. *Genes Dev.* **27**, 1596–1609
- Sepp, M., Vihma, H., Nurm, K., Urb, M., Page, S. C., Roots, K., Hark, A., Maher, B. J., Pruunsild, P., and Timmusk, T. (2017) The intellectual disability and schizophrenia associated transcription factor TCF4 is regulated by neuronal activity and protein kinase A. *J. Neurosci.* **37**, 10516–10527
- 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
- Markus, M., Du, Z., and Benezra, R. (2002) Enhancer-specific modulation of E protein activity. *J. Biol. Chem.* **277**, 6469–6477
- 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
- Shirakata, M., and Paterson, B. M. (1995) The E12 inhibitory domain prevents homodimer formation and facilitates selective heterodimerization with the MyoD family of gene regulatory factors. *EMBO J.* **14**, 1766–1772
- Greb-Markiewicz, B., Kazana, W., Zarębski, M., and Ożyhar, A. (2019) The subcellular localization of bHLH transcription factor TCF4 is

Functional consequences of TCF4 missense substitutions

- mediated by multiple nuclear localization and nuclear export signals. *Sci. Rep.* **9**, 15629
42. Jung, M., Häberle, B. M., Tschakowsky, T., Wittmann, M.-T., Balta, E.-A., Stadler, V.-C., Zweier, C., Dörfler, A., Gloeckner, C. J., and Lie, D. C. (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
 43. 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
 44. Page, S. C., Hamersky, G. R., Gallo, R. A., Rannals, M. D., Calcaterra, N. E., Campbell, M. N., Mayfield, B., Briley, A., Phan, B. N., Jaffe, A. E., and Maher, B. J. (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
 45. Pontual, L. de, Mathieu, Y., Golzio, C., Rio, M., Malan, V., Boddaert, N., Soufflet, C., Picard, C., Durandy, A., Dobbie, A., Heron, D., Isidor, B., Motte, J., Newbury-Ecob, R., Pasquier, L., et al. (2009) Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. *Hum. Mutat.* **30**, 669–676
 46. 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
 47. Kennedy, A. J., Rahn, E. J., Paulukaitis, B. S., Savell, K. E., Kordasiewicz, H. B., Wang, J., Lewis, J. W., Posey, J., Strange, S. K., Guzman-Karlsson, M. C., Phillips, S. E., Decker, K., Motley, S. T., Swayze, E. E., Ecker, D. J., et al. (2016) Tcf4 regulates synaptic plasticity, DNA Methylation, and memory function. *Cell Rep.* **16**, 2666–2685
 48. Badowska, D. M., Brzózka, M. M., Kannaiyan, N., Thomas, C., Dibaj, P., Chowdhury, A., Steffens, H., Turck, C. W., Falkai, P., Schmitt, A., Papiol, S., Scheuss, V., Willig, K. I., Martins-de-Souza, D., Rhee, J. S., et al. (2020) Modulation of cognition and neuronal plasticity in gain- and loss-of-function mouse models of the schizophrenia risk gene Tcf4. *Transl. Psychiatry* **10**, 343
 49. 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
 50. Rannals, M. D., Hamersky, G. R., Page, S. C., Campbell, M. N., Briley, A., Gallo, R. A., Phan, B. N., Hyde, T. M., Kleinman, J. E., Shin, J. H., Jaffe, A. E., Weinberger, D. R., and Maher, B. J. (2016) Psychiatric risk gene transcription factor 4 regulates intrinsic excitability of prefrontal neurons via repression of SCN10a and KCNQ1. *Neuron* **90**, 43–55
 51. 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
 52. 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
 53. Tamberg, L., Jaago, M., Säälk, K., Sirp, A., Tuvikene, J., Shubina, A., Kiir, C. S., Nurm, K., Sepp, M., Timmusk, T., and Palgi, M. (2020) Daughterless, the Drosophila orthologue of TCF4, is required for associative learning and maintenance of the synaptic proteome. *Dis. Model. Mech.* <https://doi.org/10.1242/dmm.042747>
 54. [preprint] Ruzicka, W. B., Mohammadi, S., Davila-Velderrain, J., Subburaju, S., Tso, D. R., Hourihan, M., and Kellis, M. (2020) Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. *medRxiv*. <https://doi.org/10.1101/2020.11.06.20225342>
 55. Wirgines, K. V., Sønderby, I. E., Haukvik, U. K., Mattingsdal, M., Tesli, M., Athanasiu, L., Sundet, K., Rossberg, J. I., Dale, A. M., Brown, A. A., Agartz, I., Melle, I., Djurovic, S., and Andreassen, O. A. (2012) TCF4 sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders. *Transl. Psychiatry* **2**, e112
 56. 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
 57. Marangi, G., Ricciardi, S., Orteschi, D., Tenconi, R., Monica, M. D., Scarano, G., Battaglia, D., Lettori, D., Vasco, G., and Zollino, M. (2012) Proposal of a clinical score for the molecular test for Pitt-Hopkins syndrome. *Am. J. Med. Genet. A* **158A**, 1604–1611
 58. Doostparast Torshizi, A., Armoskus, C., Zhang, H., Forrest, M. P., Zhang, S., Souaiaia, T., Evgrafov, O. V., Knowles, J. A., Duan, J., and Wang, K. (2019) Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. *Sci. Adv.* **5**, eaau4139
 59. International Schizophrenia Consortium, Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., and Sklar, P. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752
 60. 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
 61. 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**, 67
 62. 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
 63. Castro, D. S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., Drechsel, D., Lebel-Potter, M., Garcia, L. G., Hunt, C., Dolle, D., Bithell, A., Ettwiller, L., Buckley, N., and Guillemot, F. (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
 64. Krieger, E., and Vriend, G. (2014) YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics* **30**, 2981–2982
 65. Esvald, E.-E., Tuvikene, J., Sirp, A., Patil, S., Bramham, C. R., and Timmusk, T. (2020) CREB family transcription factors are major mediators of BDNF transcriptional autoregulation in cortical neurons. *J. Neurosci.* **40**, 1405–1426
 66. Chang, W., Zhou, W., Theill, L. E., Baxter, J. D., and Schaufele, F. (1996) An activation function in Pit-1 required selectively for synergistic transcription. *J. Biol. Chem.* **271**, 17733–17738
 67. Zweier, C., Sticht, H., Bijlsma, E. K., Clayton-Smith, J., Boonen, S. E., Fryer, A., Grealley, M. T., Hoffmann, L., den Hollander, N. S., Jongmans, M., Kant, S. G., King, M. D., Lynch, S. A., McKee, S., Midro, A. T., et al. (2008) Further delineation of Pitt-Hopkins syndrome: Phenotypic and genotypic description of 16 novel patients. *J. Med. Genet.* **45**, 738–744
 68. Takano, K., Lyons, M., Moyes, C., Jones, J., and Schwartz, C. E. (2010) Two percent of patients suspected of having Angelman syndrome have TCF4 mutations. *Clin. Genet.* **78**, 282–288

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 Kristelstein, 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

Elulookirjeldus

Isiklik informatsioon

Nimi	Alex Sirp
Sünniaeg	04.05.1992
E-mail	alex.sirp@taltech.ee

Hariduskäik

2017–2023	Tallinna Tehnikaülikool, geenitehnoloogia (doktoriõpe)
2014–2016	Tallinna Tehnikaülikool, geenitehnoloogia (magistriõpe)
2011–2014	Tallinna Tehnikaülikool, geenitehnoloogia (bakalaureuseõpe)
2008–2011	Estonian Business School gümnaasium (kuldmedal)

Teenistuskäik

2017–	Tallinna Tehnikaülikool, Loodusteaduskond, Keemia ja biotehnoloogia instituut, dotkorant-nooremteadur (0,10)
2015–2016	Tallinna Tehnikaülikooli üliõpilasesinduse esinduskogu liige

Stipendiumid

2022	Erkki Truve nimeline doktoriõppe stipendium, Tallinna Tehnikaülikooli Arengufond.
------	---

Teaduskraadid

Alex Sirp, magistrikraad, 2016, (juh) Jürgen Tuvikene; Tõnis Timmus, Screening of Transcription Factors Regulating BDNF Positive Feedback Loop in Cortical Neurons (BDNF geeni positiivse tagasiside regulatsioonis osalevate transkriptsioonifaktorite tuvastamine ajukoore neuronites), Tallinna Tehnikaülikool, Matemaatika-loodusteaduskond, Geenitehnoloogia instituut, Molekulaarbioloogia õppetool.

Alex Sirp, bakalaureusekraad, 2014, (juh) Teet Velling, Epidermaalse kasvufaktori retseptori, integriini-tüüpi kollageeni retseptorite ja filamiin A roll etoposiidi ja staurosporiini poolt indutseeritud rakusurmas, Tallinna Tehnikaülikool, Matemaatika-loodusteaduskond, Geenitehnoloogia instituut, Molekulaarbioloogia õppetool.

Konverentsidel osalemine

Oktoober 2022	Osalemine Eesti Inimesegeneetikaühingu 22 aastakonverentsil, Pärnu
November 2021	Osalemine Eesti Inimesegeneetikaühingu 21 aastakonverentsil, Haapsalu
February 2020	Osalemine doktorantide talvekoolis “Writing Process Reengineering”, Rakvere, Estonia

November 2019	Osalemine Eesti Inimesegeneetikaüingu 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 Inimesegeneetikaüingu 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 Inimesegeneetikaüingu 18 aastakonverentsil, Rakvere

Juhendatud väitekirjad

- Carl Sander Kiir, magistriraad, 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 äädikakärbe 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, magistriraad, 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 peajus), 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 rakusisesse 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

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-001-0 (PDF)