THESIS ON NATURAL AND EXACT SCIENCES B145

Functions of the Basic Helix-Loop-Helix Transcription Factor TCF4 in Health and Disease

MARI SEPP

TUT PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Science
Department of Gene Technology

Dissertation was accepted for the defence of the degree of Doctor of Philosophy in Gene Technology on November 13, 2012.

Supervisor: Professor Tõnis Timmusk, PhD,
Department of Gene Technology, Faculty of Natural and Exact Sciences,
Tallinn University of Technology, Estonia

Opponents: Associate professor Moritz Rossner, PhD,
Max Planck Institute for Experimental Medicine, Göttingen, Germany

Professor Eero Vasar, MD, PhD,
Department of Physiology, Faculty of Medicine, University of Tartu, Estonia

Defence of the thesis: December 21, 2012, Tallinn University of Technology

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 any academic degree.

/Mari Sepp/

The studies described in this thesis were performed at the Department of Gene Technology of Tallinn University of Technology.

The research reported in this thesis was supported by grants from Estonian Ministry of Education and Research (Grant 0140143) and Estonian Science Foundation (Grants 7257 and 8844).

Financial support for the publication of this thesis was provided by the Graduate School in Biomedicine and Biotechnology created under auspices of European Social Fund.

ISSN 1406-4723
ISBN 978-9949-23-406-6 (publication)
ISBN 978-9949-23-407-3 (PDF)

LOODUS- JA TÄPPISTEADUSED B145

Aluselise heeliks-ling-heeliks transkriptsioonifaktori TCF4 funktsioonid ja seosed haigustega

MARI SEPP

TTÜ
KIRJASTUS

CONTENTS

ABSTRACT

Transcription factor TCF4 (alias ITF2, SEF2 or E2-2) is a broadly expressed basic helix-loop-helix protein that binds to E-box DNA sequences (CANNTG) as a homo- or heterodimer. While involved in the development and functioning of many different cell types, recent studies point to important roles for TCF4 in the nervous system. Specifically, human *TCF4* gene is implicated in susceptibility to schizophrenia and deletions, truncating and missense mutations in *TCF4* cause Pitt-Hopkins syndrome (PTHS), a rare developmental disorder characterized by severe motor and mental retardation, typical facial features and breathing anomalies.

The aims of this thesis were (i) to characterize human *TCF4* gene structure and *TCF4* expression at mRNA and protein level, (ii) to study the functional characteristics and regulation of alternative TCF4 isoforms in non-neural cells and neurons, and (iii) to determine the impact of PTHS-associated mutations on TCF4 functions.

The results of the current study demonstrate that although widely expressed, human *TCF4* expression is particularly high in the brain. In addition, usage of numerous 5′ exons potentially yields in TCF4 protein isoforms with 18 different N-termini. Functional characterization of TCF4 isoforms revealed that their subcellular distribution is differentially regulated: some isoforms contain a bipartite nuclear localization signal and are exclusively nuclear, whereas distribution of other isoforms relies on heterodimerization partners. Furthermore, the ability of different isoforms to regulate E-box-controlled transcription is varied and dependent on cell type. Compared to non-neural cells where TCF4 is constitutively active in up-regulating E-box-controlled transcription, TCF4-dependent transcription in neurons requires induction of calcium influx through L-type voltage gated channels and signaling via protein kinase A (PKA). According to the present study, TCF4 is a novel direct target of PKA and PKA phosphorylation sites in TCF4 are required for full transactivation by TCF4 in neurons. Additionally, endogenous TCF4 is posttranslationally modified in response to membrane depolarization of primary neurons. Analysis of PTHS-associated mutations in *TCF4* revealed that not all deletions and truncating mutations result in complete loss-of-function and the impact of missense and reading frame extending mutations ranges from subtle deficiencies to dominant-negative effects. Particularly, some missense and reading frame extending mutations damage DNA-binding and transactivation ability in a manner dependent on dimer context, some destabilize the protein, whereas others cause no major functional deficiencies.

Altogether, this study describes the inter-tissue variability of *TCF4* expression in human, provides evidence about the functional diversity of the alternative TCF4 isoforms, identifies a new role for TCF4 as an activity-regulated transcription factor in neurons, and demonstrates that PTHS-associated mutations impair the functions of TCF4 by diverse mechanisms ranging from hypomorhic to dominant-negative effects. These results provide novel insights into the functions of TCF4 and offer a new perspective on the mechanisms underlying association of *TCF4* with disorders affecting cognitive development.

ORIGINAL PUBLICATIONS

This thesis is based on the following publications and a manuscript that are referred to in the text by their Roman numerals:

I **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. 2011;6(7):e22138.

II **Sepp M**, Pruunsild P, Timmusk T
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. 2012;21(13):2873-2888.

III **Sepp M**, Vihma H, Kannike K, Pruunsild P, Timmusk T
Basic helix-loop-helix transcription factor TCF4 is a calcium-dependent transcriptional activator in neurons.
Manuscript.

The publications are amended with permissions of the copyright owners.

Some supporting data not included in the papers is additionally presented.

Author's contribution:

I designed all and performed the majority of the experiments for studies I-III.
I analysed the data and wrote the manuscripts.

ABBREVIATIONS

AD1, -2	transcription activation domain 1, 2
ARNT	aryl hydrocarbon receptor nuclear translocator
ASCL1	achaete-scute complex homolog 1
ATOH1	atonal homolog 1
bHLH	basic helix-loop-helix
bHLH-O	basic helix-loop-helix and Orange domain
bHLH-PAS	basic helix-loop-helix and period-ARNT-single-minded domain
bHLH-Zip	basic helix-loop-helix and leucine zipper
BMP	bone morphogenetic protein
BPAD	bipolar affective disorder
CaM	calmodulin
CAMK	calcium/calmodulin-dependent protein kinase
cAMP	cyclic adenosine monophosphate
CBP	CREB binding protein
cDC	conventional/classical dendritic cell
CE	conserved element
CEn	corneal endothelium
CNS	central nervous system
CNTNAP2	contactin associated protein-like 2
COE	COLLIER-OLF1-EBF domain
CpG	cytosine-phosphate-guanine
CREB1	cAMP responsive element binding protein 1
CUX2	cut-like homeobox 2
CXCR4	chemokine (CXC motif) receptor 4
da	daughterless
dbcAMP	dibutyryl cyclic adenosine monophosphate
DES	downstream ETO-interacting sequence
dpf	days post fertilization
EBF	early B cell factor
E-box	Ephrussi box
EF-hand	E-helix-loop-F-helix in a "hand" configuration
emc	extra macrochaetae
EMSA	electrophoretic mobility shift assay
EMT	epithelial-mesenchymal transition
E-protein	Ephrussi box binding protein
ETO	eight twenty one protein
FAM96B	family with sequence similarity 96, member B
FECD	Fuchs' endothelial corneal dystrophy
FLT3L	fms-related tyrosine kinase 3 ligand
FRET	fluorescence resonance energy transfer
GBS	GAL4 binding site
G_0/G_1	cell cycle phase gap 0 or 1
GM-CSF	granulocyte-macrophage colony-stimulating-factor

GWA	genome-wide association
HEB	HeLa E-box binding protein
HEK293	human embryonic kidney 293 cell line
HES1	hairy and enhancer of split 1
HLH	helix-loop-helix
hlh-2	helix-loop-helix protein 2
ID	inhibitor of differentiation and DNA-binding
IFN	interferon
ITF2	immunoglobulin transcription factor 2
KDR	kinase insert domain receptor
MAPK	mitogen activated protein kinase
MEK	MAPK/extracellular-signal-regulated kinase kinase
miR	microRNA
MYOD1	myogenic differentiation factor 1
NEUROD1	neurogenic differentiation factor 1
NHR1, -2	nervy homology region 1, 2
NRXN1	neurexin 1
OLF1	olfactory neuronal transcription factor 1
PCR	polymerase chain reaction
pDC	plasmacytoid dendritic cell
PKA	protein kinase A
PKC	protein kinase C
PPI	prepulse inhibition
PTCRA	pre-T cell antigen receptor α-chain
PTHS	Pitt-Hopkins syndrome
RED	repeat expansion detection based on ligase chain reaction
Rep	repression region
RUVBL1	RuvB (E coli DNA helicase homolog)-like 1
SAGA	Spt-Ada-Gcn5 acetyltransferase
siRNA	small interfering RNA
SCL	stem cell leukemia protein
SCZ	schizophrenia
SNP	single nucleotide polymorphism
STAT3	signal transducer and activator of transcription 3
TAFH	TATA binding protein-associated factor 4 homology
TAL1	T cell acute lymphocytic leukemia 1
TCF3, -4, -12	transcription factor 3, 4, 12
TCF7L2	transcription factor 7-like 2
TE	transposable element
USF	upstream stimulatory factor
VDM	verbal declarative memory
VEGFR2	vascular endothelial growth factor receptor 2
VGCC	voltage-gated calcium channel
WNT	family of wingless-related/integration-site proteins
ZEB1	zinc finger E-box binding homeobox 1

LITERATURE REVIEW

1. Basic helix-loop-helix transcription factors

Transcription factors are classified into families according to the structure of their DNA-binding domains. A large eukaryotic superfamily of dimeric transcriptional regulatory proteins contains the basic helix-loop-helix (bHLH) domain that consists of a DNA-binding basic region and two dimerization-mediating α-helices separated by a loop (Massari and Murre, 2000). bHLH proteins are found in fungi, plants and metazoans, and play fundamental roles in diverse aspects of cellular physiology. In unicellular organisms, such as the yeast, bHLH proteins control metabolic pathways whereas in multicellular organisms bHLH proteins are involved in various developmental processes, regulation of cell cycle and response to environmental signals (Massari and Murre, 2000, Jones, 2004, Skinner et al., 2010).

1.1. Classification of bHLH proteins

The putative total number of bHLH protein encoding genes is eight in *Saccharomyces cerevisiae*, 39 in *Caenorhabditis elegans*, 59 in *Drosophila melanogaster*, 78 in *Branchiostoma floridae*, 117 in *Homo sapiens* (Table 1) and ~120-170 in plants *Arabidopsis thaliana* and *Oryza sativa* (Robinson and Lopes, 2000, Buck and Atchley, 2003, Li et al., 2006, Simionato et al., 2007, Skinner et al., 2010). bHLH proteins have been grouped into orthologous families, each of which contain bHLHs evolved by vertical descent from a common ancestor (Ledent and Vervoort, 2001, Simionato et al., 2007). Since there are no families of orthologues between fungi, plants and metazoans, it has been suggested that diversification of bHLH proteins has taken place independently in these clades (Ledent and Vervoort, 2001, Simionato et al., 2007). The last common ancestor for all modern metazoan possessed at least 10-14 bHLH proteins belonging to 5-6 higher-order groups (see below). This indicates that diversification of bHLH genes appeared early in the evolution of animals, possibly in relation to the acquisition of multicellularity. The later phases of expansion to at least 29-33 and 44 bHLH proteins occurred before the cladogenesis of eumetazoans and bilaterians, respectively (Simionato et al., 2007). In modern metazoans, 45 orthologous families have been defined (Table 1).

Various methods to classify bHLH proteins into higher-order groups have been devised. The historic scheme that divides bHLH proteins into 7 classes (I-VII), is based on tissue distribution, dimerization capabilities and DNA-binding specificities of the proteins (Murre et al., 1994, Massari and Murre, 2000, McLellan et al., 2002). However, phylogenetic analyses of the bHLH domain or entire protein sequences have led to categorization systems that distribute bHLH

Table 1. Metazoan families of bHLH proteins.

	Family	Class	Group	Clade	Human genes
bHLH	E-protein	I	A	b	*TCF3/E2A*, *TCF4/ITF2*, *TCF12/HEB*
bHLH	AC-S a	II	A	a	*ASCL1*, *ASCL2*
bHLH	AC-S b	II	A	a	*ASCL3*, *ASCL4*, *ASCL5*
bHLH	Atonal	II	A	a	*ATOH1*, *ATOH7*
bHLH	Net	II	A	a	*ATOH8*
bHLH	Neurogenin	II	A	a	*NEUROG1*, *NEUROG2*, *NEUROG3*
bHLH	NeuroD	II	A	a	*NEUROD1*, *NEUROD2*, *NEUROD4*, *NEUROD6*
bHLH	Mist	II	A	a	*BHLHA15/MIST1*
bHLH	Beta3	II	A	e	*BHLHE22/BETA3*, *BHLHE23/BETA4*
bHLH	Oligo	II	A	e	*OLIG1*, *OLIG2*, *OLIG3*
bHLH	PTF a	II	A	a	*FERD3L/NTWIST*
bHLH	PTF b	II	A	a	*PTF1A*
bHLH	Twist	II	A	a	*TWIST1*, *TWIST2*
bHLH	NSCL	II	A	a	*NHLH1/NSCL1*, *NHLH2/NSCL2*
bHLH	SCL	II	A	a	*TAL1/SCL*, *TAL2*, *LYL1*
bHLH	Hand	II	A	a	*HAND1*, *HAND2*
bHLH	Paraxis	II	A	a	*TCF15/PARAXIS*, *SCXA, SCXB*
bHLH	MyoR a	II	A	a	*MSC*, *TCF21/POD1*
bHLH	MyoR b	II	A	a	*TCF23/OUT*, *TCF24*
bHLH	Mesp	II	A	c; a	*MESP1*, *MESP2*; *MSGN1*
bHLH	Delilah	II	A	c	-
bHLH	MYOD	II	A	c	*MYOD1*, *MYOG*, *MYF5*, *MYF6*
bHLH	Figalpha	II	B	c	*FIGLA*
bHLH-Zip	AP4	III	B	c	*TFAP4/AP4*
bHLH-Zip	USF	III	B	b	*USF1*, *USF2*
bHLH-Zip	MITF	III	B	e	*MITF*, *TFEB*, *TFEC*, *TFE3*
bHLH-Zip	SREBP	III	B	d	*SREBF1/SREBP1*, *SREBF2/SREBP2*
bHLH-Zip	TF4		B	d	*MLX/TCFL4*
bHLH-Zip	MLX		B	e; d	*MLXIP/MONDOA*; *MLXIPL/MONDOB*
bHLH-Zip	Myc	III	B	e	*MYC*, *MYCN*, *MYCL1*, *MYCL2*
bHLH-Zip	Max	IV	B	d	*MAX*
bHLH-Zip	Mnt	IV	B	d	*MNT*
bHLH-Zip	Mad	IV	B	c	*MXD1*, *MXD3*, *MXD4*, *MXI1*
bHLH-O HLH	Emc/ID	V	D	b	*ID1*, *ID2*, *ID3*, *ID4*
bHLH-O HLH	Hairy/E(Spl)	VI	E	b	*HES1*, *HES2*, *HES3*, *HES4*, *HES5*, *HES6*, *HES7*
bHLH-O HLH	Hey		E	b	*HEY1*, *HEY2*, *HEYL*
bHLH-O HLH	Hairy/E(Spl) / Hey		E	e	*HELT*, *BHLHE40/BHLHB2/DEC1*, *BHLHE41/BHLHB3/DEC2*
bHLH-PAS	AHR	VII	C	e	*AHR*, *AHRR*
bHLH-PAS	Sim	VII	C	e	*SIM1*, *SIM2*
bHLH-PAS	Trh	VII	C	e	*NPAS3*
bHLH-PAS	HIF	VII	C	e	*HIF1A*, *HIF3A*, *NPAS1*
bHLH-PAS	Sim/Trh/HIF	VII	C	e	*EPAS1*
bHLH-PAS	Clock	VII	C	e	*CLOCK*, *NPAS2*
bHLH-PAS	ARNT	VII	C	e	*ARNT*, *ARNT2*
bHLH-PAS	Bmal	VII	C	e	*ARNTL/BMAL1*, *ARNTL2/BMAL2*
bHLH-PAS	SRC	VII	B	e	*NCOA1/SRC1*, *NCOA2/SRC2*, *NCOA3/SRC3*
	Coe		F		*EBF1/OLF1/COE1*, *EBF2, EBF3*, *EBF4*
	Not classified			e; a; ?	*TCFL5*, *SOHLH1*, *SOHLH2, NPAS4*; *BHLHA9*; *MGA*

proteins into 6 groups (A-F) (Atchley and Fitch, 1997, Ledent and Vervoort, 2001, Ledent et al., 2002) or 6 clades ('a'-'f') (Stevens et al., 2008, Skinner et al., 2010), respectively. Distribution of the metazoan orthologous bHLH families into classes, groups and clades is shown in Table 1 and briefly summarized below.

Class I includes broadly expressed bHLH factors known as E-proteins (Ephrussi box binding proteins), which are able to dimerize with themselves, and with class II and V proteins. Class II comprises tissue-specific bHLH transcription factors that, with a few exceptions, homodimerize poorly and generally require heterodimerization with E-proteins to allow DNA-binding. Class I and II bHLHs are joined into group A. Class III and IV together form most of group B and contain bHLH-Zip proteins that have a dimerization-mediating leucine zipper directly C-terminal to the bHLH domain and form homo- and heterodimers with other bHLH-Zip proteins. In addition to metazoan bHLH-Zip proteins, all fungi and plant bHLH proteins belong to this group indicating that group B is closest to the ancestral bHLH protein. Both, group A and B proteins mostly bind to the Ephrussi box (E-box) DNA (CANNTG). Class V alias group D consists of Emc/ID (extra macrochaetae/inhibitor of differentiation and DNA-binding) proteins, which lack a basic region and act as negative regulators of group A proteins by forming heterodimers that are not able to bind DNA. Class VI overlaps with group E that encompasses bHLH-O proteins containing an Orange domain C-terminal to the bHLH domain and a characteristic proline or glycine in their basic region. The role of the Orange domain is not clear, but it may function as an extended dimerization domain. bHLH-O proteins form dimers between themselves and with group A proteins, bind to E-box or N-box DNA (CACG(C/A)G), and most of them contain a WRPW 4-amino acid motif which allows interaction with Groucho/TLE (transducin-like enhancer of split) repressor proteins. In addition to recruitment of co-repressors to target sites, bHLH-O proteins may function as transcriptional repressors by titrating group A bHLH proteins via heterodimerization, similarly to the Emc/ID proteins. Class VII alias Group C corresponds to bHLH-PAS proteins that contain a dimerization-mediating and signal-sensing PAS (Per-ARNT-Sim) domain C-terminal to the bHLH domain. The bHLH-PAS group comprises co-activator proteins and transcription factors that preferentially bind to DNA at asymmetric E-box related sequences as bHLH-PAS dimers. Group F corresponds to the COE (Col-OLF1-EBF) family of proteins, which lack the basic region, have a divergent HLH sequence, and contain a COE domain involved in dimerization and DNA-binding to a palindromic sequence not related to E-box.

In case of the clade classification scheme, clade 'a' includes most of class II families; clade 'b' encompasses E-proteins, the Emc/ID family, the bHLH-Zip family USF (Upstream stimulatory factor), and the bHLH-O families Hes (Hairy/enhancer-of-split) and Hey (Hairy/enhancer-of-split related with YRPW motif). Clade 'c' comprises many plant bHLH proteins and some metazoan class II and bHLH-Zip families; clade 'd' includes almost half of the bHLH-Zip families; and clade 'e' contains all bHLH-PAS families and some class II, bHLH-O and bHLH-Zip families. Clade 'f' encompasses plant proteins only.

1.2. Structure and function of the bHLH domain

The bHLH domain was first described in murine E-proteins E12 and E47, products of *TCF3/E2a* gene (Murre et al., 1989). The domain consists of approximately 60 amino acid residues and is located at varied positions in different bHLH proteins (reviewed in Jones, 2004). The HLH part of the domain is responsible for dimerization and appropriate juxtaposition of the basic region that mediates DNA-binding (Voronova and Baltimore, 1990). The first structural information for the bHLH domain came from the co-crystal structures of DNA bound bHLH homodimers of bHLH-Zip proteins MAX (MYC associated factor X) and USF1, class I protein E47 and class II protein MYOD1 (myogenic differentiation 1) (Ferre-D'Amare et al., 1993, Ellenberger et al., 1994, Ferre-D'Amare et al., 1994, Ma et al., 1994). These studies demonstrated that two HLH domains form a parallel left-handed four-helix bundle where dimers are stabilized by hydrophobic and electrostatic interactions between the helices. α-helical basic regions bind DNA symmetrically on opposite sides in the major groove and each monomer interacts with a half-site of the E-box sequence (see Figure 1C). The most critical residue for E-box binding is a nearly invariant glutamate in the basic region that contacts cytosine and adenine bases in the CAN half-site. Several other base contacts with the half-site or 5' flanking nucleotides are made by other residues in the basic region and the interaction with DNA is stabilized by numerous contacts to the backbone of the entire E-box and flanking nucleotides (reviewed in De Masi et al., 2011). Although the basic region is α-helical when bound to DNA, it is very mobile and largely unstructured in the absence of DNA (Fairman et al., 1997).

The regulatory and biological divergence of bHLH proteins is determined by several parameters: spatiotemporal expression pattern, selection of dimerization partners and DNA-binding specificity (Grove et al., 2009). By analysing all these parameters in the network of *Caenorhabditis elegans* bHLH proteins, it has been shown that although several bHLH proteins exhibit only limited divergence in one or more of the parameters, even minor differences in at least one of the parameters lead to little overlap in candidate target genes (Grove et al., 2009). There is specificity and promiscuity in the spatiotemporal expression network of bHLH proteins, with some factors expressed broadly and others in a restricted manner, and with some tissues expressing many bHLH proteins while others expressing a few. In case of dimerization partner selection, the majority of bHLH proteins in *C. elegans* exhibit high specificity and interact with only a single other bHLH protein. However, there are two central hubs in the dimerization network, the E-protein hlh-2 and the ARNT (aryl hydrocarbon receptor nuclear translocator) family protein aha-1 (aryl hydrocarbon receptor associated protein 1), which bind many bHLH proteins and act as obligate partners for all class II and class VII bHLH proteins in *C. elegans*, respectively (Grove et al., 2009). Analysis of the conservation of the residues at the interface of bHLH dimers has revealed that residues involved in self interactions are common among all bHLH proteins, while both common and varied amino acids are present among residues involved in partner interactions, defining the specificity of dimerization (Chavali et al., 2001). Similarly, there are critical

residues in the basic region of bHLH proteins that dictate the specificity of DNA-binding (De Masi et al., 2011). For instance, bHLH proteins containing arginine at position 13 in the basic region prefer CAC or CAT half-site (depending on a residue at position 2), whereas valine, leucine or threonine at position 13 can specify binding to CAG, CAA and CAT half-sites, correspondingly. Each bHLH monomer favours particular E-box half-sites, but the preferences can be modulated by dimerization partners (De Masi et al., 2011). Additionally, most bHLH proteins exhibit preferences at the 5' flanking nucleotide position. Overall, some bHLH dimers bind to many E-boxes, whereas other bind only one, and some E-boxes are preferred by many bHLH dimers, while others by relatively few (Grove et al., 2009). In many species (but not in *C. elegans*) non-DNA-binding dimers containing the Emc/ID family proteins add to the complexity of the network. Collectively, the regulation of bHLH proteins through divergent expression patterns, various heterodimeric interactions and differences in DNA-binding specificities constitutes a dynamic network and expands the repertoire of potential target genes controlled by a given number of bHLH transcription factors.

2. E-protein family of bHLH transcription factors

There are three E-protein genes in mammals, *TCF4/ITF2*, *TCF12/HEB* (HeLa E-box binding factor) and *TCF3/E2A* (Massari and Murre, 2000). The latter encodes two alternative protein isoforms, E12 and E47, which differ in their bHLH domains. The E-protein gene family in teleost fish consists of five members, including two additional genes coding for an TCF3/E2A-related factor and EX factor (Hikima et al., 2005). Vertebrate E-proteins are homologous to the only E-proteins in *Drosophila melanogaster* and *Caenorhabditis elegans*, da (daughterless) and hlh-2 (helix-loop-helix protein 2), respectively. Although all vertebrate E-proteins have diverged considerably from da and hlh-2, TCF3/E2A proteins are more similar to invertebrate E-proteins than TCF4 and TCF12/HEB that have deviated more rapidly from the ancestral E-protein (Shain et al., 1997).

da and hlh-2 are required for embryonic development of *D. melanogaster* and *C. elegans*, correspondingly, and their mutations lead to pleiotropic effects that include defects in neurogenesis (Caudy et al., 1988, Cline, 1989, Krause et al., 1997, Portman and Emmons, 2000). In *D. melanogaster, da* null embryos lack all peripheral neurons and have abnormalities in the central nervous system (Caudy et al., 1988). Additionally, da has been demonstrated to be involved in sex determination, myogenesis, oogenesis, intestinal stem cell maintenance and development of the trachea and salivary glands (Caudy et al., 1988, Cline, 1989, Cummings and Cronmiller, 1994, Chandrasekaran and Beckendorf, 2003, Wong et al., 2008, Bardin et al., 2010).

In mammals, substantial functional overlap among E-proteins has hampered deciphering their exact roles. Postnatal lethality is observed in every transheterozygous combination of E-protein null alleles in mice, suggesting that E-proteins share common roles during development (Zhuang et al., 1996). The importance of dosage rather than identity of available E-proteins is further

supported by the demonstration that human *TCF12/HEB* cDNA can functionally replace *Tcf3*/*E2a* in mice (Zhuang et al., 1998). Yet, in addition to overlapping roles, E-proteins have been implicated in performing unique roles that may rise from finely tuned level of expression and/or from distinct functional features of the proteins. In line with this, E-proteins can regulate the expression of common and specific genes (Cisse et al., 2008, Sobrado et al., 2009). Remarkably, differences in the DNA-binding specificities of distinct E-proteins have been shown both for E-protein homo- and heterodimers (Hu et al., 1992, Chiaramello et al., 1995, Sigvardsson, 2000, Murakami et al., 2004). Additionally, different E-proteins can have opposite effects on transcription from the same promoter (Dhulipala et al., 2001). Although E-proteins have broad expression patterns, cell context has an important role in defining lineage-specific functions of distinct E-proteins by determining whether E-proteins form homo- or heterodimers, whether they associate with different co-activator or -repressor complexes, and which E-protein target genes are accessible for regulation (reviewed in Kee, 2009, de Pooter and Kee, 2010).

2.1. Transcription regulatory domains of E-proteins

E-proteins contain two transcription activation domains (ADs), N-terminal AD1 and centrally located AD2, that have been suggested to possess distinct activities in different cell types (Aronheim et al., 1993, Quong et al., 1993, Argenton et al., 1996). AD2, which is present in all E-proteins, is highly conserved whereas AD1 of vertebrate E-proteins is divergent from the N-terminal transactivation domain of invertebrate E-proteins (Zarifi et al., 2012). Additionally, AD1 is absent from alternative protein isoforms of TCF4 and TCF12/HEB. Both vertebrate E-protein transactivation domains have been suggested to form a partial amphipathic helix and are functional in yeast (Aronheim et al., 1993, Massari et al., 1996).

In yeast, ADA and SAGA (Spt-Ada-Gcn5 acetyltransferase*)* chromatin remodelling complexes are required for AD1-mediated transactivation (Massari et al., 1999). AD1 contains a LDFS motif that directly interacts with the SAGA complex and is conserved in yeast HLH protein Rtg3p (retrograde regulation protein 3). In mammals, AD1 and AD2 interact with P300, CBP (CREB binding protein) and PCAF (CBP associated factor) (Qiu et al., 1998, Bradney et al., 2003, Bayly et al., 2004, Zhang et al., 2004). The interaction with CBP/P300 has been suggested to be cooperative between AD1 and AD2, and involves a domain in CBP/P300 known as KIX and α-helical portions of both E-protein activation domains (Bayly et al., 2004). In case of AD1 the direct interaction with CBP and P300 is mediated by its LXXLL motif (Zhang et al., 2004), a protein recognition motif present in many transcription factors and cofactors including many KIX-interacting proteins (Plevin et al., 2005). In AD1 the last residue of the LXXLL motif overlaps with the first residue of the LDFS motif (Figure 1). Both motifs and their adjacent residues are strictly conserved in vertebrate E-proteins. It has been shown that the LXXLL-containing portion of AD1 undergoes a helical transition upon interacting with the KIX domain and amino acid substitutions that prevent

helix formation in AD1 interfere with the KIX interaction (Bayly et al., 2006). The formed α-helix of AD1 binds into a shallow hydrophobic groove of the KIX domain (Figure 1).

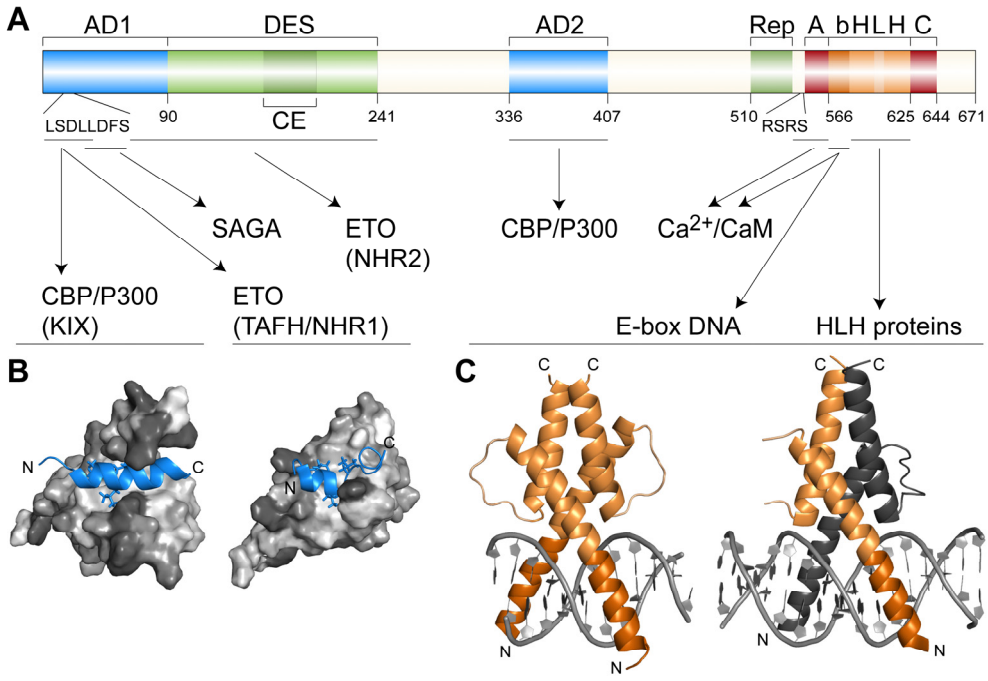


Figure 1. Conserved domains and motifs of E-proteins.
(**A**) Schematic diagram of TCF4 isoform B. Shown are the E-protein domains and motifs and their interacting molecules. Numbers denote amino acid positions. Locations of the LXXLL and LDFS motifs and RSRS insert are indicated. A, motif A; AD1, activation domain 1; AD2, activation domain 2; bHLH, basic helix-loop-helix domain; C, motif C; CE, conserved element; DES, downstream ETO-interacting sequence, Rep, repression region.
(**B**) Depiction of the solution structures of the strictly conserved peptide from the vertebrate AD1 domain (LGTDKELSDLLDFSAMFS) bound to the KIX domain of CBP (PDB: 2KWF) or TAFH/NHR1-domain of ETO (PDB: 2KNH). The AD1 peptide is shown as ribbon drawing and leucine residues of the LXXLL motif are depicted as sticks. KIX and TAFH/NHR1 domains are shown as surfaces with the hydrophobicity degree of amino acids indicated using a colour ramp from white (hydrophobic) to dark gray (charged).
(**C**) Ribbon drawings of the crystal structures of E47 bHLH homodimer and E47-NEUROD1 bHLH heterodimer (PDB: 2QL2) bound to E-box-containing DNA. N, aminoterminus; C, carboxyterminus.

The LXXLL motif in AD1 also interacts with RUNX1T1/MTG8/ETO (runt-related transcription factor 1; translocated to, 1/ myeloid translocation gene on 8q22/eight twenty one protein) co-repressor. Binding of ETO blocks the interaction of AD1 with CBP/P300 and leads to active transcription repression through formation of a histone deacetylase-containing complex (Zhang et al., 2004). E-protein-dependent transcription is repressed also by other ETO family members MTGR1 (MTG8-related protein 1) and MTGR2/ETO2 (Guo et al., 2009). The LXXLL containing peptide of AD1 binds into the hydrophobic groove of TAFH (TATA binding protein-associated factor 4 homology) alias NHR1 (nervy homology region) domain of ETO (Figure 1). The LXXLL motif is anchored to the hydrophobic groove through its 3 conserved leucines and the preceding glutamate residue that participates in an electrostatic interaction (Park et al., 2009). In addition to AD1, the DES (downstream ETO-interacting sequence) region of 150 amino acids in E-proteins binds the NHR2 and TAFH/NHR1 domains of ETO (Guo et al., 2009). This interaction allows ETO to inhibit E-protein-dependent transcription through a separate, chromatin-independent mechanism that involves interference with the assembly of the preinitiation complex. Although the ETO-binding region in DES has not been mapped precisely, a conserved element (CE) inside the DES region that represses AD1-dependent transcription in a manner dependent on three conserved lysine residues in CE, has been identified (Herbst and Kolligs, 2008). Additionally, a repression domain Rep consisting of 30 amino acids N-terminal to the bHLH domain and conserved from fly to human, has been suggested to differentially modulate the activity of AD1 and AD2 in E-protein hetero- and homodimers (Markus et al., 2002, Wong et al., 2008).

2.2. Dimerization and DNA-binding of E-proteins

It is generally believed that E-proteins bind DNA at E-box sequences mainly as homodimers or as heterodimers with other E-proteins during immune system development, whereas they form DNA-binding heterodimers with class II bHLH proteins in most other tissues (reviewed in Sun, 2004, Murre, 2005, de Pooter and Kee, 2010). In addition, E-proteins heterodimerize with ID proteins that act as dominant negative factors in regulating the DNA-binding ability of E-proteins (reviewed in Perk et al., 2005). E-protein titration by an analogous mechanism may also be carried out by bHLH-O proteins (Davis and Turner, 2001). In E-proteins, the DNA-binding basic region and dimerization-mediating HLH domain are conserved from *C. elegans* to human and are located in the C-terminal part of the protein (Figure 1). In addition to the HLH domain, motif C is required for E-protein dimerization *in vivo* (Goldfarb et al., 1998). Motif C is the most highly conserved region in E-proteins and consists of 20 amino acids directly C-terminal to the bHLH domain (Figure 1).

Preferences of both, E-protein dimerization and DNA-binding of different dimers, can be modulated by various mechanisms. For instance, dimerization specificity of E-proteins can be regulated by motif A, an acidic amino acid sequence directly N-terminal to the bHLH domain (Figure 2), that in case of E12

has been demonstrated to prevent homodimerization (Shirakata and Paterson, 1995). Additionally, alternative splicing generated 24-amino-acid insertion in the AD2 domain interferes with homo- and heterodimerization of TCF12/HEB (Klein et al., 1993). Moreover, phosphorylation of serine residues located in the DES region, Rep domain or motif A have been shown to modulate dimerization specificity of E47 by inhibiting the formation of homodimers and/or supporting heterodimerization with MYOD1 (Sloan et al., 1996, Lluis et al., 2005).

DNA-binding preferences of E-protein dimers are regulated by the Ca^{2+} binding protein calmodulin (CaM). One E-protein homodimer is able to directly bind two Ca^{2+}-loaded calmodulin molecules via a unique calmodulin binding mode that is highly dynamic and involves hydrophobic as well as electrostatic interactions (Onions et al., 2000, Larsson et al., 2001, Larsson et al., 2005). The C- and N-terminal globular domains of calmodulin, which both contain two Ca^{2+}-binding EF-hand motifs, preferentially bind to the basic sequence within the bHLH domain or to the modulatory sequence between the Rep region and bHLH domain, respectively (Onions et al., 2000). Binding of Ca^{2+}-loaded CaM to the basic sequence inhibits DNA binding of E-protein homodimers *in vitro* and *in vivo* (Corneliussen et al., 1994, Onions et al., 1997, Saarikettu et al., 2004). Heterodimers of E-proteins with ASCL1 (achaete-scute complex homolog 1, alias MASH1) or MYOD1 are less sensitive or resistant to Ca^{2+}/CaM-mediated inhibition due to the differences in the modulatory sequences N-terminal to the basic sequence that obstruct the interaction (Corneliussen et al., 1994, Onions et al., 1997). In addition to CaM, S100 EF-hand proteins interact with and inhibit DNA binding of bHLH proteins with the interaction being differentially modulated by the basic sequence and the sequence N-terminal to the bHLH domain (Onions et al., 1997). Notably, overexpression of calmodulin or increasing intracellular Ca^{2+} concentration inhibits transcriptional activation by E-protein homodimers in cells, whereas transactivation by MYOD1 heterodimers is enhanced by Ca^{2+} signalling (Corneliussen et al., 1994, Saarikettu et al., 2004, Hauser et al., 2008a). Moreover, experiments with CaM-resistant TCF3/E2A mutants have demonstrated that B-cell-receptor activation inhibits the ability of TCF3/E2A homodimers to regulate gene expression in B cells through Ca^{2+}/CaM-mediated inhibition of DNA-binding (Hauser et al., 2008b, Hauser et al., 2009, Hauser et al., 2010, Verma-Gaur et al., 2012). Additionally, preventing DNA binding of TCF3/E2A homodimers and thereby selectively promoting DNA binding of MYOD and E-protein heterodimers by Ca^{2+}/CaM signalling has been shown to be required for the progression of myogenesis (Hauser et al., 2008a). Thus, DNA binding of alternative E-protein dimers is modulated by Ca^{2+} signalling.

3. Transcription factor TCF4

TCF4 is the official designation of a transcription factor that has historically been named also ITF2, E2-2, SEF2, TFE, ME2, GE2, XE2 and bHLHb19. First, ITF2 (immunoglobulin transcription factor 2) alias E2-2 (clone name) cDNA was isolated from human B-cell derived library based on the ability of its product to

bind to μE5 element (CACCTG) of the immunoglobulin heavy chain enhancer (Henthorn et al., 1990a, Henthorn et al., 1990b). Second, SEF2 (leukaemia virus SL3-3 enhancer factor 2) was isolated by screening of human thymocyte library with glucocorticoid response element DNA (contains CAGATG) from the SL3-3 retrovirus enhancer (Corneliussen et al., 1991). Third, search for factors binding to an element (containing CAGGTG) in dog thyroglobulin gene promoter resulted in isolation of TFE (thyroid factor E) from the dog thyroid library (Javaux et al., 1991). Fourth, ME2 and GE2 (*Mus musculus* or *Gallus domesticus* E-box binding protein 2) coding sequences were amplified from developing nervous system cDNA libraries of the mouse and chick, respectively, by PCR with degenerative primers based on the bHLH domain sequence (Neuman et al., 1993). Fifth, cDNAs of related *Xenopus laevis* transcription factors termed XE2.1 and XE2.2 were isolated from frog cDNA libraries using low stringency hybridization with a mouse E-protein cDNA probe (Shain and Zuber, 1995, Shain et al., 1997). Last, an alternative name bHLHb19 (clade 'b' basic helix-loop-helix protein 19) was adopted for TCF4 in a recently introduced unified bHLH nomenclature based on phylogenetic analysis (Skinner et al., 2010).

Confusingly, in addition to transcription factor 4 the acronym TCF4 stands for T cell factor 4, a widely used alternative name for a high mobility group box transcription factor officially named TCF7L2 (transcription factor 7-like 2). TCF7L2 is one of the TCF/LEF (T cell factor/lymphoid enhancer-binding factor) factors that interact with β-catenin and participate in the WNT signalling pathway. Mix-ups of the two transcription factors, TCF4/ITF2 and TCF7L2/TCF4 have already led to data misinterpretations in several published papers (Dunn et al., 2006, Nagasawa et al., 2008, Phillips and Goodman, 2008, Pise-Masison et al., 2009, Svec et al., 2009, Zhang et al., 2009, Loe-Mie et al., 2010, Svec et al., 2010, Lu et al., 2011, Bian et al., 2012, Chiou et al., 2012, Igo et al., 2012, Lu et al., 2012, Wirgenes et al., 2012, Navarrete et al., in press).

3.1. *TCF4* expression

The gene coding for TCF4 is located on the long arm of chromosome 18 in human and mouse. Multiple mRNA species are known to be transcribed from the human and mouse *TCF4* genes (Corneliussen et al., 1991, Skerjanc et al., 1996, Liu et al., 1998, Liu et al., 2004). Accordingly, three amino-terminally distinct TCF4 protein isoforms, TCF4-A, TCF4-B and TCF4-D, have been described (Corneliussen et al., 1991, Skerjanc et al., 1996). Additionally, usage of alternative splice sites has been shown to give rise to isoforms that differ in the presence or absence of four amino acids RSRS (Corneliussen et al., 1991, Skerjanc et al., 1996). Therefore, the longest isoform TCF4-B has 667 or 671 amino acids. Compared to TCF4-B, TCF4-D and -A lack the first 130 or 183 amino acids, respectively, and TCF4-A has 23 unique amino acids in its N-terminus. All the isoforms contain the bHLH domain and transcription activation domain AD2, whereas only TCF4-B comprises the AD1 domain.

Temporal and spatial expression pattern of *TCF4* has been shown to be broad in fish, amphibian and mammalian species. During embryonic development of *Danio rerio*, *tcf4* expression has been revealed in the distal notochord at the 12-somite stage, in somites at the 20-somite stage, in dorsal telencephalon 1 day post-fertilization (dpf) and in the retina, branchial arches, telencephalon, rhombencephalon and multiple structures of diencephalon and mesencephalon at 2-3 dpf (Brockschmidt et al., 2007, Brockschmidt et al., 2011). In *X. laevis* embryogenesis, transcripts of *XE2.1* and *XE2.2* have been detected in the eye vesicle, branchial arches, olfactory region and hindbrain at the tailbud stage (Shain et al., 1997).

In mammals, *TCF4* is highly expressed during embryonic development (Soosaar et al., 1994, Pscherer et al., 1996). Widespread expression has been shown at embryonic days E8.5-E13.5 (Carnegie stages c9-c18) in *M. musculus* (Dorflinger et al., 1999, Murakami et al., 2004, Sobrado et al., 2009) and at c13-c18 (28-48 dpf) in *H. sapiens* (de Pontual et al., 2009). In mouse, the highest *Tcf4* transcript levels at E12 (c15) have been detected in the telencephalon, rhombencephalon, neural tube, branchial arches and the developing vertebral column and limbs (Soosaar et al., 1994). In E18 mouse central nervous system (CNS), *Tcf4* expression is prominent in the olfactory bulb, cerebral cortex, hippocampus, ventricular zone and cerebellum (Soosaar et al., 1994). In human, *TCF4* expression domains at c13 include prosencephalon, branchial arches and somites, by c15 expression is detected in the limb buds and gonadal ridge, and by c18 in the ventricular zone of the CNS, vertebrae, sympathetic, parasympathetic and enteric ganglia, mesonephros and the developing thyroid, thymus, pituitary gland, lung, kidney and other sites (de Pontual et al., 2009). *TCF4* expression has been found to be absent in human and mouse embryonic heart by some (Murakami et al., 2004, de Pontual et al., 2009, Sobrado et al., 2009), but not by other studies (Pscherer et al., 1996, Dorflinger et al., 1999).

In adult tissues, *TCF4* has been shown to be expressed in human brain, heart, lung, skeletal muscle, gut and placenta (Pscherer et al., 1996, Liu et al., 1998, Liu et al., 2004, Meinhardt et al., 2005, de Pontual et al., 2009); in rat brain, heart, lung, kidney, liver, skeletal muscle, spleen, testis, prostate and ovary, (Muir et al., 2006); and in mouse brain, heart, lung, liver, skeletal muscle, mammary gland and placenta (Pscherer et al., 1996, Skerjanc et al., 1996, Scott et al., 2000, Itahana et al., 2008). Differentially from other E-protein genes, expression of which decreases substantially in the adult compared to embryonic mouse nervous system (Neuman et al., 1993, Uittenbogaard and Chiaramello, 1999), *Tcf4* expression persists in adult and aging brain (Soosaar et al., 1994, Uittenbogaard and Chiaramello, 2000, Ravanpay and Olson, 2008, Brzozka et al., 2010). In particular, high *Tcf4* expression is sustained in the areas of neuronal plasticity including olfactory bulb, cerebral cortex, hippocampus and cerebellum (Soosaar et al., 1994, Dorflinger et al., 1999, Brzozka et al., 2010).

TCF4 gene expression has been found to be regulated during the development of various cell lineages suggesting that it influences the balance between cellular proliferation and differentiation. However, the direction of regulation appears to

depend on cell type. Specifically, upregulation of *TCF4* expression has been seen upon *in vitro* differentiation of human neuroblastoma cells (Persson et al., 2000), mouse myoblasts (Chen and Lim, 1997) and osteoblasts (Beck et al., 2001). In addition, *Tcf4* transcript levels in rat Sertoli cells increase during pubertal maturation and are maintained high in adults (Muir et al., 2006). In contrast, downregulation of *Tcf4* expression has been observed in differentiated trophoblasts in mouse placenta (Scott et al., 2000) and a decline in TCF4 mRNA and/or protein levels has been described upon maturation of mouse B and T cells (Bain et al., 1993, Wikstrom et al., 2006, Wikstrom et al., 2008) and lineage commitment of hematopoietic stem cells (Lemieux et al., 2011). In several cell types including mouse mammary epithelial and rat pheochromocytoma cell lines, TCF4 levels remain constant during differentiation (Einarson and Chao, 1995, Parrinello et al., 2001).

Mechanisms underlying the regulation of cell lineage- and differentiation stage-dependent *TCF4* expression have not been thoroughly elucidated, but it is known that *TCF4* expression is modulated by multiple signalling pathways and cell environmental conditions. For instance, *TCF4* mRNA levels are upregulated by transforming growth factor-β in human lung cancer cells (Vicent et al., 2008), downregulated transiently in response to follicle-stimulating hormone or cAMP in rat cultured Sertoli cells (Muir et al., 2006) and reduced 24 hours after global ischemia in CA1 subfield of hippocampus in rat (Wigle et al., 1999). *Tcf4* transcript levels are increased in response to agouti signal protein and decreased in response to melanocyte stimulating hormone in mouse melanocytes. Accordingly, *Tcf4* expression is risen during a switch to pheomelanogenesis (production of pheomelanin instead of eumelanin) in mouse hair follicles and cultured melanocytes (Furumura et al., 2001). Additionally, *Tcf4* levels are upregulated in hematopoietic common dendritic cell progenitors by FLT3L (fms-related tyrosine kinase 3 ligand) mediated signalling and downregulated by GM-CSF (granulocyte-macrophage colony-stimulating-factor) mediated signalling (Li et al., in press). Moreover, *TCF4-B* transcript levels have been shown to be increased in response to CXCR4 (chemokine CXC motif receptor 4) and CXCL12 (chemokine CXC motif ligand 12) mediated signalling in human breast cancer cells (Appaiah et al. 2010) as well as WNT/β-catenin pathway in rat immortalized kidney cells and human colon cancer cells (Kolligs et al., 2002, Mologni et al., 2010). The latter, together with differential expression of *TCF4* alternative transcripts in colorectal cancer cell lines and colorectal carcinomas (Herbst et al. 2009A, Herbst et al 2009B), indicates that the transcription of alternative *TCF4* mRNAs can be independently regulated.

Out of the signalling molecules mentioned above, the mode of transcriptional regulation of *TCF4* has been described in more detail only for FLT3L and WNT/β-catenin pathways. Namely, STAT3 (signal transducer and activator of transcription 3) has been demonstrated to activate *Tcf4-B* promoter in reporter assays and to bind a *cis*-regulatory element in mouse *Tcf4-B* promoter *in vitro* and *in vivo* in response to FLT3L-signalling (Li et al., in press). Additionally, in rat and human *TCF4-B* promoters, proximal T-cell factor-binding elements have been identified that play a

key role in β-catenin's ability to activate transcription from *TCF4-B* promoter in reporter assays (Kolligs et al., 2002). Using chromatin immunoprecipitation it has been shown that transcription factor TCF7L2 binds to the region containing this element at the endogenous human *TCF4-B* promoter, and recruits co-factors such as β-catenin interacting ATPase/DNA helicase RUVBL1/TIP49 (RuvB-like 1/49 kDa TATA box-binding protein-interacting protein), histone acetyltransferase KAT5/TIP60 (K(lysine) acetyltransferase 5/60 kDa Tat-interacting protein) and TRRAP (transformation/transcription domain-associated protein) (Feng et al., 2003).

Regulation of *TCF4* expression at the chromatin-level has been investigated by a few studies. Firstly, it has been shown that expression of a dominant negative form of RUVBL1 or silencing of *RUVBL1* inhibits histone acetylation at *TCF4-B* proximal promoter region and interferes with its expression (Feng et al., 2003). Secondly, RNAi-mediated decrease in the levels of REST/NRSF (RE1-silencing transcription factor/neuron restrictive silencer factor), a transcription factor that maintains repressed chromatin state, increases *Tcf4* expression in a neuroblastoma cell line (Loe-Mie et al., 2010). Thirdly, *TCF4-B* expression in primary gastric tumours and gastric cancer cell lines correlates with histone acetylation and/or unmethylated CpG levels at *TCF4-B* 5′ exon, and can be upregulated by inhibiting histone deacetylation or DNA methylation (Kim et al., 2008, Herbst et al., 2009a). *TCF4* regulation has been demonstrated to occur also at the level of pre-mRNA processing, timing and degree of which is selectively modulated by CBP in hematopoietic stem cells (Lemieux et al., 2011).

MicroRNA-mediated post-transcriptional regulation adds another layer to the control of *TCF4* mRNA levels and/or translation. *TCF4* is an experimentally validated target of brain-enriched *miR-137* that is implicated in neuron maturation, dendritic morphogenesis, adult neurogenesis and schizophrenia (Smrt et al., 2010, Szulwach et al., 2010, Ripke et al., 2011, Kwon et al., in press), *miR-155* that is linked to the regulation of epithelial-mesenchymal transition (Xiang et al., 2011), and *miR-204*, which is involved in apoptosis and stress response (Li et al., 2011a). Additionally, *TCF4* is predicted to be a target of *miR-221* and *miR-222* that have been shown to participate in cell fate decision of bone marrow-derived dendritic cells (Kuipers et al., 2010) and *miR-200* family microRNAs that have been implicated in the regulation of epithelial-mesenchymal transition (Castilla et al., 2010).

3.2. TCF4 functions

Homozygous *Tcf4* knockout mice die shortly after birth for unknown causes, whereas heterozygous knockout mice are viable (Zhuang et al., 1996, Flora et al., 2007). By analyzing $Tcf4^{+/-}$ $Tcf4^{-/-}$ mosaic mice it has been found that the presence of more than 30% of *Tcf4* null cells is not compatible with life (Bergqvist et al., 2000). In addition, reduction of *tcf4* expression in zebrafish leads to a general developmental delay in embryogenesis (Brockschmidt et al., 2011). Moreover, heterozygous mutations in *TCF4* in human cause Pitt-Hopkins syndrome (PTHS), a

rare disorder characterized by intellectual disability and developmental delay (Amiel et al., 2007, Brockschmidt et al., 2007, Zweier et al., 2007).

The requirement of TCF4 for postnatal survival, its involvement in developmental processes and the presence of TCF4 in a variety of tissues suggest that it plays important roles in many cell types. Indeed, TCF4 has been implicated in the development and functioning of immune system cells (Zhuang et al., 1996, Bergqvist et al., 2000, Cisse et al., 2008), neurons (Persson et al., 2000, Flora et al., 2007), melanocytes (Furumura et al., 2001), skeletal muscle cells (Apone and Hauschka, 1995, Skerjanc et al., 1996, Chen and Lim, 1997, Petropoulos and Skerjanc, 2000), smooth muscle cells (Dhulipala et al., 2001, Kumar et al., 2003), Sertoli cells (Muir et al., 2006), endothelial cells (Tanaka et al., 2009, Tanaka et al., 2010), mammary gland (Parrinello et al., 2001, Itahana et al., 2008) and placenta (Scott et al., 2000, Liu et al., 2004). However, the precise roles of TCF4 in many of the above mentioned cell lineages have remained poorly characterized. Additionally, the genes regulated by TCF4 in different tissues and cell types are not well defined and the only lineage for which TCF4 targets have been systematically investigated is the plasmacytoid dendritic cell (Cisse et al., 2008, Ghosh et al., 2010). Therefore, the accumulated data about TCF4-regulated genes, as summarized in Table 2, is fragmented and should not be considered as a comprehensive set. In addition, given that TCF4 forms heterodimers with many other bHLH factors and interacts with different proteins involved in transcriptional regulation (Table 3) in a cell-type-dependent manner, it is estimated that TCF4-comprising complexes collectively harbour a potential of regulating a large number of target genes involving both overlapping as well as unique targets in different tissues. The functions of TCF4 in the immune and nervous system as well as its roles in the regulation of cell cycle and migration are further described below.

Table 2. TCF4 target-genes. Listed genes meet at least one of the following criteria: (i) the promoter or enhancer of a target gene is regulated by TCF4 in reporter assays, (ii) ectopic expression or silencing/deleting of *TCF4* regulates the transcription of a target gene and TCF4 binds to the target gene as determined by chromatin immunoprecipitation. TCF4-mediated up- or down-regulation is shown with arrows. If necessary, TCF4 isoform and additional partner proteins are specified after the arrow. For genes in upper case, there is at least some evidence for the regulation of the human gene by TCF4.

[a] The promoter or enhancer linked to reporter gene is regulated by TCF4;

[b] endogenous gene is regulated by ectopic expression or silencing/deleting TCF4;

[c] TCF4 binding to the locus determined by chromatin immunoprecipitation;

[d] E-box is implicated in the regulation.

Gene	Mode	a	b	c	d	References
Transcription regulator activity						
ARID5A	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
ATF3	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
Bcl11a	↑		+	+		(Ghosh et al., 2010)
CBL	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
DPF3	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)

Table 2. Continued

Gene	Mode	a	b	c	d	References
FOSB	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
ID2	↓, ↑B	+	+	+	+	(Neuman et al., 1995, Appaiah et al., 2010, Brzozka et al., 2010, Ghosh et al., 2010)
ID4	↑	+			+	(Pagliuca et al., 1998)
IRF4	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
IRF7	↑		+	+	+	(Cisse et al., 2008, Ghosh et al., 2010)
IRF8	↑		+	+	+	(Cisse et al., 2008)
JUNB	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
KDM4C/ JMJD2C	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
MITF	↓B	+	+			(Furumura et al., 2001)
MYC	↑		+	+		(Ghosh et al., 2010, Mologni et al., 2010)
NFAT5	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
NR2C1	↑		+	+		(Ghosh et al., 2010)
RBMS1	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
RORA	↑		+	+		(Cisse et al., 2008, Sobrado et al., 2009, Ghosh et al., 2010)
SPIB	↑		+	+	+	(Cisse et al., 2008, Ghosh et al., 2010)
TGFB1	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
Kinase and kinase regulator activity						
ADPGK	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
AKT1	↓		+	+		(Ghosh et al., 2010, Mologni et al., 2010)
CCND3	↑		+	+		(Ghosh et al., 2010)
CDK20/ CCRK	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
CDKN1A P21/CIP1	↓, ↑A, B	+	+	+	+	(Pagliuca et al., 2000, Cisse et al., 2008, Herbst et al., 2009a, Herbst et al., 2009b, Ghosh et al., 2010, Deliri et al., 2011)
CDKN2A P16/INK4A	↑	+			+	(Pagliuca et al., 2000)
CDKN2B P15/INK4B	↑	+		+	+	(Pagliuca et al., 2000, Ghosh et al., 2010)
Ckm	↓B of MYOD1	+			+	(Apone and Hauschka, 1995, Chen and Lim, 1997)
GMFB	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
MAP3K5	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
MAP3K11	↑		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
NUAK2	↓		+	+		(Ghosh et al., 2010)
PHLDA1	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
SGL	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
SPHK1	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
Receptor, channel and transmembrane proteins						
CCR9	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
CD14	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
CD22	↓		+	+		(Ghosh et al., 2010)
CD27	↓		+	+		(Ghosh et al., 2010, Mologni et al., 2010)
CLEC4C	↑		+	+	+	(Cisse et al., 2008, Ghosh et al., 2010)
Dll1	↑B +ASCL1	+			+	(de Pontual et al., 2009)
FCRLA	↑		+	+		(Ghosh et al., 2010)

Table 2. Continued

Gene	Mode	a	b	c	d	References
FGFR4	↑		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
IGF2R	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
IGLL1/λ5	↑ +EBF1	+		+		(Sigvardsson, 2000, Ghosh et al., 2010)
IL2RA	↓B	+		+		(Lu et al., 2005)
ITGA4	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
ITGAX	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
KDR/VEGFR2	↑B	+	+	+		(Tanaka et al., 2009, Tanaka et al., 2010)
KISS1R	↑		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
LILRA4	↑		+	+	+	(Cisse et al., 2008, Ghosh et al., 2010)
P2RX1/P2X1	↑B, →A	+		+	+	(Dhulipala et al., 2001, Ghosh et al., 2010)
PLXNC1	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
PMEPA1/ TMEPAI	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
PTCRA	↑		+	+	+	(Cisse et al., 2008, Wikstrom et al., 2008, Ghosh et al., 2010)
SLC2A3	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
SLC2A9	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
SLC4A1	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
SLC12A1	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
SPINT2	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
SSTR2	↑B	+		+	+	(Pscherer et al., 1996, Ghosh et al., 2010)
TLR7	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
TLR9	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
TMEM176A	↓		+	+		(Ghosh et al., 2010)
Signal, adapter and transport proteins						
ACBD4	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
ANXA2	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
BTRC	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
CARD8	↓		+	+		(Ghosh et al., 2010, Mologni et al., 2010)
CDC6	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
CGA	↓B	+				(Liu et al., 2004)
FGF1	↑, ↓B^{-}	+	+		+	(Liu et al., 1998, Cisse et al., 2008)
GAB2	↑		+	+		(Ghosh et al., 2010)
GNAL	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
IGFBP5	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
NFKBIA	↓		+	+		(Ghosh et al., 2010)
PARD6B	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
PYCARD	↓		+	+		(Ghosh et al., 2010, Mologni et al., 2010)
RGS2	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
SH3BGRL	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
SLA2	↑		+	+		(Ghosh et al., 2010)
Tf	↑B +cAMP	+				(Muir et al., 2006)
TNPO1	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
YWHAB	↑		+	+		(Ghosh et al., 2010)
Catalytic activity						
ADAM23	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
ALDH2	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
ASNS	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
CASP10	↓		+	+		(Ghosh et al., 2010, Mologni et al., 2010)
DNAJB4	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
DNASE1L3	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
DCT	↓A, B	+	+			(Furumura et al., 2001)

Table 2. Continued

Gene	Mode	a	b	c	d	References
DHFR	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
DNTT	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
EGLN3	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
FBP1	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
GLT25D1	↓		+	+		(Ghosh et al., 2010)
ISG20	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
IVD	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
LDHB	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
LIAS	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
MTHFR	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
MTMR4	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
OGT	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
PJA2	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
SQRDL	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
Th	↑ +CUX2	+			+	(Yoon and Chikaraishi, 1994)
TRUB2	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
Tyr	↓A, B	+	+		+	(Furumura et al., 2001)
Tyrp1	↓A, B	+	+		+	(Furumura et al., 2001)
UHRF2	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
Structural proteins						
Acta2	↑	+				(Kumar et al., 2003)
ACTC1	↑A, ↓B +MYOD1	+				(Skerjanc et al., 1996, Petropoulos and Skerjanc, 2000)
CAV2	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
EMILIN1	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
FN1	↑A, B		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
MARCKS	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
SNTB2	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
RNA binding						
MRPL11	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
RPS12	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
SNRPF	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
SRP14	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
ZFP36	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
ZFP36L2	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
Various						
ANKIB1	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
BBS7	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
BBX	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
CCDC109B	↓		+	+		(Ghosh et al., 2010)
HMHA1	↓		+	+		(Ghosh et al., 2010)
LETMD1	↑		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
MID1IP1	↓		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
MINA	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
PSMG2/ TNFSF5IP1	↓		+	+		(Cisse et al., 2008, Ghosh et al., 2010)
USMG5	↑		+	+		(Sobrado et al., 2009, Ghosh et al., 2010)
Viral						
EBV BZLF1	↓	+			+	(Thomas et al., 2003)
HIV1 5'LTR	↓B	+				(Lu et al., 2005)

Table 3. TCF4 interacting proteins.
[a] interaction demonstrated *in vitro*,
[b] interaction demonstrated in yeast cells with over-expressed proteins,
[c] interaction demonstrated in mammalian cells with over-expressed proteins,
[d] interaction demonstrated in mammalian cells between endogenous TCF4 and over-expressed (1) or endogenous (2) partner protein.

Protein	Binding region	a	b	c	d	References
E-proteins (class I, group A, clade b)						
TCF4/ITF2	bHLH	+	+	+		(Goldfarb and Lewandowska, 1995, Persson et al., 2000, Murakami et al., 2004, Rual et al., 2005, Tanaka et al., 2009)
TCF12/HEB		+				(Hu et al., 1992, Murakami et al., 2004)
TCF3/E2A:E47		+				(Murakami et al., 2004)
Class II, group A, clade a bHLH proteins						
ASCL1/MASH1	bHLH	+	+	+		(Persson et al., 2000, Jogi et al., 2002)
ASCL3/SGN1	bHLH		+			(Yoshida et al., 2001)
ATOH1/MATH1		+	+	+		(Flora et al., 2007)
HAND1	HLH			+		(Scott et al., 2000)
HAND2		+				(Jogi et al., 2002, Murakami et al., 2004)
LYL1			+			(Langlands et al., 1997)
MSC/ABF1	bHLH		+			(Massari et al., 1998, Rual et al., 2005)
NEUROD1/BETA2				+	2	(Suzuki et al., 2001, Brzozka et al., 2010)
NEUROD2/NDRF					2	(Brzozka et al., 2010)
SCX/SCLERAXIS			+			(Muir et al., 2006)
TAL1/SCL	HLH	+	+	+	1	(Hsu et al., 1994, Goldfarb and Lewandowska, 1995, Goldfarb et al., 1996, Langlands et al., 1997, Goldfarb et al., 1998, Schuh et al., 2005, Goardon et al., 2006, Tanaka et al., 2009)
TAL2			+			(Langlands et al., 1997)
TCF21/POD1	bHLH	+	+	+		(Miyagishi et al., 2000)
TWIST1			+	+		(Singh and Gramolini, 2009)
Class II, group A, clade c bHLH proteins						
MYOD1/MYF3	bHLH	+	+	+		(Braun and Arnold, 1991, Goldfarb and Lewandowska, 1995, Goldfarb et al., 1996, Goldfarb et al., 1998)
MYOG/MYF4		+				(Braun and Arnold, 1991, Hu et al., 1992)
MYF5		+				(Braun and Arnold, 1991)
MYF6		+				(Braun and Arnold, 1991)
ID proteins (class V, group D, clade b)						
ID1	HLH	+	+	+	2	(Einarson and Chao, 1995, Goldfarb and Lewandowska, 1995, Langlands et al., 1997, Parrinello et al., 2001, Suzuki et al., 2001, Jogi et al., 2002, Liu et al., 2004, Muir et al., 2006, Itahana et al., 2008, Tanaka et al., 2010)
ID2	bHLH	+	+	+		(Goldfarb et al., 1996, Langlands et al., 1997, Massari et al., 1998, Suzuki et al., 2001, Jogi et al., 2002, Liu et al., 2004, Rual et al., 2005, Muir et al., 2006, Itahana et al., 2008)
ID3		+	+	+		(Loveys et al., 1996, Chen and Lim, 1997, Langlands et al., 1997, Suzuki et al., 2001, Jogi et al., 2002, Rual et al., 2005, Muir et al., 2006)
ID4				+		(Jogi et al., 2002)

Table 3. Continued.

Protein	Binding region	a	b	c	d	References
Transcription factors and co-regulators						
CREB1			+			(Muir et al., 2008)
ATF4/CREB2			+			(Muir et al., 2008)
GTF2B/TFIIB	AD1-2	+				(Pscherer et al., 1996)
HIVEP2/MIBP1	bHLH	+				(Dorflinger et al., 1999)
RUNX1T1/ETO	AD1	+	+			(Zhang et al., 2004, Rual et al., 2005)
EF-hand calcium-binding proteins						
calmodulin	basic	+	+			(Corneliussen et al., 1994, Onions et al., 1997, Onions et al., 2000, Larsson et al., 2005, Rual et al., 2005)
S100α1	basic	+				(Onions et al., 1997)
S100β	basic	+				(Onions et al., 1997)
E3 protein ligases						
PIAS1	523-665		+	+		(Kawai-Kowase et al., 2005)
PIAS2	523-665		+			(Kawai-Kowase et al., 2005)
TRIM33/TIF1G					1	(Bai et al., 2010)
Other						
FAM96B				+		(Yang et al., 2011)

3.2.1 Involvement of TCF4 in the regulation of cell growth and migration

TCF4 has been linked to the control of cell cycle progression in various experimental systems. On the one hand, TCF4 has been shown to inhibit proliferation of several human cancer cell lines and endothelial cells (Pagliuca et al., 2000, Kim et al., 2008, Herbst et al., 2009a, Herbst et al., 2009b, Tanaka et al., 2010). Both, promotion of cell death associated with apoptosis and induction of cell cycle arrest in G_0/G_1 phase by TCF4 have been described (Pagliuca et al., 2000, Herbst et al., 2009a, Herbst et al., 2009b). The mechanisms by which TCF4 interferes with cell cycle progression may be partially explained by its ability to regulate the expression of cyclin-dependent kinase inhibitors *CDKN1A/P21*^{cip1}, *CDKN2B/P15*^{INK4B} and *CDKN2A/P16*^{INK4A} (Table 2). On the other hand, it has been demonstrated that overexpression of TCF4-B promotes neoplastic transformation of an immortalized epithelial cell line (Kolligs et al., 2002). Additionally, TCF4 supports the expansion of B-cell progenitors (Wikstrom et al., 2006) and proliferation of a colon cancer cell line (Mologni et al., 2010). Collectively, the paradoxical effects of TCF4 on cell cycle progression may relate to the differences in the target genes activated/repressed in different cell types.

TCF4 has been identified as a regulator of epithelial-mesenchymal transition (EMT), a complex process whereby epithelial cells lose their intercellular contacts and baso-apical polarity while acquiring a mesenchymal phenotype with a motile capacity (Sobrado et al., 2009). Both TCF4-A and -B induce EMT, but are dispensable for the maintenance of the EMT induced by themselves or other EMT mediators such as SNAI1 (snail homolog 1) or E47. TCF4-A and -B indirectly downregulate the expression of cell adhesion protein E-cadherin (CDH1), induce

mesenchymal markers N-cadherin (CDH2) and fibronectin (FN), and increase expression of several EMT-related genes including *SPARC* (secreted protein, acidic, cysteine-rich) and *ZEB1* (zinc finger E-box binding homeobox 1) (Sobrado et al., 2009, Xiang et al., 2011 and Table 2). In addition, compared to control epithelial cells, clones overexpressing TCF4 display higher migratory ability (Sobrado et al., 2009).

Conversely, migration of gastric cancer cells and endothelial cells has been shown to be increased by silencing of *TCF4* (Kim et al., 2008, Tanaka et al., 2009). Moreover, TCF4 overexpression inhibits endothelial cell migration, network formation and proliferation *in vitro* and angiogenesis *in vivo*, indicating that TCF4 is a negative regulator of endothelial cell activation. TCF4-B has been demonstrated to directly repress *KDR/VEGFR2* (kinase insert domain receptor, vascular endothelial growth factor receptor 2) promoter, possibly by binding to an imperfect E-box (Tanaka et al., 2010). Activation of VEGFR2 by VEGF (vascular endothelial growth factor) plays a central role in angiogenesis. At various stages of vascular development *KDR/VEGFR2* expression is induced by bone morphogenetic proteins (BMPs) at least partly through upregulation of ID1 levels (reviewed in Watabe, 2011). It has been demonstrated that ID1 and also TAL1/SCL counteract TCF4-mediated repression of VEGFR2 promoter by disrupting DNA-binding of TCF4 (Tanaka et al., 2009, Tanaka et al., 2010). Additionally, another BMP-inducible protein of largely unknown function, FAM96B (family with sequence similarity 96, member B), interacts with TCF4, decreases the protein levels of TCF4, relieves *KDR/VEGFR2* promoter repression and potentiates endothelial cell activation (Yang et al., 2011). In sum, these studies reveal a role for TCF4 in the regulation of angiogenetic switch as a maintainer of endothelial cell quiescence, whereas ID1, TAL1/SCL1 and FAM96B promote endothelial cell activation by counteracting TCF4.

3.2.2 Role of TCF4 in the immune system

TCF4 has been implicated in the development of both adaptive and innate immune system. By analyzing $Tcf4^{+/-}$ $Tcf4^{-/-}$ mosaic mice it has been found that *Tcf4* null cells have a developmental disadvantage in populating the spleen and thymus, but not many other tissues such as the lung, kidney, skin, liver, cerebral cortex or midbrain (Bergqvist et al., 2000). In transfusion experiments with irradiated wild-type host mice, *Tcf4* null fetal liver cells mediate efficient hematopoietic reconstitution giving rise to lymphocytes, myeloid cells and most conventional or classical dendritic cells (cDCs), but the emergence of plasmacytoid dendritic cells (pDCs) is completely blocked (Cisse et al., 2008). Additionally, the role of TCF4 in lymphocyte development is underscored by the demonstration that *Tcf4* deficiency leads to a partial block in B and T lymphocyte development in mice (Zhuang et al., 1996, Bergqvist et al., 2000).

Although TCF4 is not required for lineage establishment of B lymphocytes, it plays a role in the generation of normal numbers of pro-B cells, as supported by the following findings: first, B-lineage cells are present in *Tcf4* knockout fetuses

although the number of fetal liver pro-B cells is approximately half of the wild-type levels (Zhuang et al., 1996); second, *Tcf4* null fetal liver cells can reconstitute all B cell subpopulations in irradiated mice, albeit with a reduction in cellularity (Wikstrom et al., 2006). Specifically, in the bone marrow of reconstituted mice the numbers of *Tcf4*$^{-/-}$ pro- and pre-B cells are reduced to 50% of the wild-type levels, and in the spleen a 30% reduction is seen in the number of follicular B cells, but not marginal zone B cells. Additionally, *Tcf4* null fetal liver cells have a disadvantage to reconstitute B cell compartments when developing together with competing wild-type cells and *Tcf4*$^{-/-}$ cells are underrepresented among B cells in the spleen of *Tcf4*$^{+/-}$ *Tcf4*$^{-/-}$ mosaic mice, pointing to a role for TCF4 in pro-B cell expansion (Bergqvist et al., 2000, Wikstrom et al., 2006). In line with the latter, experiments with cultured mouse fetal liver cells have demonstrated that TCF4 is required for interleukin-7 responsiveness and optimal expansion of B cells in a dose dependent manner (Wikstrom et al., 2006).

Similar competitive disadvantage has been attributed to *Tcf4*$^{-/-}$ cells in reconstituting T cell compartments in irradiated wild-type mice and contributing to T cell populations in spleen and thymus of *Tcf4*$^{+/-}$ *Tcf4*$^{-/-}$ mosaic mice (Bergqvist et al., 2000, Wikstrom et al., 2008). Using fetal thymus organ culture, it has been demonstrated that *Tcf4*-deficient thymocytes display a dose-dependent partial block at the 3rd double negative (CD4^{-} CD8^{-} CD44^{-} CD25^{+}) stage of development (Wikstrom et al., 2008). At this stage β-chain of the T cell receptor has undergone rearrangement and is paired with the invariant pre-T cell antigen receptor α-chain (PTCRA) for β-selection. As a partial cause of the phenotype, *Ptcra* levels have been found to be reduced in the fetal thymi of *Tcf4*-deficient mice (Wikstrom et al., 2008).

It has to be noted that all E-proteins are intimately involved in lymphocyte development and there is considerable amount of redundancy in their functions (reviewed in Sun, 2004, Murre, 2005, Kee, 2009). The dose of E-proteins rather than specific E-protein appears to be important for lymphocyte development (Zhuang et al., 1996, Zhuang et al., 1998). Additionally, the major E-box-binding proteins in B cells are E47 homodimers and E-box-binding complexes in T cells are predominantly composed of TCF3/E2A and TCF12/HEB heterodimers (reviewed in de Pooter and Kee, 2010). Therefore, many of the TCF4 functions in lymphocyte development described above could at least partially arise from the modulation of the DNA-binding pool of other E-protein dimers through competitive dimerization with ID proteins.

In contrast to the above, TCF4 is a major component of E-box-binding complexes in plasmacytoid dendritic cells, an innate immune cell type specialized in the production of type I interferons (IFN) in response to viral nucleic acids (Cisse et al., 2008). pDCs share common bone marrow progenitors with antigen-presenting cDCs, but differ from cDCs by their secretory morphology resembling plasma B cells (Reizis, 2010). Compared to other thymic and splenic cells, *TCF4* expression is especially high in pDCs in human and mouse (Cisse et al., 2008, Nagasawa et al., 2008, Robbins et al., 2008). It has been demonstrated that in mouse dendritic cell progenitors *Tcf4* expression is upregulated by cytokine FLT3L

and STAT3 mediated signalling that supports pDC development, and downregulated by GM-CSF mediated signalling that inhibits pDC maturation (Li et al., in press). Importantly, as described below, TCF4 is an essential and specific transcriptional regulator of pDC development, pDC-mediated interferon response, and maintenance of pDC fate.

Although the bone marrow progenitors in germline *Tcf4* knockout mouse initiate development towards the pDC cell lineage, they fail to differentiate and accumulate as an immature population. Loss of TCF4 abolishes pDC development selectively and in a cell-intrinsic fashion since conditional knockout of *Tcf4* in the adult mouse or specifically in the DC-lineage cells impairs pDC development but does not affect the relative content of other major cell types in spleen and bone marrow (Cisse et al., 2008, Cervantes-Barragan et al., 2012). Additionally, by inducing simultaneous loss of *Tcf3/E2a* and *Tcf12/Heb* in adult mice, other E-proteins have been found to be dispensable for pDC development. Moreover, it has been demonstrated that *TCF4* hemizygosity in mice and in human patients with Pitt-Hopkins syndrome (PTHS) results in impaired development and abnormal phenotype of pDCs, indicating that *TCF4* is haploinsufficient for pDC development and that the role of TCF4 in pDC generation is conserved between mice and humans (Cisse et al., 2008). The latter is also supported by the demonstration that in cultures of human postnatal thymic progenitors, forced expression of TCF4-B stimulates pDC development whereas silencing of *TCF4* impairs pDC generation (Nagasawa et al., 2008).

In all above mentioned *Tcf4* deficient animal models and PTHS patients, lack of pDCs is directly linked to eliminated or impaired pDC-dependent interferon response to the viral mimetic unmethylated CpG containing DNA (Cisse et al., 2008, Cervantes-Barragan et al., 2012). This in turn should be reflected in deficits in immune responses to viral infections. Indeed, loss of pDCs induced by conditional targeting of *Tcf4* in mice results in defective protection against mouse hepatitis virus, an acute cytopathic virus, and impaired adaptive immune response to chronic infection of lymphocytic choriomeningitis virus, a rapidly replicating noncytopathic virus (Cervantes-Barragan et al., 2012).

TCF4 directly controls the expression of multiple genes enriched in pDCs or necessary for pDC function as assessed by gene expression profiling in heterozygous knockout mice and chromatin immunoprecipitation in a human pDC cell line (Cisse et al., 2008, Ghosh et al., 2010 and Table 2). Notably, these include genes encoding for a key IFN-inducing factor IRF7 (interferon regulatory factor 7), Toll-like receptors TLR7 and TLR9, and pDC-enriched transcription factors SPIB and IRF8. In addition, in pDCs TCF4 represses the expression of a number of cDC-enriched genes and regulates transcription from several genes commonly expressed in lymphoid cells. Therefore, TCF4 has been suggested to be responsible for the induction of the lymphoid-cell like features of pDCs that differentiates them from other innate immune cells such as NK cells and cDCs that depend on E-protein inhibition for their development (Reizis, 2010).

Activation of pDCs by unmethylated CpG containing DNA induces downregulation of TCF4 expression, consistent with their differentiation into cDCs

(Cisse et al., 2008). Likewise, deletion of *Tcf4* from mature peripheral pDCs causes their spontaneous differentiation into cells with cDC properties including morphology, capacity to activate $CD4^+$ T cells and expression of cDC-associated genes (Ghosh et al., 2010). Additionally, reduced TCF4 levels in heterozygous knockout mice cause phenotypic conversion of peripheral pDCs. Thus, continuous expression of TCF4 is required to maintain pDC cell fate and to prevent spontaneous cell fate conversion and eventual loss of pDCs. Intriguingly, a subgroup of cDCs that share developmental pathways and similarities in gene expression profile with pDCs, are absent from chimeric mice established using *Tcf4* deficient fetal livers, suggesting that TCF4 may be crucial for other immune cell types in addition to pDCs (Bar-On et al., 2010).

3.2.3 Functions of TCF4 in the nervous system

Functioning of neurogenesis initiating proneural bHLH proteins (e.g. AC-S, Atonal and Neurogenin families) and bHLH proteins involved in terminal neuronal differentiation (e.g. NeuroD family), is dependent on heterodimerization with E-proteins, including TCF4 (reviewed in Guillemot, 2007). Although E-proteins have been suggested to compensate functionally for each other during the development of the nervous system (Ravanpay and Olson, 2008), recent evidence points towards distinct non-redundant roles carried out by specific E-protein containing heterodimers (Flora et al., 2007). Additionally, it has not been ruled out that E-proteins may participate in neurogenesis independently of heterodimerization with other bHLH proteins, e. g. by supporting the migration of precursor cells (Ik Tsen Heng and Tan, 2003).

The described target genes of TCF4 in the nervous system include *FGF1* (fibroblast growth factor 1) (Liu et al., 1998), *TH* (tyrosine hydroxylase) (Yoon and Chikaraishi, 1994) and *SSTR2* (somatostatin receptor 2) (Pscherer et al., 1996), but given its heterodimerization capability, numerous other genes are most likely regulated by TCF4 in the brain. TCF4 has been demonstrated to heterodimerize with several HLH proteins that play important roles in the development of the nervous system (Einarson and Chao, 1995, Persson et al., 2000, Jogi et al., 2002, Flora et al., 2007, Ravanpay and Olson, 2008, Brzozka et al., 2010). These are NEUROD1 (neurogenic differentiation 1, alias BETA2), NEUROD2 (alias NDRF), ATOH1 (atonal homolog 1, alias MATH1), HAND2 (heart and neural crest derivatives expressed transcript 2, alias DHAND), ASCL1, HES1 (hairy and enhancer of split 1) and all ID factors (Table 3). The importance of TCF4 in the nervous system has been revealed by studies on *tcf4* morphant zebrafish, *Tcf4* knockout and transgenic mice, and human patients with Pitt-Hopkins syndrome.

In zebrafish, reduction of *tcf4* expression leads to delayed development of the brain and eyes, but does not affect early neural patterning or regionalization of the forebrain (Brockschmidt et al., 2011). *tcf4* morphant embryos have an enlarged rhombencephalic ventricle and their neuronal development is impeded in several anatomical areas. The eyes of *tcf4* morphants are small and characterized by less structured inner nuclear layer of bipolar and amacrine cells, thinner inner plexiform

layer (a network of processes between the inner nuclear layer and ganglion cells), increased ocular length and a delay in the differentiation of the lens (Brockschmidt et al., 2011).

Neurological phenotypes of *Tcf4* null mice have not been characterized in-depth, but it is known that they have grossly normal brains that do not show any evident morphological defect at postnatal day P0 (Zhuang et al., 1996, Flora et al., 2007). However, the development of ATOH1-dependent neurons of the pontine nucleus is selectively and in a dose-dependent manner disrupted in heterozygous and homozygous *Tcf4* knockout mice due to a delay or interruption in the migration of neuronal progenitors in the anterior extramural migratory stream (Flora et al., 2007). *Tcf4* and other E-protein genes are coexpressed with *Atoh1* in the embryonic rhombic lip containing a population of neuronal progenitors that gives rise to the granule cell layer of the cerebellum, external cuneate nucleus, pontine nucleus and lateral reticulate nucleus. Deletion of *Tcf4*, but not *Tcf3/E2a* or *Tcf12/Heb* has been demonstrated to cause a differentiation defect only in the pontine nucleus, leaving the other ATOH1-dependent structures intact. Additionally, ASCL1-dependent progenitors that give rise to *locus coeruleus*, the main noradrenergic structure of the brain, differentiate properly in the absence of *Tcf4* (Flora et al., 2007). Altogether this reveals an exclusive function for TCF4-ATOH1 heterodimer in the development of the nervous system, uncovering intrinsic molecular differences between E-proteins and illustrating the potential complexity of the functions carried out by different TCF4 heterodimers in the nervous system.

To study the potential function of TCF4 in higher-order cognitive processes, transgenic mice with mild postnatal *Tcf4-B* overexpression in the forebrain have been produced (Brzozka et al., 2010). These mice have altered circadian expression profile of *Per2* (period homolog 2), *Id2*, *Neurod1*, *-2* and *-6* in the hippocampus, reduced *Fos* (FBJ murine osteosarcoma viral oncogene homolog) expression in the hippocampus and anterior cingulate cortex, and several behavioural deficits (Brzozka et al., 2010, Brzozka and Rossner, 2013). First, *Tcf4* transgenic mice have impaired sensorimotor gating monitored in prepulse inhibition paradigm. Second, the transgenic mice show reduced freezing in contextual and cued fear-conditioning tests in both delay and trace paradigms, suggesting deficient fear-related learning and memory. Third, these mice have a minor impairment in reversal training in Morris water maze indicating slight impairment in cognitive flexibility. And fourth, moderate *Tcf4* overexpression may lead to altered level of anxiety as revealed in open field test performed in dark phase. These observations by Brzozka et al. implicate TCF4 in the regulation of cognitive functions and circadian processes.

Heterozygous *TCF4* mutations in humans with Pitt-Hopkins syndrome cause widespread neurological defects without obvious defects in the anatomical structure of the nervous system. This syndrome, as well as other human disorders associated with TCF4 is further discussed in the following chapter.

3.3. Association of TCF4 with human disorders

TCF4 has been implicated in several illnesses ranging from neurodevelopmental disorders to tumorigenesis. Particularly, *TCF4* gene is mutated in Pitt-Hopkins mental retardation syndrome, its loss contributes to the phenotype of patients with chromosome 18q deletions, and variation in *TCF4* confers susceptibility for schizophrenia, Fuchs' corneal dystrophy and possibly for bipolar disorder. Additionally, TCF4 plays a role in both development and suppression of cancer. The underpinnings of these associations are further described below.

3.3.1 Pitt-Hopkins syndrome

Mutations in *TCF4* gene cause Pitt-Hopkins syndrome (PTHS; OMIM 610954), a rare autosomal dominant disorder characterized by motor and mental retardation, typical facial features and breathing anomalies. PTHS was first described already in 1978 (Pitt and Hopkins, 1978), but in almost 30 years following the initial study only a few patients were reported as possible cases of PTHS (Singh, 1993, Van Balkom et al., 1998, Orrico et al., 2001, Peippo et al., 2006). In 2007 the genetic cause of PTHS was identified (Amiel et al., 2007, Brockschmidt et al., 2007, Zweier et al., 2007) simplifying the diagnosis, and to date more than 130 genetically confirmed patients have been described (Andrieux et al., 2008, Giurgea et al., 2008, Zweier et al., 2008, de Pontual et al., 2009, Rosenfeld et al., 2009, Taddeucci et al., 2010, Takano et al., 2010, Hasi et al., 2011, Lehalle et al., 2011, Marangi et al., 2011, Takano et al., 2011, Armani et al., 2012, Marangi et al., 2012, Need et al., 2012, Takenouchi et al., 2012, Talkowski et al., 2012, Verhulst et al., 2012, Whalen et al., 2012, Willemsen et al., 2012, Ghosh et al., in press, Maini et al., in press, Steinbusch et al., in press). The mutational spectrum of PTHS includes frameshift (26% of patients), splice-site (9%), nonsense (14.5%) and missense (17.5%) mutations in *TCF4*, as well as translocations (2%) and large deletions encompassing all (21%) or part (10%) of the *TCF4* gene. Additionally, three patients with milder clinical phenotype than classical PTHS have been demonstrated to carry a reading frame extending frameshift mutation, a nonsense mutation or a translocation involving *TCF4* gene (Kalscheuer et al., 2008, Hamdan et al., in press, Steinbusch et al., in press). The PTHS-associated mutations are mostly of *de novo* origin, although three cases have been reported when the mutation has been inherited from a mother with somatic mosaicism (de Pontual et al., 2009, Steinbusch et al., in press). Apart from one splice site mutation, three nonsense and three missense mutations that have been reported recurrently in unrelated patients, most of the PTHS-associated mutations are private. Since the symptoms of PTHS seem to be similar in patients carrying whole gene deletions or point mutations in one *TCF4* allele, haploinsufficiency has been proposed as an underlying mechanism of PTHS regardless of mutation type. Some support for this concept has been provided by two studies that have found the ability of TCF4-B to activate E-box-dependent transcription together with ASCL1 to be similarly reduced by two different frameshift, two nonsense and three missense mutations in

TCF4 (Zweier et al., 2007, de Pontual et al., 2009). Additionally, compared to controls, *TCF4* expression has been shown to be reduced by half in lymphoblastoid cell lines from monozygotic twins carrying a balanced *TCF4* disrupting translocation (Talkowski et al., 2012).

Nearly all of the PTHS patients have a specific facial appearance that includes deep-set eyes, broad nasal bridge, wide mouth with widely spaced teeth, Cupid's bow and/or tented upper lip, everted lower lip and fleshy ears (Zweier et al., 2008, Marangi et al., 2011, Whalen et al., 2012). The frequent limb anomalies present in more than half of the PTHS patients include long fingers, single palmar crease, fetal finger pads and flat feet. Most patients are born after an uneventful pregnancy with normal growth parameters. The height and weight of the majority PTHS patients is in normal range, although more frequently below than above the average (Whalen et al., 2012). All PTHS patients have severe intellectual disability with absent language, and impaired motor development with walking delayed or not achieved (Van Balkom et al., 2012, Whalen et al., 2012). Eighty *per cent* of PTHS patients have hypotonia, 75% have constipation and 60% have unsteady or ataxic gait. Out of the ocular anomalies strabismus is present in 60% and myopia in 55% of patients. Fifty-five *per cent* of PTHS patients have breathing anomalies including hyperventilation episodes and apnoea, and 40% of patients have a history of seizures (Whalen et al., 2012). The behavioural phenotype of PTHS is characterized by impairments in communication and socialization ability, smiling appearance (90% of PTHS patients), stereotypic movements (77%), agitation (68%), anxiety (57%) and harm to self (52%) (Van Balkom et al., 2012, Whalen et al., 2012). The behavioural difficulties of PTHS patients have been found to be beyond what would be expected for the low cognitive level, fitting a classification of autism spectrum disorder (Van Balkom et al., 2012). The most common abnormalities found in cerebral magnetic resonance imaging are agenesis or hypoplasia of *corpus callosum* (40%), ventricular dilatation (31%) and minor posterior fossa abnormalities (28%) (Whalen et al., 2012). Additionally, it has been suggested that PTHS patients have an increased vulnerability to recurrent respiratory tract infections, consistent with low numbers of functional blood plasmacytoid dendritic cells and defective interferon response to unmethylated DNA (Cisse et al., 2008).

PTHS-like clinical phenotype can be caused by recessive defects in *CNTNAP2* (contactin associated protein-like 2) and *NRXN1* (neurexin 1) genes that belong to *NRXN* gene superfamily and encode neuronal cell–cell interactions-mediating transmembrane proteins. However, because of different genetic basis these disorders are denoted as Pitt-Hopkins-Like syndromes 1 (OMIM 610042) and 2 (OMIM 600565), respectively (Zweier et al., 2009). Additionally, PTHS has some phenotypical overlap with Angelman syndrome (OMIM 105830), Rett syndrome (OMIM 312750), Mowat-Wilson syndrome (OMIM 235730) and X-linked alpha-thalassemia/mental retardation syndrome (ATR-X; OMIM 301040) (Zweier et al., 2008, Takano et al., 2010, Marangi et al., 2011, Takano et al., 2011). Intriguingly, a patient with developmental delay has been demonstrated to carry a 18q21.2-18q21.31 duplication that encompasses *TCF4* gene (Talkowski et al., 2012).

So far, the treatment of PTHS patients has focused on epileptic seizure management and habilitation. In addition, valproate and acetazolamide have been found to improve PTHS breathing abnormalities (Verhulst et al., 2012, Maini et al., in press).

3.3.2 Role of *TCF4* in partial hemizygosity of chromosome 18q

Absence of one of the *TCF4* alleles can contribute to the phenotype of individuals with interstitial or terminal deletions of the long arm of chromosome 18. Partial hemizygosity of 18q (18q-) is a relatively common chromosomal segmental aneusomy with an estimated frequency of 1 per 40,000 live births. 18q- stands for a heterogeneous group of deletions that differ greatly in size and content, and therefore give rise to varied clinical features (Cody et al., 2009a, Cody et al., 2009b, Heard et al., 2009). When a deletion includes the *TCF4* gene, many of the physical characteristics and the impairment of motor and cognitive development are similar to those of PTHS (Gustavsson et al., 1999, Andrieux et al., 2008, Kato et al., 2010, Hasi et al., 2011, Stavropoulos et al., 2011, Takenouchi et al., 2012). By comparing the $TCF4^{+/-}$ patients with the $TCF4^{+/+}$ patients in the context of 18q-, it has been suggested that for predicting the severity of the physical and mental impairment of the individuals with 18q deletions, the presence or absence of *TCF4* is more important than the size of the deletion (Hasi et al., 2011). Particularly, hemizygosity of *TCF4* has been found to result in the developmental ceiling of twelve months of age. The phenotypic components that have been shown to be unique to those with *TCF4* hemizygosity also include *corpus callosum* abnormalities and higher probability of premature death due to aspiration-related complications (Hasi et al., 2011). In addition, individuals with 18q deletions that include *TCF4* together with two other genes *NETO1* (neuropilin- and tolloid-like protein 1) and *FBXO15* (F-box protein 15), are more likely to exhibit autistic-like behaviour (O'Donnell et al., 2010). Conversely, *TCF4* hemizygosity is not involved in many other common features of 18q- patients e.g. kidney malformation, congenital ear deformities, dysmyelination or growth hormone response deficiency (Cody et al., 2009b).

3.3.3 *TCF4* and genetic anticipation in the major psychoses

The traditional Kraepelinian classification system divides major psychoses of adulthood into schizophrenia (SCZ) and bipolar affective disorder (BPAD), although this dichotomy is now being gradually replaced by a concept of continuum of conditions (Craddock and Owen, 2010). Both SCZ and BPAD have severe effects on thought, perception, emotion and behaviour, but are clinically set apart based mainly upon whether psychotic or affective symptoms predominate and whether the psychosis is of chronic or fleeting nature (Ivleva et al., 2008). SCZ is characterized by the occurrence of positive symptoms such as hallucinations, delusions and disorganized thoughts; negative symptoms such as affective flattening, social withdrawal and lack of motivation; and cognitive symptoms such

as poor executive functioning, trouble focusing and decline in working memory. On the other hand, BPAD has been principally considered as a mood disorder featuring major depressive episodes alternated with phases of mania, with cognitive, behavioural and psychotic symptoms worsening at the peaks of the stages. According to the World Health Organization SCZ and BPAD are among the leading 10 causes of the global burden of disease in the age group of 15-44 year-olds. Both disorders have a lifetime prevalence of ~1% and estimated heritability of ~80% (Sullivan et al., 2003, Craddock and Forty, 2006). Research into the genetic basis of SCZ and BPAD has provided evidence for an overlap in genetic susceptibility across the traditional psychosis categories and for common genetic variation as a substantial component to the risk of the disorders involving thousands of allelic variants with very small effect (Lichtenstein et al., 2009, Moskvina et al., 2009, Purcell et al., 2009).

Several studies have reported suggestive evidence for genetic anticipation in SCZ and BPAD, but there are a number of methodological issues related to ascertainment biases and statistical analyses that have rendered the findings inconclusive (reviewed in Vincent et al., 2000, Goossens et al., 2001, Fortune et al., 2003, O'Donovan et al., 2003). The concept of anticipation refers to earlier age at onset or increased disease severity in successive generations. At the molecular level this phenomenon could be explained by unstable DNA mutations, for example triplet repeat expansions. Indeed, by using a ligase chain reaction based method of repeat expansion detection (RED) initial studies found that people with BPAD or SCZ carry larger CAG/CTG repeats than controls (Lindblad et al., 1995, O'Donovan et al., 1995). This finding was confirmed by several but not all of the following studies (reviewed in Vincent et al., 2000, O'Donovan et al., 2003).

A substantial portion (7-19%) of the extended repeats detected by RED map to an intron of *TCF4* gene (Lindblad et al., 1998, Guy et al., 1999, Verheyen et al., 1999). The CTG repeat length in this locus is highly polymorphic with the heterozygosity of 84% (Breschel et al., 1997). Alleles ranging from 10-37 repeats are common and considered stable, whereas expanded alleles of more than 53 repeats are unstable (Breschel et al., 1997). The control frequency of 3-7% has been observed for the *TCF4* alleles with >40 repeat sequences (McInnis et al., 2000, Del-Favero et al., 2002). The association between alleles with more than 40 repeats and BPAD was found in an early study (Lindblad et al., 1998), but was not replicated in many of the following (reviewed in Goossens et al., 2001, O'Donovan et al., 2003). In a later study that used the largest sample size of more than 400 European cases and controls, modest evidence for association was seen after stratification for first degree family history and disease severity (Del-Favero et al., 2002). Additionally, *TCF4* alleles with expanded repeats have been found to segregate with BPAD in one family that shows genetic anticipation (Pato et al., 2000). In case of SCZ, most studies have found no association with *TCF4* alleles carrying expanded CTG repeats (Vincent et al., 1999, Bowen et al., 2000, McInnis et al., 2000, Ayton et al., 2002), though one study suggested a link between the expansion and childhood-onset SCZ in males (Burgess et al., 1998).

Altogether, although some evidence exists for association of the extended CTG repeats in *TCF4* with BPAD, it has not yet been convincingly demonstrated. Given the complex inheritance pattern of the major psychoses and small risks conferred by individual variants, it remains possible that the performed studies suffer from lack of power due to insufficient subject numbers. Of note, a recent copy number variation analysis identified a duplication of 3′ part of the *TCF4* gene in a patient with major depressive disorder (Ye et al., 2012), supporting the potential link between *TCF4* and mood disorders.

3.3.4 Association of *TCF4* with schizophrenia

Recent genome-wide association (GWA) studies with the sample sizes in the order of tens of thousands have shown that common variants at *TCF4* locus are among the most prominent susceptibility factors for schizophrenia (Stefansson et al., 2009, Ripke et al., 2011, Steinberg et al., 2011). GWA studies are a potentially powerful approach for identification of common risk factors of complex polygenic diseases, but have so far ascertained only 11 SCZ-associated loci, variants in which have surpassed the level of genome-wide significance (Stefansson et al., 2009, Ripke et al., 2011, Steinberg et al., 2011, Williams et al., 2011b, Yue et al., 2011). Notably, these include the extended *MHC* (major histocompatibility complex), *miR-137* (microRNA 137) and *TCF4* loci.

In *TCF4* locus, four SCZ-associated single nucleotide polymorphisms (SNPs) have been identified by GWA studies in different European-ancestry populations – rs9960767 and rs17512836 in *TCF4* intronic sequence have odds ratios of 1.20 and the risk allele control frequencies are 6% and 2%, respectively; rs4309482 and rs12966547 located ~140 kb downstream of *TCF4* gene have odds ratios of 1.09 and the risk allele control frequencies are 60% (Stefansson et al., 2009, Ripke et al., 2011, Steinberg et al., 2011). The association of the intronic SNP rs9960767 is further supported by the finding that a neighbouring SNP rs2958182, which is in complete linkage disequilibrium with rs9960767, confers risk of schizophrenia in the population of Han Chinese (odds ratio 0.78, Li et al., 2010). The importance of the region downstream of *TCF4* encompassing rs4309482 and rs12966547 is highlighted by the demonstration that the same region contains the breakpoint of a pericentric inversion inv(18)(p11.31;q21.2) occurring in a Scottish family in which carriers have been diagnosed with SCZ and severe learning disability (Pickard et al., 2005). Additionally, *TCF4* has been suggested to be a target of another schizophrenia-associated gene that encodes *miR-137* (Ripke et al., 2011, Kwon et al., in press).

It has not been established yet whether there is a correlation between SCZ-associated *TCF4* variants and *TCF4* expression. On the one hand, *TCF4* expression in *post-mortem* adult cortex has been found not to be influenced by the SCZ-associated intronic SNP rs9960767 (Williams et al., 2011a). On the other hand, several lines of evidence suggest that *TCF4* expression is influenced by genetic variation *in cis* and altered in SCZ. First, allelic expression imbalance at *TCF4* exonic SNP rs8766 has been demonstrated in different brain regions (Buonocore et

al., 2010), and *TCF4* transcript levels in frontal cortex and thalamus have been found to be associated with three SNPs in *TCF4* introns and 3′ UTR (Kim et al., 2012). This collectively indicates that *cis*-regulatory variation in *TCF4* exists. Second, compared to the control levels, *TCF4* mRNA expression has been found to be increased by ~50% in *post-mortem* adult cerebellar cortices of SCZ patients (Mudge et al., 2008), by ~2% in the blood of psychosis patients (Wirgenes et al., 2012), and by 2.8-fold in induced pluripotent stem cells-derived neurons of SCZ patients (Brennand et al., 2011). Additionally, treatment of mice with phencyclidine, an inducer of SCZ-like symptoms, leads to increased *Tcf4* expression in the nucleus accumbens (Le-Niculescu et al., 2007, Ayalew et al., 2012). Conversely, decreased *TCF4* mRNA levels in blood have been connected to high delusion state in patients with SCZ or related disorders (Kurian et al., 2011), and ~20% decrease in *TCF4* mRNA expression has been reported in the blood of SCZ patients compared to controls (Navarrete et al., in press).

Despite that the mechanisms underpinning the association of *TCF4* risk variants with SCZ have remained unclear, the SNPs have been demonstrated to have different effects on several endophenotypes of SCZ – impaired sensorimotor gating, sensory gating, verbal fluency and verbal declarative memory (VDM). First, among both patients and healthy individuals the carriers of the rs9960767 risk allele have decreased prepulse inhibition (PPI) of the acoustic startle response (Quednow et al., 2011). Moreover, the association of rs9960767 genotype with PPI levels is stronger (odds ratio 4.81 using criterion of 1 SD below normal for decreased PPI) than association with SCZ *per se* (Quednow et al., 2011). Second, auditory sensory gating is reduced in carriers of the risk alleles of intronic SNPs rs9960767 and rs17512836, especially among heavy smokers, as measured by P50 suppression in paired click paradigm in healthy subjects (Quednow et al., 2012). The odds ratios using criterion of 1 SD below normal for decreased P50 suppression are 1.8-1.9 in total population and 3.2-5.5 among heavy smokers (Quednow et al., 2012). Third, the risk variants at rs12966547 and rs4309482 downstream of *TCF4* have been shown to be associated with poorer verbal fluency in the combined sample of psychosis patients and controls (Wirgenes et al., 2012). And fourth, there is a significant association with better performance in recognition and a trend toward higher performance in delayed recall in a test for VDM in SCZ patients carrying the rs9960767 risk allele (Lennertz et al., 2011). Although elevated risk for SCZ appears not to be conferred by rs9960767-mediated VDM impairment, the results of Lennertz et al. 2011 suggest a role for *TCF4* in learning and memory.

3.3.5 Association of *TCF4* with Fuchs' endothelial corneal dystrophy

Fuchs' endothelial corneal dystrophy (FECD; MIM 136800) is a slowly progressive degenerative disorder of the corneal endothelium (CEn) that affects 4-11% of the population over the age of 40 years (Klintworth, 2009, Thalamuthu et al., 2011, Eghrari et al., 2012). CEn localized at the posterior corneal surface of the eye consists of a monolayer of non-dividing hexagonal cells, developmentally

derived from the neural crest (Hayashi et al., 1986). CEn secretes collagen that forms basal lamina termed Descemet's membrane, draws water from the corneal stroma to maintain it in slightly dehydrated state that is required for optical transparency, and allows nutrients to enter the stroma that lacks blood vessels (reviewed in Bourne, 2003). FECD is characterized by decrease in CEn cell density and presence of outgrowths of Descemet's membrane called guttae. Oxidative stress leading to apoptosis of CEn cells has been suggested to play a key role in FECD pathogenesis (Elhalis et al., 2010, Jurkunas et al., 2010). The disorder results in corneal edema with a loss of corneal clarity, painful episodes of corneal erosions, loss of visual acuity and blindness (reviewed in Klintworth, 2009, Elhalis et al., 2010, Schmedt et al., 2012). At present, corneal transplantation using *post mortem* donor tissue is the only definitive treatment for FECD (Schmedt et al., 2012).

FECD heritability has been estimated to be almost 40% (Louttit et al., 2012). The disease has a female predominance at a 2.5-3:1 ratio and can be divided into familial early-onset and familial or sporadic late-onset variants (Kannabiran, 2009, Elhalis et al., 2010, Schmedt et al., 2012). The rare early-onset FECD starts at first decade of life and some cases have been found to be caused by mutations in *COL8A2* (collagen 8 alpha 2) gene, whereas the more common late-onset FECD begins in the fifth decade of life. A variety of SNPs in *TCF4* locus have been found to be associated with late-onset FECD of sporadic type. An exceptionally strong association of genome-wide significance was identified for intronic SNP rs613872 by GWA study in a relatively small discovery group of less than 500 cases and controls (Baratz et al., 2010) and replicated by several independent studies in different European ancestry populations (Li et al., 2011b, Riazuddin et al., 2011, Eghrari et al., 2012, Igo et al., 2012, Kuot et al., 2012, Stamler et al., in press). Rs613872 has an odds ratio of ~4-5.5 and the risk allele frequency is 20%. Besides rs613872, 4 more intronic SNPs in *TCF4* (rs618869, rs581653, rs17089887, rs17595731) and 3 upstream SNPs (rs2286812, rs9954153, rs17089925) have been found to be associated with sporadic FECD in European ancestry or Chinese populations, indicating that multiple independent *TCF4* haplotypes confer FECD susceptibility (Baratz et al., 2010, Riazuddin et al., 2011, Thalamuthu et al., 2011, Kuot et al., 2012).

Familial late-onset FECD has been mapped to several loci by linkage studies (Elhalis et al., 2010), the most common of these loci, *FCD2* located at 18q21.2-21.3, includes the *TCF4* gene (Sundin et al., 2006, Li et al., 2011b). Although the risk allele of rs613872 was found not to co-segregate with the disease-phenotype in three *FCD2*-linked families (Riazuddin et al., 2011) and a missense change in *LOXHD1* (lipoxygenase homology domains 1) was identified as a responsible variant in one of the *FCD2*-linked families (Riazuddin et al., 2012), further research is needed to ascertain the possible role of *TCF4* variants in familial FECD.

The pathogenic mechanisms underlying the involvement of *TCF4* in FECD have not been elucidated. Studies in healthy young adults have shown that the rs613872 risk allele is not associated with CEn cell density (Mackey et al., 2012). It has been proposed that rs613872 affects disease progression with the minor allele

acting as a promoting and/or the major allele representing a protecting factor for a shift from subclinical to clinical disease (Eghrari et al., 2012). Interestingly, TCF4 is known to upregulate *ZEB1* (Sobrado et al., 2009), mutations in which are also a rare cause of sporadic and familial late-onset FECD (Mehta et al., 2008, Riazuddin et al., 2010), and both TCF4 and ZEB1 are implicated in the regulation of EMT by repressing E-cadherin expression (Grooteclaes and Frisch, 2000, Eger et al., 2005, Sobrado et al., 2009). Therefore, a common pathway relevant to the FECD pathogenesis has been suggested, possibly involving impaired migration of corneal stem cells or increased production of extracellular matrix by CEn cells (Wright and Dhillon, 2010).

3.3.6 TCF4 and tumorigenesis

As explicated in previous sections, TCF4 has been implicated in the regulation of many normal cellular processes that are frequently misregulated in cancer progression, e.g. cell cycling, EMT and angiogenesis. The controversial consequences of TCF4 on cell cycle progression parallel its opposite effects on tumorigenesis reported by different studies that have found TCF4 to participate in the relay of metastatic, oncogenic or tumour suppressing signals.

Support for a potential role of TCF4 in metastasis comes from studies that have identified TCF4 as a potent EMT inducer in epithelial cells (Sobrado et al., 2009) and as one of the factors mediating high metastatic activity in mouse xenograft models (Vicent et al., 2008, Appaiah et al., 2010). Particularly, TCF4 in combination with PRKD3 (protein kinase D3) and SUSD5 (sushi domain containing 5) or MCAM (melanoma cell adhesion molecule) has been demonstrated to increase bone colonization activity of lung large cell carcinoma cells by enhancing tumour-induced osteoclastic bone resorption and metalloproteolytic activity at tumour-host interface without affecting tumour growth or cell homing to bone (Vicent et al., 2008). Increased *TCF4* expression has been reported in invasive cervical carcinoma cell lines compared to non-invasive ones (Sobrado et al., 2009), in subset of breast cancer cell lines with mesenchymal traits (Blick et al., 2008), and in the endometrial carcinosarcomas' mesenchymal component compared to epithelial component (Castilla et al., 2010). Additionally, in transplantation experiments *Tcf4* levels have been demonstrated to be upregulated in mouse mammary tumour cells recovered from lung metastasis compared to cells recovered from primary tumour (Xiang et al., 2011) and silencing of *TCF4* in a CXCR4-enriched subpopulation of breast cancer cells has been found to reduce the capacity of these cells to form metastases in the lung (Appaiah et al., 2010).

The indication that TCF4 acts as an oncogenic transcription factor arises from the finding that TCF4-B is a downstream target of WNT/β-catenin signalling pathway (Kolligs et al., 2002) mutational defects in which have been detected in many cancers (reviewed in Polakis, 2000). Remarkably, it has been demonstrated that overexpression of TCF4-B enhances tumour growth of a specific subsets of breast cancer cells (Appaiah et al., 2010). Additionally, combined silencing of

TCF4 and *β-catenin* has been reported to induce apoptosis of colorectal cancer cells and regression of bulky tumours in nude mice (Mologni et al., 2010). Studies on human malignancies have revealed that *TCF4* expression is upregulated in primary ovarian endometroid adenocarcinomas with defective β-catenin signalling (Kolligs et al., 2002, Zhai et al., 2002). Moreover, high *TCF4* expression has been found to correlate with lower overall survival in lung squamous cell carcinoma (Cano and Portillo, 2010) and in estrogen receptor negative breast cancer (Appaiah et al., 2010).

In contrast to the above, several lines of evidence support the notion that *TCF4* acts as a tumour suppressor gene. Downregulation of *TCF4* expression has been reported in colon and gastric cancer cell lines (Kim et al., 2008, Herbst et al., 2009a), colorectal carcinomas (Herbst et al., 2009a, Herbst et al., 2009b) and gastric cancers (Kim et al., 2008). The underlying mechanisms involve *TCF4* allelic loss and histone deacetylation in case of colorectal cancers (Herbst et al., 2009a, Herbst et al., 2009b) or CpG hypermethylation and histone deacetylation at a *TCF4* 5' exon in case of gastric tumours (Kim et al., 2008, Joo et al., 2010). Since TCF4-B expression is low in normal intestinal epithelium, increased in colorectal adenomas but lost in carcinomas, a model has been proposed according to which *TCF4* is upregulated in early stages of carcinogenesis but is selected against in later stages to allow cancer progression (Herbst et al., 2009a). Thus, the effect of TCF4 depends to a large extent on cellular context and the question as to how *TCF4* can function as an oncogene and a tumour suppressor gene requires further investigation.

AIMS OF THE STUDY

Although the existence of multiple *TCF4* transcripts has been demonstrated, systematic analysis of human *TCF4* gene structure and alternative TCF4 isoforms has not been performed. Additionally, *TCF4* has been associated with several neurodevelopmental disorders, yet the functions of TCF4 in neurons have remained poorly described. Lastly, haploinsufficiency of *TCF4* has been proposed as an underlying mechanism for Pitt-Hopkins syndrome regardless of specific mutation, but extensive functional analyses to validate this concept are lacking. Thus, the aims of the current study were to:

(i) characterize human *TCF4* gene structure and *TCF4* expression at mRNA and protein level;

(ii) study the functional characteristics and regulation of alternative TCF4 isoforms in non-neural cells and neurons;

(iii) determine the impact of PTHS-associated mutations on TCF4 functions.

MATERIALS AND METHODS

Experimental procedures used in the current study are described in the original papers as indicated. The methods regarding the additional data not included in the papers are explicated in what follows.

- Bioinformatic analyses of gene, mRNA and protein sequences (I-III)
- Comparative modelling of protein structures (II)
- Ribonuclease protection assay (I)
- RNA isolation, reverse transcription and polymerase chain reaction (I-II)
- *In situ* hybridization (I)
- Recombinant DNA construction and site directed mutagenesis (I-III)
- *In vitro* translation (I-III)
- Electrophoretic mobility shift assay (II-III)
- Culturing of immortalized mammalian cells and rat primary neurons (I-III)
- Transfection of mammalian cells (I-III)
- Subcellular fractionation (II)
- Preparation of cell lysates and western blotting (I-III)
- Isoelectric focusing of proteins (III)
- Cyto- and histochemical immunostaining (I-III)
- Reporter assays (I-III)
- Immunoprecipitation and *in vitro* kinase assay (III)

Immunocytochemical analysis of TCF4 in rat primary neurons
Neurons were cultured on poly-l-lysine coated coverslips and transfected as described in studies II-III. In case of siRNA co-transfections 5 pmol siRNA and 0.25 µg pEGFP-C1 (Clontech) were used per well of a 48-well plate. Immunocytochemistry was performed according to the described protocols (I-II) using antibodies specific for TCF4 (1:200, CeMines), V5 (1:2000, Sigma-Aldrich), tubulin III (1:2000, Chemicon) and mouse/rabbit IgG conjugated with Alexa 405, 488, 546 or 568 (1:2000, Molecular Probes). If necessary, DNA was stained with DRAQ5 (1:5000, Biostatus Limited). Samples were mounted in ProLong Gold antifade reagent with or without DAPI (Molecular Probes) and analysed by confocal microscopy (LSM Duo, Zeiss).

RESULTS AND DISCUSSION

1. Structure of human *TCF4* gene (I)

Human gene for the transcription factor TCF4 is located on chromosome 18q21.2 and spans 437 kbs. In publication I, we demonstrated that it is transcribed using 21 mutually exclusive 5′ initial exons, many of which contain more than one transcription start sites and/or splice donor sites. These 5′ exons are located at various positions in the gene interspersed with internal exons 1–9, followed by constitutive internal exons 10–20 and 3′ exon 21. We designated *TCF4* 5′ exons with a lowercase letter preceded by a number that shows the following internal exon in the gene (e.g. 1a), and named *TCF4* transcripts according to the initial exon they contain. Usage of alternative 5′ exons gives rise to *TCF4* transcripts containing different number of internal exons and potentially encoding for TCF4 protein isoforms with 18 different N-termini. We named the isoforms TCF4-A – TCF4-R in agreement with the previous studies that have described three N-terminally distinct isoforms, i.e. TCF4-A, -B and -D (Corneliussen et al., 1991, Skerjanc et al., 1996).

Apart from differences in the N-termini, the number of TCF4 isoforms is increased by in-frame alternative splicing of internal exons. First, simultaneous skipping of cassette exons 8 and 9 gives rise to TCF4 Δ isoforms, as opposed to full-length isoforms that contain the amino acids coded by exons 8–9. Second, according to our and several previous studies (Corneliussen et al., 1991, Yoon and Chikaraishi, 1994, Skerjanc et al., 1996, Liu et al., 1998) there are two alternative splice donor sites at exon 18 that enable to splice in or out a 12 bps sequence encoding amino acid sequence RSRS present in + isoforms and absent in − isoforms. Third, exons 8 and 15 contain two alternative splice acceptor sites that lead to optional inclusion of the first three nucleotides (CAG) of an exon in mRNA and a glutamine or alanine residue, respectively, in the corresponding position in protein sequence.

There are several differences between the alternative TCF4 isoforms that can potentially lead to their functional divergence. The majority of TCF4 isoforms have a stretch of unique amino acids in their N-termini and the whole aminoterminal transactivation domain AD1 (coded by exons 3–6) is present only in isoforms with the longest N-termini such as TCF4-J, -K, -L and TCF4-B. The full-length DES region (coded by exons 6-11) is present in most isoforms except in the isoforms with the shortest N-termini, TCF4-D, -G, -H, -A and –I. Additionally, the CE part of the DES sequence is absent from the Δ isoforms. The RSRS insert is located in the calmodulin interacting region, opening a possibility that + and − isoforms could be differentially regulated by Ca^{2+} signalling. All TCF4 isoforms contain transactivation domain AD2 (coded by exons 14–16), Rep region (coded by exon 18) and the bHLH domain (coded by exon 19).

2. *TCF4* mRNA and protein expression (I)

To determine the abundance of mRNAs initiated at different positions within the *TCF4* gene we used ribonuclease protection assays. Relatively similar levels of transcripts containing different number of internal exons were detected in human cerebellum and muscle, indicating that transcription initiation at alternative sites within the *TCF4* gene is common.

We studied the usage of different 5′ exons in a variety of human tissues and brain regions by reverse transcription polymerase chain reaction. Most alternative 5′ exons containing transcripts were broadly expressed. Nevertheless, expression of transcripts containing 5′ exons 1a, 1b, 3a and 5c was detected only in a few tissues including testis and prostate. These exons originate from exonization of different transposable elements (TEs) or are located immediately behind a TE in the human genome. In accordance with the restricted expression of TE-dependent *TCF4* transcripts, TEs are known to be transcriptionally silenced in most mammalian tissues as a defense mechanism against potentially deleterious effects of their activity (Yoder et al., 1997, Zamudio and Bourc'his, 2010). The examination of mRNA levels of alternatively spliced *TCF4* transcripts in human tissues revealed that inclusion of exons 8-9 was prevalent and skipping of exons 8-9 a rare event in most tissues analysed and that the levels of transcripts containing or lacking the extra 12 bps of exon 18 were roughly equal in most tissues analysed.

To compare *TCF4* expression levels in different human tissues, we performed quantitative PCR with three pairs of primers designed to amplify all *TCF4* transcripts and determined the levels of four house-keeping genes for normalization. *TCF4* mRNA levels in the studied 23 tissues varied by two orders of magnitude. We detected the highest *TCF4* mRNA levels in fetal brain and adult cerebellum, the levels were high also in adult cerebral cortex, spleen, uterus, lung, thymus and placenta. The lowest quantities of *TCF4* transcripts were present in fetal liver, and adult pancreas and colon. *In situ* hybridization experiments on sections from adult human hippocampus and cerebellum demonstrated that *TCF4* mRNA is expressed in hippocampal neurons in dentate gyrus and CA1-CA3 regions, neurons of subiculum and parahippocampal gyrus of the cortex, and cerebellar granule neurons.

We studied the expression of endogenous TCF4 protein isoforms in different human tissue extracts by western blotting with TCF4-specific antibodies that recognize an epitope present in all described TCF4 isoforms. We validated the specificity of the TCF4 antibodies using a RNAi based approach in Neuro2A mouse neuroblastoma cells. By analysing extracts from human brain, lung, liver, kidney, muscle and testis we detected three prominent bands that were assigned as high, medium and low molecular weight (Mw) TCF4. In extracts of frontal cortex, hippocampus and cerebellum low Mw TCF4 was the major form, the levels of high or medium Mw TCF4 were elevated relative to other forms of TCF4 in testis or lung, respectively. Comparisons to individual TCF4 isoforms, produced using cloned coding regions of isoforms TCF4-A – TCF4-I, allowed us to make the following conclusions about the nature of the three bands in human tissue extracts:

the high Mw TCF4 fractionates similarly to TCF4-B and -C, the medium Mw TCF4 similarly to TCF4-D, and low Mw TCF4 to TCF4-A and TCF4-H.

Collectively our analyses corroborated the concept of TCF4 being a ubiquitous transcription factor. However, the levels of *TCF4* transcripts and the ratio of different TCF4 protein isoforms vary considerably between tissues. The highest *TCF4* mRNA levels are present in fetal brain, but expression remains elevated also in adult brain where we detected *TCF4* mRNA in neurons. These findings are consistent with earlier studies that have demonstrated *TCF4* expression in human or rodent nervous system (Soosaar et al., 1994, Chiaramello et al., 1995, Pscherer et al., 1996, Skerjanc et al., 1996, Uittenbogaard and Chiaramello, 2000, de Pontual et al., 2009, Brzozka et al., 2010). Given that a translocation involving 9 upper 5′ exons of *TCF4* has been described in a patient with mental retardation (Kalscheuer et al., 2008), and transgenic mice with mild overexpression of *Tcf4* in forebrain display deficits in fear conditioning and sensorimotor gating (Brzozka et al., 2010), it is evident that a narrow range of TCF4 protein activity is compatible with correct brain development and functioning. Since TCF4 acts as a central hub in the dimerization network of bHLH proteins in the central nervous system, it is likely that even slight changes in its levels can transform the balance in the system. All this substantiates the complex structure and the high number of 5′ exons in *TCF4* gene as these probably enable precise spatiotemporal regulation of *TCF4* expression.

3. TCF4 subcellular localization (I and unpublished)

To gain insight into possible functional variation among the alternative TCF4 isoforms we investigated the intracellular localization of over-expressed TCF4 proteins in HEK293 cells using confocal microscopy. By comparing the distribution of different TCF4 isoforms, and by analyzing mutated and EGFP-fused TCF4 proteins we discovered in the region coded by exons 8–9 a bipartite NLS that was able to direct nuclear transport when fused to EGFP. The NLS is responsible for nuclear localization of TCF4 isoforms with longer N-termini. It is conserved among E-proteins and mutating the second half of the NLS has been demonstrated to reduce nuclear import of TCF3/E2A (Lingbeck et al., 2005). Of note, the NLS partially overlaps with the CE sequence implicated in transcriptional repression (Herbst and Kolligs, 2008). TCF4 isoforms with shorter N-termini (TCF4-A, -H and -I) and Δ isoforms do not contain the NLS. According to our results from redirection assays with NLS- or NES-bearing TCF4 heterodimerization partners, the NLS-lacking TCF4 proteins are transported between the nucleus and cytoplasm by a piggy-back mechanism that is dependent on the integrity of the HLH domain. When TCF4 isoforms were over-expressed in HEK293 cells, the NLS-bearing full-length isoforms with longer N-termini (TCF4-B^+, -B^-, -C^-, -D^-, -E^-, -F^- and -G^-) were restricted to cell nucleus whereas NLS-lacking shorter isoforms (TCF4-A^-, -H^- and -I^-) and Δ isoforms (TCF4-$B\Delta^+$, -$B\Delta^-$ and -$C\Delta^-$) localized to the nucleus and cytoplasm. The nuclear import of the latter was likely mediated by endogenous NLS-bearing bHLH proteins that heterodimerize with TCF4. Comparable patterns

of distribution were observed when TCF4 isoforms were expressed in rat primary cortical and hippocampal neurons (Figure 2A). Additionally, the identified NLS in TCF4 was able to mediate nuclear transport of EGFP protein and structural integrity of the HLH was required for nuclear entry of EGFP-bHLH fusion protein in rat primary neurons (Figure 2B). Taken together, our results demonstrate that the intracellular localization of TCF4 is controlled in a similar manner in different cell types, but the distribution of TCF4 isoforms is differentially regulated depending on the presence or absence of the NLS. This in turn may have an effect on the stability, selection of dimerization partners, and ultimately the target genes of the alternative TCF4 isoforms. Our data are in accordance with previous studies demonstrating that nuclear localization for HLH complexes is a dominant process requiring only a single NLS per complex (Goldfarb and Lewandowska, 1994, Deed et al., 1996).

Examination of the intracellular distribution of endogenous TCF4 in human hippocampus and cerebellum by immunohistochemical staining using TCF4 antibodies revealed that TCF4-like signal was mainly in the nuclei but also in the cytoplasm of neurons. To demonstrate the specificity of the immunostaining achieved with TCF4 antibodies we studied the effect of *Tcf4* silencing on the signal intensity in rat primary cortical and hippocampal neurons. The siRNA transfected neurons were identified by the signal of EGFP produced from a co-transfected DNA vector. In mock and scrambled siRNA transfected neurons we detected most of the TCF4 signal in the nuclei, whereas in neurons transfected with *Tcf4*-specific siRNAs the nuclear staining was severely reduced (Figure 2C). This validates the specificity of the nuclear signal achieved with TCF4 antibodies in neurons. Considering the predominance of NLS-lacking low Mw TCF4 isoforms in brain tissue and mainly nuclear localization of TCF4 in neurons in human tissue slices and rat primary cultures, we suggest that most of endogenous TCF4 in neurons is heterodimerized with NLS-containing bHLH partners. Previously, distribution of endogenous TCF4 has been studied only in human placental tissue, where TCF4-like signal was detected in the cytoplasm of trophoblasts and endothelial cells; however, the specificity of the antibodies used was not ascertained (Liu et al., 2004).

4. Differential capacity of TCF4 isoforms to activate transcription (I and III)

E-proteins function as transcription activators or repressors (Murre, 2005). To study the transcription regulatory activity of alternative TCF4 isoforms we carried out reporter assays in HEK293 cells (paper I) and rat primary cortical and hippocampal neurons (paper III) using firefly luciferase constructs carrying tandem μE5 (CACCTG) E-boxes (Henthorn et al., 1990a). To test whether the functions of TCF4 in neurons are modulated by neuronal activity, we applied high potassium

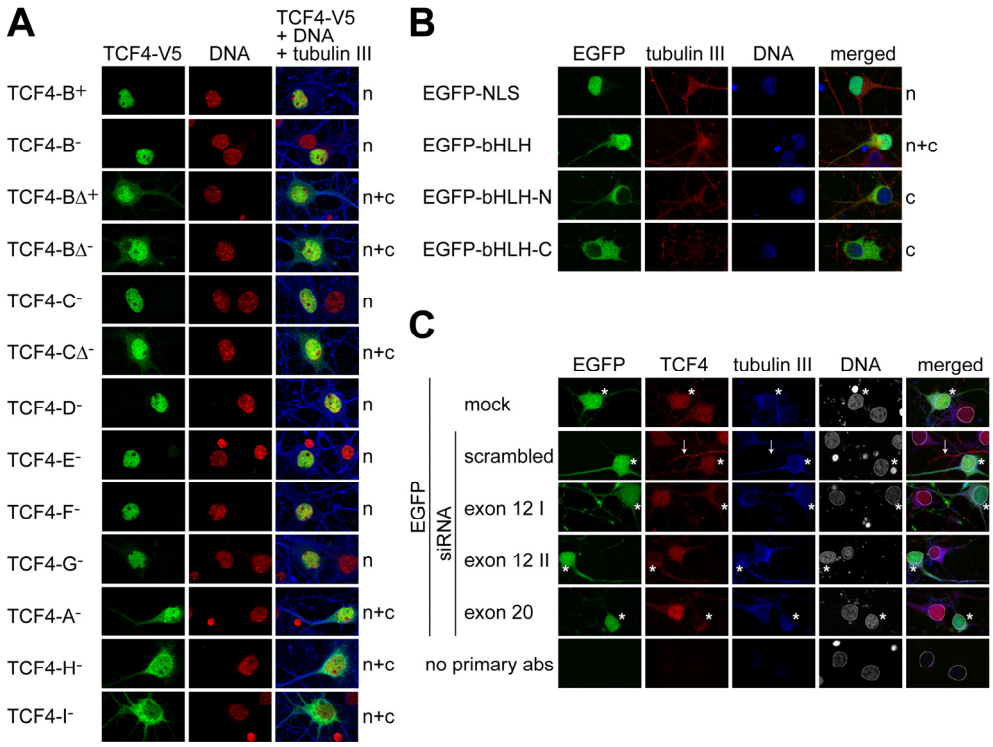


Figure 2. Intracellular distribution of TCF4 proteins in rat primary neurons.
(**A**) Immunocytochemical analysis of V5-tagged TCF4 isoforms. NLS-bearing TCF4 isoforms are nuclear while NLS-lacking isoforms localize to nucleus and cytoplasm.
(**B**) Localization EGFP fusion proteins with TCF4 NLS, bHLH, N-terminal (N) or C-terminal (C) half of TCF4 bHLH domain. EGFP-NLS and EGFP-bHLH localize completely or partially into neuronal nuclei, whereas EGFP-bHLH-N and -C are cytoplasmic. In A and B localization pattern is indicated at the right; n, nuclear; n+c, nuclear and cytoplasmic; c, cytoplasmic. DNA was counterstained with DAPI and neurons were identified by tubulin III staining.
(**C**) Immunocytochemical detection of endogenous TCF4. Neurons were co-transfected with an EGFP-encoding construct and siRNAs. Three different siRNAs specific for *Tcf4* exon 12 or 20 were used; mock and scrambled siRNA transfections were performed in control. Neuronal nuclei are encircled in DNA and merged images; transfected cells are marked with asterisks and arrows indicate glial processes. Neurons were identified by tubulin III staining and DNA was visualized by DRAQ5 staining. Compared to neighbouring untransfected cells, immunostaining of neuronal nuclei with TCF4 antibodies is severely reduced in EGFP-positive cells co-transfected with *Tcf4*-specific siRNAs, but not in mock or scrambled siRNA transfected cells.

treatment that is known to provoke membrane depolarization and Ca^{2+} influx through L-type voltage gated calcium channels in neurons (VGCCs, Morgan and Curran, 1986).

Different TCF4 isoforms were over-expressed at variable levels in HEK293 cells. The levels of isoforms TCF4-BΔ^+, -BΔ^-, -B$^+$, -B$^-$, -CΔ^-, -A$^+$, -A$^-$, -H$^-$ and -I$^-$ were high, those of TCF4-C$^-$ and -D$^-$ medium, those of TCF4-E$^-$ and -F$^-$ low, and the expression of TCF4-G$^-$ was not detected by western blotting. Compared to empty vector transfected cells, the normalized μE5 E-box-dependent luciferase activity was 100-700 times higher in cells expressing TCF4 isoforms TCF4-BΔ^+, -BΔ^-, -B$^+$ or -B$^-$, and 50-70 times higher in cells expressing isoforms TCF4-CΔ^-, -H$^-$ or -I$^-$. TCF4-A$^+$, -A$^-$, -D$^-$, and -F$^-$ activated reporter gene transcription ~20 times, and TCF4-C$^-$ and -E$^-$ ~10 times. In neurons grown in basal conditions we observed only slight differences (less than 3-fold changes) in normalized reporter gene activity when comparing control and TCF4-expressing cells. Neuronal activity increased normalized μE5-dependent luciferase signals 1.7-fold in empty vector transfected cells. In cells expressing TCF4 isoforms, KCl-treatment induced levels ranged from 4 to 44 times above the signal obtained from vector-transfected cells grown in basal conditions. Strongest transactivation was achieved with isoforms TCF4-B$^+$, -B$^-$, -BΔ^+ and -A$^+$, medium with isoforms TCF4-BΔ^-, -C$^-$, -CΔ^-, -D$^-$, A$^-$ and -F$^-$, and weaker with isoforms TCF4-H$^-$, I$^-$ and E$^-$.

It is known that the activity of *CMV* promoter, used for expression of TCF4 isoforms in the above experiments, is upregulated by neuronal activity (Wheeler and Cooper, 2001). Therefore, depolarization induced rise in TCF4-mediated E-box-dependent reporter transcription could be caused by increased TCF4 expression or increased TCF4 activation. To differentiate between these possibilities, we subcloned selected TCF4 isoforms into a depolarization non-inducible *EF1α* promoter containing vector. We performed comparative analysis of *CMV* and *EF1α* promoter-controlled TCF4 proteins by determining the expression levels of C-terminally V5/His-tagged TCF4-A$^-$ by western blotting and by studying the effects of TCF4-B$^-$ and TCF4-A$^-$ on E-box-controlled reporter transcription in untreated and depolarized primary neurons. The levels of TCF4-A$^-$-V5/His were considerably higher when expressed from *EF1α* promoter compared to when expressed from *CMV* promoter. Importantly, *EF1α*-controlled TCF4-A$^-$-V5/His expression did not change in response to KCl-treatment, while in case of *CMV* promoter the protein could be detected only in depolarized neurons. Despite low expression levels, TCF4-mediated E-box-dependent reporter transcription was higher in depolarized neurons expressing *CMV*-TCF4 proteins than in untreated neurons expressing *EF1α*-TCF4 proteins. Additionally, transactivation by *EF1α*-TCF4 proteins was increased by membrane depolarization in spite of stable TCF4 expression. Thus, we concluded that depolarization-induced increase in reporter activity reflects activation of TCF4.

We noticed increased luciferase activity in Δ isoforms expressing HEK293 cells compared to the respective full-length isoform expressing cells that could be caused by higher expression levels of Δ isoforms or the absence of the CE region implicated in transcriptional repression (Herbst and Kolligs, 2008). Additionally,

there was a strong correlation between the presence of full-length AD1 and the isoform's ability to transactivate in HEK293 cells (point biserial correlation coefficient r=0.87, p<0.00011). This relationship was weaker in case of primary neurons (r=0.62, p<0.03). Full-length AD1 includes exon 3 encoded LDFS and LXXLL motifs that are known to be involved in AD1-mediated recruitment of histone acetyltransferases (Massari et al., 1999, Bayly et al., 2004, Zhang et al., 2004). In neither cell type there were significant correlations between the isoform's capacity to activate transcription and the presence of NLS, presence of only the C-terminal part of AD1 or presence of the extra four amino acids in the + isoforms. The latter is in contrast to a study on *FGF-1* promoter in glioblastoma cells that proposed differential roles for TCF4 + and − isoforms (Liu et al., 1998). Altogether, our data show that all TCF4 isoforms analysed are able to activate transcription controlled by μE5 E-boxes both in HEK293 cells and primary neurons, but in neurons the transactivation is dependent on neuronal activity. Regulation at several levels, e.g. posttranslational modification, dimerization specificity and DNA-binding, could be part of the mechanisms by which neuronal activity modulates the capacity of TCF4 to activate transcription. Since all alternative TCF4 isoforms were able to mediate neuronal activity induced transactivation from E-boxes, we reasoned that the effect could be dictated by the functional domains AD2 and/or bHLH that are present in all isoforms.

To study the individual role of TCF4 transcription activation domains AD1 and AD2, we compared the abilities of artificial TCF4-B$^-$ without AD2 (ΔAD2) and native TCF4-A$^-$ to regulate E-box-controlled transcription, and heterologic proteins where AD1 or AD2 is fused with GAL4 DNA binding domain and E2 epitope-tag to activate reporter transcription from a reporter vector that carries GAL4 binding sites (GBS). The compared proteins were over-expressed at similar levels in HEK293 cells. In HEK293 cells comparable degrees of reporter gene activation were achieved with TCF4-B$^-$ΔAD2 and TCF4-A$^-$, whereas in neurons only TCF4-A$^-$ was able to mediate statistically significant induction of reporter gene transcription after membrane depolarization. Both GAL4-AD1-E2 and GAL4-AD2-E2 activated transcription in HEK293 cells, but only GAL4-AD2-E2 raised normalized GBS-dependent luciferase activity above control levels in primary neurons. However, this induction was not activity-dependent since similar upregulation of reporter levels was recorded both in cells grown in basal conditions and cells treated with KCl. In neurons the transactivation potency of AD2 was comparable to that of VP16 when fused to GAL4 DNA-binding domain. In sum these results indicate that (i) in HEK293 cells both activation domains are able to mediate transactivation when joined to either TCF4 bHLH or GAL4 DNA-binding domain; (ii) in cultured primary neurons only AD2 and not AD1 separately functions in activating transcription; (iii) transactivation by AD2 is under the control of neuronal activity only when joined to TCF4 bHLH domain and not GAL4 DNA-binding domain. In contrast to our results, it has been suggested for TCF3/E2A that only GAL4-fused AD1 and not GAL4-fused AD2 is active in the nervous system of zebrafish embryos (Argenton et al., 1996). This discrepancy

could be related to differences in the studied proteins, organisms and/or experimental systems.

We noted that in HEK293 cells when both activation domains were present in a single protein as is the case for TCF4-B isoforms, the activation of transcription exceeded the additive effect of proteins containing AD1 and AD2 separately, suggesting that the two domains may act synergistically. To test this we studied the transactivation capacity of GAL4 fusion proteins that contain both TCF4 activation domains joined by native or heterologic amino acids. In both cases the activation fold exceeded the additive effect of single TCF4 activation domain containing GAL4 fusion proteins, thus confirming the synergistic functioning of the two domains in HEK293 cells. Similar cooperation between AD domains in HEK293 derivative cells has also been described for E12 (Bhalla et al., 2008), suggesting that the phenomenon could be characteristic of mammalian E-proteins in general. One of the underlying mechanisms explaining the synergistic effect could likely be the cooperative interaction of AD1 and AD2 with CBP/P300 histone acetyltransferases (Bayly et al., 2004).

Collectively, we showed that alternative TCF4 isoforms have differential capacity to activate transcription and their functioning is dependent on cell context. In HEK293 cells the two TCF4 transactivation domains function cooperatively and therefore the transcription activation capacity diverges considerably between isoforms encompassing both or only one of the transactivation domains. Additionally, since AD1- and DES-containing isoforms with longer N-termini can interact with ETO family of co-repressors, it is possible that these isoforms may act as repressors in certain conditions, whereas isoforms with shorter N-termini could function as constitutive activators (Guo et al., 2009). In primary neurons the transactivation capacity of TCF4 is dependent on AD2 and controlled by neuronal activity independent of isoformal specificity or expression levels. The latter reveals a novel layer of regulation for TCF4.

5. Regulation of TCF4 transcriptional activity in neurons (III)

To elucidate the mechanisms by which neuronal activity regulates TCF4, we first made constructs coding for VP16 transactivation domain fused TCF4 proteins with serial deletions of N-terminal sequences starting with the shortest TCF4 isoform TCF4-I$^-$. We verified that all TCF4 deletant proteins were expressed at comparable levels in HEK293 cells, were able to bind μE5 E-box *in vitro* and upregulated μE5 E-boxes-controlled transcription in HEK293 cells. However, in rat primary neurons overexpression of the shortest TCF4 deletant protein VP16-TCF4-P498 did not elicit induction of reporter activity in KCl-treated cells compared to untreated cells, whereas all other deletant proteins, including VP16-TCF4-M430, were able to mediate membrane depolarization-provoked induction of reporter activity. These results indicate that the region of TCF4 between amino acids M430 and P498 (according to isoform TCF4-B$^-$) together with the bHLH domain are critical for activity-dependent regulation of TCF4.

To elucidate which signalling pathways play a role in the capacity of TCF4 to activate transcription from μE5 E-boxes-controlled promoter in response to neuronal activity, we first confirmed that Ca^{2+} influx through L-type VGCCs is a prerequisite for the induction using L-type VGCC blocker nifedipine. In neurons rise in intracellular Ca^{2+} activates many protein kinases including Ca^{2+}-calmodulin kinases (CAMK), protein kinase A (PKA), protein kinase C (PKC) and mitogen-activated protein kinases (MAPKs) (Kotaleski and Blackwell). We therefore analysed these pathways for their ability to regulate TCF4-dependent transcription in neurons by coexpressing constitutively active or dominant-negative forms of the kinases and by treating neurons with different pharmacological activators or inhibitors. According to our results TCF4-dependent transcription in neurons was activated by PKA. Particularly, TCF4-dependent transcription was strongly induced by constitutively active PKA catalytic subunit but not constitutively active CAMK2B, CAMK4 or MEK (MAPK/extracellular-signal-regulated kinase kinase) proteins. In addition, treatment with a PKA activator dibutyryl cAMP (dbcAMP) was able to upregulate TCF4-dependent transcription, and both pharmacological inhibition of PKA and expression of a dominant negative form of PKA regulative subunit were effective in significantly reducing neuronal activity induced TCF4-dependent transcription. Roles for other kinases, especially for PKC, constitutively active form of which increased and pharmacological inhibition of which reduced TCF4-dependent transcription, cannot be completely ruled out. Nevertheless, our results strongly suggest that TCF4 activity in neurons depends primarily on signalling via the cAMP-PKA pathway.

We next asked whether TCF4 is a direct substrate for PKA. To evaluate this we performed *in vitro* kinase assays with immunoprecipitated Flag-TCF4-I^{-} deletant proteins. E2-PKA-Cα was not able to phosphorylate the shortest TCF4 deletant Flag-TCF4-P498, whereas phosphorylation of all other deletant proteins, including Flag-TCF4-M430 was detected in the presence of PKA. Using bioinformatic analysis we identified S448 and S464 as potential PKA targets in a region between amino acids M430-P498. Both of these sites are conserved in mammals, reptiles and amphibians and S448 also in teleost fishes. To determine whether S448 and/or S464 are phosphorylated by PKA *in vitro* we substituted one or both of these serines with alanines in the context of Flag-TCF4-I^{-} and performed kinase assays. Compared to wt protein, phosphorylation by PKA was severely reduced in case of S448A mutant, slightly reduced in case of S464A mutant and abolished in case of double mutant. This indicates that S448 is the major and S464 a minor site for phosphorylation by PKA. Altogether these data identify TCF4 as novel direct substrate for PKA.

To analyse the effects of PKA phosphorylation site mutations on the ability of membrane depolarization or E2-PKA-Cα co-expression to induce TCF4 transcriptional activity in primary neurons, we performed reporter assays using constructs encoding wt, S448A, S464A or double mutant TCF4-A^{-}. The fold induction of μE5 E-box-dependent luciferase activity in KCl-treated neurons relative to untreated neurons was significantly reduced in case of S448A and double mutant proteins compared to wt TCF4-A^{-}. Additionally, in contrast to

results with wt TCF4-A$^-$, there was no significant fold induction of E-box-dependent luciferase activity in neurons co-expressing E2-PKA-Cα relative to empty vector transfected cells in case of S448A, S464A or double mutant proteins. Taken together, our results indicate that the S448 phosphorylation site is required for membrane depolarization induced TCF4-dependent transcription, whereas both S448 and S464 are needed for PKA-Cα induced TCF4-dependent transcription.

S448 and S464 phosphorylation sites in TCF4 are conserved in different vertebrate species, but not in other E-proteins. Therefore, we asked whether the activity of other E-proteins is regulated by membrane depolarization in primary neurons. To answer this we compared the transactivation capacity of TCF12/HEB, the E-protein most similar to TCF4, to that of TCF4-B$^-$ in HEK293 cells and primary neurons. TCF12/HEB and TCF4-B$^-$ upregulated μE5 E-box-dependent reporter transcription at comparable levels in HEK293 cells. Conversely, TCF4-B$^-$ was significantly more potent than TCF12/HEB in activating E-box-controlled transcription in response to high potassium treatment in primary neurons. These data point towards differential regulation of E-proteins' activity in neurons and support the importance of the identified PKA phosphorylation sites present in TCF4 and not TCF12/HEB.

To study whether endogenous TCF4 is phosphorylated in a neuronal activity-dependent manner, we analysed the fractionation profiles of endogenous TCF4 in untreated and KCl-treated primary neurons by isoelectric focusing combined with polyacrylamide gel electrophoresis. For comparison we included HEK293 cells left untreated or treated with the PKA inhibitor H89 or the PKA activator dbcAMP in our analysis. We detected four forms of TCF4-A$^-$ (1-4) and three forms of TCF4-A$^+$ (5-7) differing in their isoelectric points in primary neurons and HEK293 cells. Starting from the basic side we numbered the TCF4-A forms 1-7 according to their isoelectric points, which fell in the range of pH 5.0-5.3. The predominant form of TCF4-A in untreated neurons was form 2, whereas in KCl-treated neurons the relative abundance of forms 1, 3 and 4 was notably higher than in untreated neurons. In untreated HEK293 cells the distribution of TCF4-A between different forms was similar to that seen in KCl-treated neurons, and the relative abundance of forms 3 and 4 was decreased by H89 and increased by dbcAMP treatment. The rise in more acidic forms 3 and 4 in KCl-treated neurons compared to untreated neurons suggests that endogenous TCF4-A$^-$ might be phosphorylated in response to membrane depolarization. The change in the relative abundance of forms 3 and 4, induced by PKA inhibition or stimulation in HEK293 cells, implies that endogenous TCF4-A$^-$ could be phosphorylated by PKA.

The cAMP-PKA pathway, which upon synaptic activity in neurons is activated by Ca^{2+}/calmodulin-sensitive adenylyl cyclases, is known to be involved in Ca^{2+}-regulated synaptic function, plasticity and the late phase of long term potentiation that requires gene expression (Poser and Storm, 2001, Nguyen and Woo, 2003). The prototypic activity-stimulated transcription factor CREB1 was first described to be phosphorylated and activated by the cAMP-PKA pathway (Gonzalez and Montminy, 1989). However, in neurons CREB1 activity has been found to be primarily stimulated by the CAMK cascade and the MAPK pathway in response to

synaptic activity-induced Ca^{2+} influx through N-methyl-D-aspartate receptors and L-type VGCCs (Bading et al., 1993, Chawla et al., 1998, Dolmetsch et al., 2001). In addition to CREB1 these two signalling modules have been demonstrated to regulate neuronal gene expression by controlling the activity of several other Ca^{2+}-dependent transcription factors, including methyl CpG binding protein 2 (MECP2) (Chen et al., 2003, Martinowich et al., 2003), serum response factor (Misra et al., 1994), myocyte enhancer factor 2 (MEF2) (Linseman et al., 2003), nuclear factor κB (Meffert et al., 2003) and NEUROD1 (Gaudilliere et al., 2004). cAMP-PKA signalling is considered to play a modulatory role in Ca^{2+}-dependent gene regulation in neurons. For example, the MAPK pathway and the CAMK cascade are positively and negatively regulated by PKA, respectively (Wayman et al., 1997, Impey et al., 1998), localization of CREB-regulated transcriptional coactivator (CRTC1) in the nucleus is dependent on cAMP-PKA activity (Ch'ng et al., 2012) and Ca^{2+}-dependent gene regulation by MEF2D and NFAT (nuclear factor of activated T cells) proteins is modulated by PKA (Chow and Davis, 2000, Sheridan et al., 2002, Wang et al., 2005, Belfield et al., 2006). In contrast, we found here that TCF4 appears to be activated predominantly by PKA in response to Ca^{2+} influx through L-type VGCCs in neurons.

The mechanisms by which phosphorylation of TCF4 regulates its transactivational capacity remain undefined. There are several non-exclusive possibilities. For instance, phosphorylation could stabilize TCF4 protein, similarly to what has been demonstrated in case of phosphorylation of bHLH protein MYOD1 by MEK (Jo et al., 2011), TWIST1 by MAPKs (Hong et al., 2011) and HES1 by c-Jun N-terminal kinase (Lin and Lee, 2012). Additionally, phosphorylation could modulate the dimerization specificity of TCF4 via a mechanism similar to that described for E47, phosphorylation of which at different serine residues located upstream of the bHLH domain inhibits the formation of homodimers and/or supports heterodimerization with MYOD1 (Sloan et al., 1996, Lluis et al., 2005). Moreover, regulated dimerization could be coupled to modulation of DNA-binding of different dimers by Ca^{2+}-loaded calmodulin that has been shown to bind to the basic sequence of E-proteins' bHLH domain and inhibit DNA-binding of E-protein homodimers and thereby selectively promoting DNA-binding of heterodimers (Corneliussen et al., 1994, Saarikettu et al., 2004, Hauser et al., 2008a, Hauser et al., 2008b). In view of the above, it would be of great importance to determine the heterodimerization partners of TCF4 in the nervous system. So far, only NEUROD1 and NEUROD2 have been demonstrated to interact with endogenous TCF4 in brain (Brzozka et al., 2010). Notably, both NEUROD1 and NEUROD2 have been shown to function as Ca^{2+}-regulated transcription factors in regulating cerebellar dendritogenesis or thalamocortical and hippocampal synaptic maturation, respectively (Gaudilliere et al., 2004, Ince-Dunn et al., 2006, Wilke et al., 2012). However, the role of NEUROD proteins in our experiments remains uncertain since the CACCTG E-box used in this study is not a preferred binding site for E-protein-NEUROD heterodimers (De Masi et al., 2011).

Neuronal activity induced transcription is a substantial component in adaptive functions of the brain, such as learning and memory, and hence it is essential for

cognitive development. This is illustrated by neurodevelopmental disorders like autism and Rubinstein-Taybi and Rett syndromes that are related to dysregulation of neuronal activity-dependent gene expression program in the brain (Morrow et al., 2008, West and Greenberg, 2011). Our finding that *TCF4*, variation in which is associated with schizophrenia and possibly bipolar disorder, and mutations in which cause Pitt-Hopkins syndrome, codes for a Ca^{2+}-dependent transcriptional activator further links neuronal activity-responsive gene regulation to cognitive development. We suggest that impaired cognitive functions seen in PTHS, and possibly in psychosis patients, could in part be caused by deranged TCF4-dependent transcriptional response to neuronal activity.

6. Mapping of Pitt-Hopkins syndrome causing deletions and truncating mutations in *TCF4* (II and unpublished)

To date more than 130 PTHS patients have been described with large deletions encompassing all or part of *TCF4* gene, or with splice-site, indel, nonsense and missense mutations in *TCF4* (Andrieux et al., 2008, Giurgea et al., 2008, Zweier et al., 2008, de Pontual et al., 2009, Rosenfeld et al., 2009, Taddeucci et al., 2010, Takano et al., 2010, Hasi et al., 2011, Lehalle et al., 2011, Marangi et al., 2011, Takano et al., 2011, Armani et al., 2012, Marangi et al., 2012, Need et al., 2012, Takenouchi et al., 2012, Talkowski et al., 2012, Verhulst et al., 2012, Whalen et al., 2012, Willemsen et al., 2012, Ghosh et al., in press, Maini et al., in press, Steinbusch et al., in press). To specify the impact of the PTHS-associated mutations, we mapped all the described mutations to the composite structure of *TCF4* gene. Besides the mutations analysed in publication II, 16 additional genetically confirmed cases have been described in the literature for the moment. Among the latter, four patients carry large deletions that include the whole *TCF4* gene, and 12 have smaller deletions or point mutations in *TCF4* (Table 4).

Out of sixteen PTHS-associated translocations and large intragenic deletions, 3 lead to loss of all 5′ exons. Three of the intragenic deletions are anticipated to allow the synthesis of functional *TCF4* transcripts from one or four downstream 5′ exons. Six of the PTHS-associated intragenic deletions encompass at least one of the constitutive exons 10–21 and since these deletions are out-of-phase and/or lead to the loss of the AD2 or bHLH domain, no functional *TCF4* transcripts can be produced from these alleles. It is uncertain whether and how many utilizable transcription initiation sites are present in the two patients with deletions extending from exons 7a-7 to exons 9-10c (Whalen et al., 2012). In a patient with deletion extending from exons 8d-10a to exon 10c (Talkowski et al., 2012), it remains possible that functional mRNAs encoding TCF4 Δ-isoforms could be produced using 5′ exons upstream of the deletion. The translocation in a patient with milder clinical phenotype than classical PTHS (Kalscheuer et al., 2008) encompasses 9 upper 5′ exons, but fully functional transcripts could be transcribed from the remaining 12 initial exons.

Table 4. Novel PTHS-associated mutations in *TCF4*.
The coordinates are given according to human genome assembly hg19. The mutation denoted with '#' is from a patient with milder phenotype than classical PTHS. Recurrent mutations are indicated with '†'.

Mutation	exons involved	Reference
Intragenic deletions and translocations		
t(3;18):(q13.32;q21.2)(Chr18: 52,909,158)	exon.1a_16	(Talkowski et al., 2012) ×2
Chr18:g.(_53,192,528)_(52,936,991_)del	exon.(_5b)_(12_)del	(Talkowski et al., 2012)
Chr18:g.(_53,192,528)_(53,024,848_)del	exon.(_5b_8c_)del	(Talkowski et al., 2012)
Chr18:g.(53,028,773_53,014,539)_ (52,961,841_52,946,888)del	exon.(8d_10a)_10c del	(Talkowski et al., 2012)
Splice-site mutations		
c.655G>A p.D219?fs*37	exon 10 3′	(Need et al., 2012)
c.1487-2A>G (p.G496Dfs*12)	exon 18 5′	(Marangi et al., 2012)
Indels and nonsense mutations		
c. 469C>T p. R157*	exon 8	(Hamdan et al., in press) #†
c.1269dupA p.I421Nfs*7	exon 16	(Marangi et al., 2012)
Missense mutations		
c. 1732G>T p.R578C	exon 19	(Marangi et al., 2012)
c. 1739G>A p.R580Q	exon 19	(Willemsen et al., 2012) †
c. 1753A>G p.N585D	exon 19	(Marangi et al., 2012)

All eleven reported PTHS-associated splice-site mutations affect the constitutive exons of *TCF4* and are predicted to shift the reading frame as a result of out-of-phase exon skipping or novel splice-site selection. Forty-three of the 49 nonsense and small indel mutations described in 55 patients map to the constitutive exons 10–19 and lead to the generation of a premature stop codon. The above mutations are predicted to be subjected to the nonsense-mediated decay pathway. Additionally, most of the transcripts containing these mutations would encode a protein without the bHLH domain and only 2 nonsense mutations would allow translation of a protein with the majority of the bHLH present. There are four premature stop-causing mutations reported in upper internal exons 8–9 that should have no effect on the Δ-isoforms encoding transcripts, which have exons 8–9 spliced out, and the shorter *TCF4* transcripts initiated at 5′ exons 10a, 10b and 10c. In addition, two PTHS-associated indels in exon 20 do not cause a premature stop, but instead extend TCF4 reading frame into the 3′ UTR exon 21.

Collectively the above data reveals that although many of the deletions and premature stop-inducing mutations probably lead to complete loss-of-function of the affected allele, there are some that are expected to permit the production of the Δ-isoforms of TCF4 and/or the shorter isoform(s) TCF4-D, TCF4-A, TCF4-I and/or TCF4-H. This indicates that even partial loss-of-function of the mutated *TCF4* allele is sufficient to cause PTHS. There are several non-exclusive explanations for the underlying mechanisms. First, it is possible that all *TCF4* alternative promoters are needed to produce sufficient quantity of the protein to maintain normal function. Second, the deficit of AD1 and/or NLS containing TCF4 isoforms might not be compensated by the remaining isoforms that lack these

functional regions. Third, the cell-type- and age-dependent regulation of *TCF4* alternative promoters could also play a role in the functional divergence of different isoforms. Given that a patient with a translocation that leaves 12 of the 21 5′ exons in place has a milder phenotype (Kalscheuer et al., 2008), we tried to analyse whether the PTHS patients with mutations that result in partial loss of TCF4 isoforms have less severe symptoms. Unfortunately, such correlations are difficult to make since systematic graded characterization of the patients is not available. However, in a study where a PTHS clinical diagnosis score was introduced, the patient with hypomorphic deletion had the lowest relative score (Whalen et al., 2012). Additionally, a patient with moderate intellectual disability carries a hypomorphic nonsense mutation (Hamdan et al., in press) located in *TCF4* exon 8 that is spliced out in the Δ-isoforms encoding transcripts.

7. Impairment of TCF4 functions by PTHS-associated missense and reading-frame extending mutations (II)

Twenty-five PTHS patients with missense mutations giving rise to 14 different substitutions at 10 amino acid positions in the constitutive part of TCF4 have been described. The majority of the missense mutations are located in the bHLH domain and only two (p.G358V and p.D535G) are situated outside of the bHLH domain. G358 is situated in the AD2 domain and D535 in the Rep region.

To examine the location of the missense mutation sites in the bHLH domain 3D structure, we used the published co-crystal structures of DNA-bound E47 bHLH homodimer and E47-NEUROD1 bHLH heterodimer (Ellenberger et al., 1994, Longo et al., 2008) as templates to build the structure models of DNA-bound TCF4 bHLH homo- and heterodimer. Examination of the specific functions of the PTHS-affected residues in the structure models revealed that out of the four arginines R569, R576, R578 and R580 in the basic DNA-binding region, the three latter make direct contacts with DNA bases or backbone. The residue respective to R569, which does not bind DNA in our models, has been found to make contacts with E-box flanking DNA in related bHLH co-structures (De Masi et al., 2011), suggesting that this arginine may also contact DNA. According to our models PTHS-mutated arginine R582 and asparagine N585 in helix 1 of the bHLH domain form salt bridge(s) within the first helix or hydrogen bonds with DNA backbone, respectively. The PTHS-affected alanines A587 in the first helix and A614 in the second helix form packing interactions at the dimer interface in the TCF4 structure models. Based on the location, we hypothesized that PTHS-associated missense mutations could have an effect on the protein's DNA-binding, dimerization and/or transactivation ability.

In order to elucidate the underlying molecular mechanisms of the PTHS-associated missense mutations, we introduced seven different missense mutations to the TCF4-B$^-$ expression construct. One substitution was made for each TCF4 position that had been shown to be mutated in PTHS at the time of the design of publication II – these were G358V, D535G, R576Q, R578H, R580W, R582P and A614V. In addition, we introduced the reading frame extending S653Lfs*57

frameshift mutation to a construct containing the TCF4-B^{+} protein coding and 3′ UTR sequences. To differentiate between loss-of-function and gain-of-function effects of the frameshift, we also generated the S653* mutant.

We carried out functional analyses of the PTHS-associated missense and reading frame extending mutations to study the effect of the mutations on TCF4 protein expression, subcellular localization, dimerization, DNA-binding and transactivation ability using a range of techniques. We applied western blotting to assess the levels of PTHS-associated mutant TCF4 proteins translated *in vitro* or over-expressed in HEK293 cells. We measured mRNA levels of *TCF4* over-expressed in HEK293 cells by reverse transcription and quantitative PCR. We studied the subcellular distribution of mutant TCF4-B proteins over-expressed in HEK293 cells or rat primary cortical and hippocampal neurons by indirect immunofluorescence analysis. To examine the dimerization ability of the mutants, we performed glutaraldehyde crosslinking experiments with *in vitro*-translated TCF4 proteins and nuclear redirection assays with over-expressed TCF4 proteins in HEK293 cells and primary neurons, taking advantage of the knowledge that TCF4 proteins lacking the NLS (e. g. EGFP-bHLH or TCF4-A^{-}) can be transported to the cell nucleus through heterodimerization with NLS-containing bHLH partner proteins. We included ASCL1 and NEUROD2 in the latter analysis since they represent two separate phylogenetically defined families of the bHLH transcription factors and have been demonstrated to interact with TCF4 (Atchley and Fitch, 1997, Persson et al., 2000, Simionato et al., 2007, Brzozka et al., 2010). Additionally, these factors might be of relevance to PTHS because of parallels between the phenotypes of the disorder and *Ascl1* or *Neurod2* null mice (Hirsch et al., 1998, Olson et al., 2001).

To examine the effect of the PTHS-associated mutations on the ability of TCF4 to bind DNA, we used *in vitro*-translated proteins and determined their binding to E-box containing oligonucleotides by electrophoretic mobility shift assay (EMSA). We studied the μE5 E-box (CACCTG) binding ability of TCF4-B homodimers, intra-TCF4 heterodimers (that contain one mutant TCF4-B subunit and one wt TCF4 bHLH subunit fused with VP16 transactivation domain) and TCF4:ASCL1 heterodimers. We examined the effect of PTHS-associated mutations on the DNA-binding ability of TCF4:NEUROD2 heterodimers at the CATCTG E-box from the insulinoma-associated antigen-1 (IA-1) promoter (Breslin et al., 2003). In our assays, wt TCF4 homodimers were able to bind μE5 and IA-1 E-boxes, whereas ASCL1 and NEUROD2 bound DNA only in heterodimeric context. In addition to *in vitro* DNA-binding ability, we monitored the capacity of the PTHS-associated TCF4 mutant proteins to interact with chromatin in HEK293 cells by differential detergent fractionation. The proportions of the studied proteins in the soluble fraction including cytoplasm and nucleoplasm, and the chromatin containing pellet fraction were determined by western blotting and densitometric quantification.

We investigated the impact of the PTHS-associated mutations on the ability of TCF4-B to activate μE5 E-box-dependent reporter transcription when over-expressed alone or together with ASCL1 in HEK293 cells or rat primary neurons grown in basal conditions or treated with high potassium. For some mutants we

performed additional experiments in the context of shorter TCF4-A^{-} isoform and GAL4-AD2 heterologic protein. For each wt and PTHS-associated mutant TCF4-B protein we calculated the cooperation index that expresses how many times the transactivation fold is increased in cells co-expressing TCF4-B and ASCL1 compared with the sum of transactivation folds from cells expressing both proteins separately. The index value of 1.0 is indicative of simple summation of the independent effects of both proteins, whereas a value above 1.0 implies synergism and below 1.0 antagonism. The cooperation index was almost 14 for wt TCF4-B^{-} and almost 6 for wt TCF4-B^{+}, indicating that wt TCF4 acts cooperatively with ASCL1.The effects of each studied PTHS-associated mutation on TCF4 functions are summarised in Table 5 and further discussed below.

Table 5. Effects of PTHS-associated mutations on TCF4 functions.
[a] levels of over-expressed protein in HEK293 cells;
[b] levels of over-expressed mRNA in HEK293 cells;
[c] levels of *in vitro* translated protein;
[d] nuclear localization in cells;
[e] homodimerization *in vitro*;
[f] heterodimerization with ASCL1 and NEUROD2 in cells;
[g] TCF4 homodimer and intra-TCF4 heterodimer DNA binding *in vitro*;
[h] TCF4 and ASCL1/NEUROD2 heterodimer DNA binding *in vitro*;
[i] co-fractionation with cellular chromatin;
[j] activation of E-box-dependent transcription by over-expressed TCF4 in HEK293 cells/primary neurons;
[k] activation of E-box-dependent transcription by over-expressed TCF4 and ASCL1 in HEK293 cells.
++ denotes similar, +++ increased, + reduced, (+) severly reduced, and - absent activity of a mutant protein compared to the activity wild-type TCF4. nd, not determined.

Mutation	a	b	c	d	e	f	g	h	i	j	k
None (wt)	++	++	++	++	++	++	++	++	++	++	++
N-terminal to bHLH											
G358V	++	nd	++	++	nd	++	++	++	++	++	++
D535G	++	nd	++	++	nd	++	+++	+/++	++	+++	++
Arginines in bHLH											
R576Q	++	nd	++	++	nd	++	-	(+)	+	-	+
R578H	++	nd	++	++	nd	++	-	-/(+)	+	-	-
R580W	++	nd	++	++	nd	++	-	-	+	-	-
R582P	++	nd	++	++	+	++	-	-	+	-	-
Dimerization interface											
A610V	+	++	++	++	(+)	++	(+)	++	+	(+)	++
C-terminal to bHLH											
S653Lfs*57	-	+++	++	++	nd	++	-	-	++	-	(+)
S653*	++	nd	++	++	nd	++	++	+/++	++	++/+++	+

We found that similarly to wt TCF4-B, all mutant TCF4-B proteins localized to the cell nucleus in HEK293 cells and rat primary neurons as detected by indirect fluorescence analysis. This result was expected since none of the PTHS-associated mutations affected the NLS of TCF4-B. Compared to wt protein, differences in subnuclear distribution were observed by microscopy only for S653Lfs*57 mutant that formed aggregates. However, our differential detergent fractionation experiments demonstrated that the subnuclear distribution of some other PTHS-associated mutants was also altered. Namely, R576Q, R578H, R580W, R582P and A614V mutants, were shifted from chromatin containing fraction to soluble nucleoplasm fraction. These results are in accordance with a recent study that observed altered subnuclear localization of EGFP-fused TCF4-B$^+$ proteins carrying mutations at R578, R580 or A614 (Forrest et al., 2012).

Using nuclear redirection assay we demonstrated that all studied PTHS-associated mutants were able to heterodimerize with ASCL1 and NEUROD2 in cells. Additionally, glutaraldehyde cross-linking of *in vitro* translated proteins revealed reduced homodimerization of A614V mutant compared to wt TCF4-B$^-$. Forrest et al. 2012 found by quantitative fluorescence resonance energy transfer (FRET) methodology that in cell lysates both homodimerization and heterodimerization ability of TCF4-B$^+$ is reduced by mutations at R578, R580 and A614. These results are not completely contradictory. First, our analysis did not exclude the possibility of quantitative effects. Second, dimerization properties of TCF4 could be altered in cell lysates as compared to intact cells. For example, homodimerization of TCF4 can be detected by FRET analysis in cell lysates (Forrest et al., 2012) and glutaraldehyde cross-linking of *in vitro* translated proteins (II), but not by nuclear redirection assay in intact cells (I).

7.1. Effects of PTHS-associated mutations N-terminal to the bHLH domain of TCF4

We studied two PTHS-associated missense mutations, G358V and D535G, located upstream of the bHLH domain. No major deficiencies were revealed for G358V or D535G mutant TCF4-B$^-$ proteins that were able to dimerize, bind DNA and activate E-box-dependent transcription in HEK293 cells and rat primary neurons. We additionally analysed G358V and D535G mutations in the context of shorter TCF4-A$^-$ isoform, and G358V mutation in the context of a protein where TCF4 AD2 is fused to heterologic GAL4 DNA-binding domain, but did not observe any significant impairment in the ability of the G358V and D535G mutants to activate transcription from μE5 E-boxes or GAL4-binding sites containing promoter, respectively, in HEK293 cells or rat primary neurons. Conversely, D535G mutant TCF4-B$^-$ displayed slightly increased ability to activate transcription compared to wt protein when over-expressed alone in HEK293 cells or primary neurons. Similarly, in a recent study the D535G mutation was found to enhance transactivation by TCF4-B$^+$ (Forrest et al., 2012).

Nevertheless, D535G mutant displayed shifted affinity of homo- versus heterodimerization *in vitro*. Namely, in co-translation and EMSA experiments, we

detected wt TCF4 predominantly in DNA-bound heterodimeric complexes with ASCL1, whereas in case of D535G mutant the prevalence of heterodimers was lost or even surpassed by TCF4 homodimers when translated together with ASCL1. The change in homodimer–heterodimer equilibrium was not evident in reporter assays, but it is possible that slight differences in dimerization preferences that are less influential in experiments with over-expressed proteins could be of importance in an endogenous context. Several previous studies have demonstrated that regions outside the HLH domain affect dimerization of E-proteins. For instance, amino acids C-terminal to the bHLH domain are required for the dimerization of E-proteins *in vivo* (Goldfarb et al., 1998), acidic sequences preceding the bHLH domain impact the dimerization specificity of E12 (Shirakata and Paterson, 1995), insertion upstream of the bHLH domain interferes with dimerization and DNA-binding of TCF12/HEB (Klein et al., 1993) and phosphorylation of a residue outside of the bHLH domain regulates dimerization specificity of E47 (Sloan et al., 1996, Lluis et al., 2005).

Our finding that D535G mutant has only a minor deficiency and G358V mutant functions similarly to wt TCF4 raises questions of whether and how these mutations are responsible for the PTHS phenotype in the patients. Further studies are needed to uncover the pathogenic mechanisms of these mutations and to verify that the patients do not carry any additional mutations that may contribute to the disorder.

7.2. Effects of PTHS-associated mutations at arginines in the bHLH domain of TCF4

We demonstrated that R576Q, R578H and R580W and R582P mutations abrogate *in vitro* DNA-binding of TCF4 homodimers regardless of whether the mutation is present in one or both subunits. In accordance with this, the mutant proteins were not able to associate effectively with cellular chromatin or mediate activation of E-box-dependent transcription in HEK293 cells or primary neurons. Therefore, loss of even a single DNA contact made by any of the basic region arginines R576, R578 and R580 is detrimental for DNA binding of TCF4. Consistently, previous studies have demonstrated that mutating the basic region arginines in bHLH-PAS protein ARNTL/BMAL1 (ARNT-like/brain and muscle ARNT-like 1) or E-proteins E12 or E47 has no effect on dimerization but impairs DNA binding (Voronova and Baltimore, 1990, Hosoda et al., 2004, Beltran et al., 2005). For R582P mutation located in the beginning of helix 1 of TCF4 HLH domain, we propose that the α-helix destabilizing proline substitution (Pace and Scholtz, 1998) interferes with the conformational change of the basic region from a random coil to a structured α-helix that has been suggested to be induced in bHLH factors upon DNA-binding in solution (Fairman et al., 1997).

It has been suggested that two intact basic regions are required for the interaction of bHLH dimers with DNA (Voronova and Baltimore, 1990). However, according to our results, not all mutations in the TCF4 basic region arginines are *trans*-dominant in the context of heterodimers. Specifically, although R578H,

R580W and R582P mutants impaired DNA-binding and transactivation ability of TCF4 heterodimers with ASCL1 or NEUROD2, the DNA-binding and transactivation ability of R576Q mutant was partially rescued in heterodimeric complex with ASCL1. The index of cooperation with ASCL1 (i) was increased above wt levels by R576Q mutation, highlighting synergism with ASCL1; (ii) did not differ from 1.0 in case of R578H, R580W mutants, meaning that these mutants did not act synergistically with ASCL1; and (iii) was below 1.0 in case of R582P mutant, revealing an antagonistic dominant-negative effect. Our data are consistent with reporter assays performed with R580 mutations encompassing TCF4-ASCL1 heterodimers in previous studies (Zweier et al., 2007, de Pontual et al., 2009), but differ from the results obtained with R576G substitution that did not allow TCF4 to cooperate with ASCL1 in activating transcription from the *Delta1* promoter in SKNBE(2)C cells (de Pontual et al., 2009). The latter could reflect substitution-, cell-line- or promoter-specific effects. Notably, across all bHLH families, the position corresponding to R576 in TCF4 is less conserved than other arginines in the basic region, and substitutions corresponding to PTHS-associated mutations R576G or R576Q in TCF4 are present in some natural bHLH proteins (De Masi et al., 2011). In sum, as the PTHS-associated mutations in the bHLH domain arginines impair the DNA-binding and transactivation ability of TCF4 to a varying degree, these mutations should not be regarded as functional equivalents.

7.3. Effects of PTHS-associated mutation at dimerization interface in the bHLH domain of TCF4

We observed that A614V mutation, located in the middle of helix 2 of the HLH domain, leads to reduced steady-state levels of TCF4-B$^-$ protein, whereas mRNA expression levels of transcripts coding for A614V mutant were not decreased compared with wt transcripts in transfected HEK293 cells. Treatment with the proteasome inhibitor MG132 partially rescued the expression levels of TCF4-B$^-$ A614V in HEK293 cells and there were no differences in the translatability of A614V mutant compared to wt TCF4-B$^-$ *in vitro*. Together these results indicate that A614V mutation destabilizes TCF4 in cells by inducing proteasomal degradation.

We found that A614V mutation has differential effects on TCF4 homo- and heterodimers. Specifically, the mutation reduced *in vitro* homodimerization and DNA-binding of homodimers in a *trans*-dominant manner, whereas it had no effect on heterodimerization, DNA-binding of TCF4-ASCL1 and TCF4-NEUROD2 heterodimers or activation of μE5 E-box-dependent transcription by TCF4-ASCL1 dimers in cells. This differential effect was reflected also in the increased cooperation index of A614V mutant compared to that of wt TCF4-B$^-$ protein. Importantly, proteasome inhibition did not rescue the impaired transactivation capacity of A614V mutant expressed individually. These data suggest that deficiencies other than reduced expression levels are sufficient to damage the ability of A614V mutant to activate transcription.

Our results are in agreement with a previous study showing that mutating the alanine, corresponding to A614 in TCF4, to aspartate in E47 does not abrogate the formation of homodimers, but leads to weakened DNA-binding ability of homodimers (Voronova and Baltimore, 1990). The alanine at the position corresponding to A614 in TCF4 is found in proteins from many bHLH families, except bHLH-PAS and ID factors (Chavali et al., 2001), indicating that the packing interactions formed by this residue at the dimer interface could be influential in majority of bHLH homo- and heterodimers. However, it is plausible that the extended helix 1 of E-proteins places more constraints on this position in homodimers, whereas the substitution is tolerated in heterodimers because of shorter helix 1 of the partner protein. All this suggests that, in E-proteins, the alanine corresponding to A614 in TCF4 is not essential for heterodimer formation, but possibly contributes to the dynamic properties of HLH homodimers that in turn influences the efficiency of conformational transition of the basic region required for DNA-binding.

7.4. Effects of PTHS-associated frame-shift mutation C-terminal to the bHLH domain of TCF4

We demonstrated that S653Lfs*57 extension triggeres destabilization and aggregation of TCF4 in cells. First, expression of TCF4-B^+ S653Lfs*57 mutant became evident only after MG132 treatment in transfeted HEK293 cells. Second, in MG132 treated cells the mutant protein accumulated into nuclear dots. Third, S653Lfs*57 transcript levels were increased compared to wt TCF4-B^+ mRNA levels in transfected HEK293 cells. Fourth, the quantities of *in vitro* translated protein were similar in case of wt and S653Lfs*57 mutant TCF4-B^+ proteins.

Although proteasome inhibition partially alleviated the instability of S653Lfs*57 mutant protein, it did not rescue the impaired ability of the mutant to activate transcription in reporter assays. This was expected, since S653Lfs*57 mutant had severely reduced ability to bind DNA in homo- and heterodimeric complexes *in vitro*. Though the mutant protein co-fractionated with cellular chromatin, this probably resulted from aggregation and not its ability to bind DNA in cells. Gain of 56 amino acids and not loss of 19 native C-terminal amino acids was responsible for the impaired stability and DNA-binding ability of S653Lfs*57 mutant protein since these deficiencies were not observed for TCF4 with nonsense mutation S653*, which was analysed as a control. However, both S653Lfs*57, S653* mutations in TCF4 shifted the affinity of homo- versus heterodimerization with ASCL1 *in vitro*, suggesting that the native C-terminal amino acids might play a role in this balance. This change in homodimer–heterodimer equilibrium was reflected also in the reduced ability of S653Lfs*57 mutant to cooperate with ASCL1 in activating transcription.

Notably, the same C-terminal amino acids are introduced by another PTHS-associated mutation A634Dfs*74, which was described during the preparation of publication II (Steinbusch et al., in press). We suggest that the pathogenic mechanisms of both of the C-terminal reading-frame extending mutations might be similar.

In sum, our results demonstrate that impacts of PTHS-causing mutations vary in severity from partial loss-of-function to dominant-negative effects and functionally unequal mutations are found among deletions and truncating mutations as well as missense and reading-frame extending mutations in *TCF4*. Given that there is considerable variability of symptoms among the PTHS patients and many of the features considered characteristic of the disorder are not uniformly present in all patients, it would be of importance to investigate whether there are associations between the effects of mutations on TCF4 function and phenotype severity in patients. The only phenotype–genotype correlation suggested so far is that individuals with missense mutations are more likely to develop seizures than other PTHS patients (Rosenfeld et al., 2009), but this was disproved in a recent report (Whalen et al., 2012). Our study provides a starting point for making PTHS phenotype–genotype correlations that would not be based on mutation type solely, but would take into account the functional impacts of the mutations.

8. Closing remarks

Several lines of evidence from human patients and mouse models indicate that precise regulation of *TCF4* expression levels is crucial for correct brain development and function, and deviations in expression result in cognitive disturbances. This substantiates the high number of 5′ exons in the complexly structured *TCF4* gene, characterized in the current study.

The results presented here indicate that hypomorhic mutations in *TCF4* are sufficient to cause severe developmental delay in PTHS patients. However, given that a translocation in *TCF4*, which leaves more than half of the 5′ exons in place, leads to mild mental retardation (Kalscheuer et al., 2008), it is possible that a continuum of phenotypes from moderate mental retardation to full-scale PTHS might be caused by different mutations in the *TCF4* gene. This would mean that, in addition to PTHS, *TCF4* should be considered a candidate gene in cases of milder mental retardation that could be caused by mutations in individual TCF4 5′ exons or in regulatory elements of the alternative promoters. Moreover, since *TCF4* duplication has been described in a patient with developmental delay (Talkowski et al., 2012), it is possible that increased *TCF4* expression is also a risk factor for disturbed cognitive development.

Moderate *Tcf4* overexpression in postnatal forebrain has been shown to result in decreased sensorimotor gating, deficient fear-related learning and memory, and alterations in neuronal activation (Brzozka et al., 2010, Brzozka and Rossner, 2013), suggesting that elevated *Tcf4* levels may be causal for established endophenotypes of SCZ in mice. Nevertheless, it has not been conclusively

ascertained whether and how SCZ-associated variation in *TCF4* is linked to *TCF4* expression in the human nervous system. Evidence for both decreased *TCF4* expression in blood (Kurian et al., 2011, Navarrete et al., in press), and increased expression in cerebellum, blood and induced pluripotent stem cells-derived neurons (Mudge et al., 2008, Brennand et al., 2011, Wirgenes et al., 2012) of psychosis patients has been published. It is noteworthy that according to the *TCF4* gene structure described here, the SCZ-associated intronic SNPs rs17512836 or rs9960767 and rs2958182 (Stefansson et al., 2009, Li et al., 2010, Ripke et al., 2011, Steinberg et al., 2011) are located upstream of *TCF4* 5′ exons 5b or 5c, respectively, indicating that they could play a role in the regulation of transcription from *TCF4* alternative promoters. In addition, the CTG repeat implicated in BPAD (Del-Favero et al., 2002), is situated immediately in front of *TCF4* 5′ exon 4a, suggesting that the repeat length could have a direct impact on transcription from this exon. Besides alterations in overall *TCF4* expression, it is possible that the the level of individual TCF4 isoforms or the changed ratio of different isoforms could be a factor in the psychoses. The latter is supported by the findings of the present study showing that alternative TCF4 isoforms have functional differences in their nuclear import and transactivational capacity. Notably, a partial duplication of the *TCF4* gene, which is predicted to increase the levels of TCF4 isoforms with shorter N-termini (TCF4-D, -A, -H and –I), has been identified in a patient with major depressive disorder (Ye et al., 2012).

The current study identified TCF4 as a Ca^{2+}-dependent transcriptional activator in neurons, providing novel insights into the mechanisms that underlie the association of *TCF4* with disorders affecting cognitive function and development. Several biological processes that modulate neuronal connectivity, e.g. dendritic development, synaptic elimination, regulation of excitatory-inhibitory balance and synaptic plasticity, are regulated by neuronal activity (Greer and Greenberg, 2008). Possible role of TCF4-dependent transcription in any of these processes awaits vigorous proof from *in vivo* functional studies.

CONCLUSIONS

The present thesis examined several aspects of TCF4 functions in health and disease. The principal findings of this study are summarized below:

- Human *TCF4* gene contains 21 5′ exons that potentially enable production of TCF4 isoforms with 18 different N-termini. The diversity of *TCF4* transcripts is further increased by alternative splicing at several internal exons.
- Human *TCF4* mRNAs are expressed ubiquitously but not equally in different tissues. *TCF4* expression is particularly high in the nervous system where the transcripts and protein are detected in neurons.
- Multiple TCF4 protein isoforms with different molecular mass are present in human tissues.
- Intracellular distribution of TCF4 isoforms is differentially regulated. TCF4 isoforms with longer N-termini contain a bipartite NLS and are localized to the nucleus. TCF4 isoforms with shorter N-termini and Δ isoforms are transported to the nucleus via heterodimerization with NLS-bearing partner proteins.
- The ability of different TCF4 isoforms to activate E-box-controlled transcription is varied.
- TCF4 transactivation domains AD1 and AD2 function synergistically in HEK293 cells, whereas only AD2 is able to activate transcription in rat primary neurons.
- TCF4-dependent transcription in primary neurons requires induction of calcium influx through L-type voltage gated channels and signalling via protein kinase A (PKA).
- PKA directly phosphorylates TCF4 at serines 448 and 464 *in vitro*. Endogenous TCF4 is posttranslationally modified in response to membrane depolarization in primary neurons and by modulating cAMP-PKA pathway in HEK293 cells.
- C-terminal part of TCF4, containing the bHLH domain and the PKA-phosphorylated serines, is necessary for regulation of TCF4 by neuronal activity.
- Hypomorphic mutations in *TCF4* are sufficient to cause Pitt-Hopkins syndrome since not all PTHS-associated deletions and truncating mutations in *TCF4* result in complete loss-of-function of the affected allele.
- PTHS-associated missense mutations in TCF4 bHLH domain and the reading frame extending mutation impair DNA-binding and transactivation ability to a varying extent, ranging from hypomorphic to dominant-negative effects, and in a manner dependent on dimer context.
- PTHS-associated reading frame extending mutation and missense mutation at the dimer interface of the HLH domain destabilize TCF4 protein.
- For PTHS-associated missense mutations outside of the bHLH domain no major deficiencies in TCF4 functions were detected.

REFERENCES

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.

Andrieux, J., Lepretre, F., Cuisset, J. M., Goldenberg, A., Delobel, B., Manouvrier-Hanu, S. and Holder-Espinasse, M., 2008. Deletion 18q21.2q21.32 involving TCF4 in a boy diagnosed by CGH-array. Eur J Med Genet. 51, 172-177.

Apone, S. and Hauschka, S. D., 1995. Muscle gene E-box control elements. Evidence for quantitatively different transcriptional activities and the binding of distinct regulatory factors. J Biol Chem. 270, 21420-21427.

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.

Argenton, F., Arava, Y., Aronheim, A. and Walker, M. D., 1996. An activation domain of the helix-loop-helix transcription factor E2A shows cell type preference in vivo in microinjected zebra fish embryos. Mol Cell Biol. 16, 1714-1721.

Armani, R., Archer, H., Clarke, A., Vasudevan, P., Zweier, C., Ho, G., Williamson, S., Cloosterman, D., Yang, N. and Christodoulou, J., 2012. Transcription Factor 4 and Myocyte Enhancer Factor 2C mutations are not common causes of Rett syndrome. Am J Med Genet A. 158A, 713-719.

Aronheim, A., Shiran, R., Rosen, A. and Walker, M. D., 1993. The E2A gene product contains two separable and functionally distinct transcription activation domains. Proc Natl Acad Sci U S A. 90, 8063-8067.

Atchley, W. R. and Fitch, W. M., 1997. A natural classification of the basic helix-loop-helix class of transcription factors. Proc Natl Acad Sci U S A. 94, 5172-5176.

Ayalew, M., Le-Niculescu, H., Levey, D. F., Jain, N., Changala, B., Patel, S. D., Winiger, E., Breier, A., Shekhar, A., Amdur, R., Koller, D., Nurnberger, J. I., Corvin, A., Geyer, M., Tsuang, M. T., Salomon, D., Schork, N. J., Fanous, A. H., O'Donovan, M. C. and Niculescu, A. B., 2012. Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. Mol Psychiatry. 17, 887-905.

Ayton, A., Morris, A. G., Tyson, P. J., Hunt, D., Mortimer, A. M. and Cottrell, D., 2002. Early development and unstable genes in schizophrenia: preliminary results. Eur Psychiatry. 17, 332-338.

Bading, H., Ginty, D. D. and Greenberg, M. E., 1993. Regulation of gene expression in hippocampal neurons by distinct calcium signaling pathways. Science. 260, 181-186.

Bai, X., Kim, J., Yang, Z., Jurynec, M. J., Akie, T. E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H., DiBiase, A., Zhou, Y., Grunwald, D. J., Lin, S., Cantor, A. B.,

Orkin, S. H. and Zon, L. I., 2010. TIF1gamma controls erythroid cell fate by regulating transcription elongation. Cell. 142, 133-143.

Bain, G., Gruenwald, S. and Murre, C., 1993. E2A and E2-2 are subunits of B-cell-specific E2-box DNA-binding proteins. Mol Cell Biol. 13, 3522-3529.

Baratz, K. H., Tosakulwong, N., Ryu, E., Brown, W. L., Branham, K., Chen, W., Tran, K. D., Schmid-Kubista, K. E., Heckenlively, J. R., Swaroop, A., Abecasis, G., Bailey, K. R. and Edwards, A. O., 2010. E2-2 protein and Fuchs's corneal dystrophy. N Engl J Med. 363, 1016-1024.

Bardin, A. J., Perdigoto, C. N., Southall, T. D., Brand, A. H. and Schweisguth, F., 2010. Transcriptional control of stem cell maintenance in the Drosophila intestine. Development. 137, 705-714.

Bar-On, L., Birnberg, T., Lewis, K. L., Edelson, B. T., Bruder, D., Hildner, K., Buer, J., Murphy, K. M., Reizis, B. and Jung, S., 2010. CX3CR1+ CD8alpha+ dendritic cells are a steady-state population related to plasmacytoid dendritic cells. Proc Natl Acad Sci U S A. 107, 14745-14750.

Bayly, R., Chuen, L., Currie, R. A., Hyndman, B. D., Casselman, R., Blobel, G. A. and LeBrun, D. P., 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.

Bayly, R., Murase, T., Hyndman, B. D., Savage, R., Nurmohamed, S., Munro, K., Casselman, R., Smith, S. P. and LeBrun, D. P., 2006. Critical role for a single leucine residue in leukemia induction by E2A-PBX1. Mol Cell Biol. 26, 6442-6452.

Beck, G. R., Jr., Zerler, B. and Moran, E., 2001. Gene array analysis of osteoblast differentiation. Cell Growth Differ. 12, 61-83.

Belfield, J. L., Whittaker, C., Cader, M. Z. and Chawla, S., 2006. Differential effects of Ca2+ and cAMP on transcription mediated by MEF2D and cAMP-response element-binding protein in hippocampal neurons. J Biol Chem. 281, 27724-27732.

Beltran, A. C., Dawson, P. E. and Gottesfeld, J. M., 2005. Role of DNA sequence in the binding specificity of synthetic basic-helix-loop-helix domains. Chembiochem. 6, 104-113.

Bergqvist, I., Eriksson, M., Saarikettu, J., Eriksson, B., Corneliussen, B., Grundstrom, T. and Holmberg, D., 2000. The basic helix-loop-helix transcription factor E2-2 is involved in T lymphocyte development. Eur J Immunol. 30, 2857-2863.

Bhalla, S., Spaulding, C., Brumbaugh, R. L., Zagort, D. E., Massari, M. E., Murre, C. and Kee, B. L., 2008. Differential roles for the E2A activation domains in B lymphocytes and macrophages. J Immunol. 180, 1694-1703.

Bian, F., Liu, W., Yoon, K. C., Lu, R., Zhou, N., Ma, P., Pflugfelder, S. C. and Li, D. Q., 2012. Molecular signatures and biological pathway profiles of human corneal epithelial progenitor cells. Int J Biochem Cell Biol. 42, 1142-1153.

Blick, T., Widodo, E., Hugo, H., Waltham, M., Lenburg, M. E., Neve, R. M. and Thompson, E. W., 2008. Epithelial mesenchymal transition traits in human breast cancer cell lines. Clin Exp Metastasis. 25, 629-642.

Bourne, W. M., 2003. Biology of the corneal endothelium in health and disease. Eye (Lond). 17, 912-918.

Bowen, T., Guy, C. A., Cardno, A. G., Vincent, J. B., Kennedy, J. L., Jones, L. A., Gray, M., Sanders, R. D., McCarthy, G., Murphy, K. C., Owen, M. J. and O'Donovan, M. C., 2000. Repeat sizes at CAG/CTG loci CTG18.1, ERDA1 and TGC13-7a in schizophrenia. Psychiatr Genet. 10, 33-37.

Bradney, C., Hjelmeland, M., Komatsu, Y., Yoshida, M., Yao, T. P. and Zhuang, Y., 2003. Regulation of E2A activities by histone acetyltransferases in B lymphocyte development. J Biol Chem. 278, 2370-2376.

Braun, T. and Arnold, H. H., 1991. The four human muscle regulatory helix-loop-helix proteins Myf3-Myf6 exhibit similar hetero-dimerization and DNA binding properties. Nucleic Acids Res. 19, 5645-5651.

Brennand, K. J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., McCarthy, S., Sebat, J. and Gage, F. H., 2011. Modelling schizophrenia using human induced pluripotent stem cells. Nature. 473, 221-225.

Breschel, T. S., McInnis, M. G., Margolis, R. L., Sirugo, G., Corneliussen, B., Simpson, S. G., McMahon, F. J., MacKinnon, D. F., Xu, J. F., Pleasant, N., Huo, Y., Ashworth, R. G., Grundstrom, C., Grundstrom, T., Kidd, K. K., DePaulo, J. R. and Ross, C. A., 1997. A novel, heritable, expanding CTG repeat in an intron of the SEF2-1 gene on chromosome 18q21.1. Hum Mol Genet. 6, 1855-1863.

Breslin, M. B., Zhu, M. and Lan, M. S., 2003. NeuroD1/E47 regulates the E-box element of a novel zinc finger transcription factor, IA-1, in developing nervous system. J Biol Chem. 278, 38991-38997.

Brockschmidt, A., Filippi, A., Charbel Issa, P., Nelles, M., Urbach, H., Eter, N., Driever, W. and Weber, R. G., 2011. Neurologic and ocular phenotype in Pitt-Hopkins syndrome and a zebrafish model. Hum Genet. 130, 645-655.

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.

Brzozka, 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.

Brzozka, M. M. and Rossner, M. J., 2013. Deficits in trace fear memory in a mouse model of the schizophrenia risk gene TCF4. Behav Brain Res. 237, 348-356.

Buck, M. J. and Atchley, W. R., 2003. Phylogenetic analysis of plant basic helix-loop-helix proteins. J Mol Evol. 56, 742-750.

Buonocore, F., Hill, M. J., Campbell, C. D., Oladimeji, P. B., Jeffries, A. R., Troakes, C., Hortobagyi, T., Williams, B. P., Cooper, J. D. and Bray, N. J.,

2010. Effects of cis-regulatory variation differ across regions of the adult human brain. Hum Mol Genet. 19, 4490-4496.

Burgess, C. E., Lindblad, K., Sidransky, E., Yuan, Q. P., Long, R. T., Breschel, T., Ross, C. A., McInnis, M., Lee, P., Ginns, E. I., Lenane, M., Kumra, S., Jacobsen, L., Rapoport, J. L. and Schalling, M., 1998. Large CAG/CTG repeats are associated with childhood-onset schizophrenia. Mol Psychiatry. 3, 321-327.

Cano, A. and Portillo, F., 2010. An emerging role for class I bHLH E2-2 proteins in EMT regulation and tumor progression. Cell Adh Migr. 4, 56-60.

Castilla, M. A., Moreno-Bueno, G., Romero-Perez, L., Van De Vijver, K., Biscuola, M., Lopez-Garcia, M. A., Prat, J., Matias-Guiu, X., Cano, A., Oliva, E. and Palacios, J., 2010. Micro-RNA signature of the epithelial-mesenchymal transition in endometrial carcinosarcoma. J Pathol. 223, 72-80.

Caudy, M., Grell, E. H., Dambly-Chaudiere, C., Ghysen, A., Jan, L. Y. and Jan, Y. N., 1988. The maternal sex determination gene daughterless has zygotic activity necessary for the formation of peripheral neurons in Drosophila. Genes Dev. 2, 843-852.

Cervantes-Barragan, L., Lewis, K. L., Firner, S., Thiel, V., Hugues, S., Reith, W., Ludewig, B. and Reizis, B., 2012. Plasmacytoid dendritic cells control T-cell response to chronic viral infection. Proc Natl Acad Sci U S A. 109, 3012-3017.

Chandrasekaran, V. and Beckendorf, S. K., 2003. senseless is necessary for the survival of embryonic salivary glands in Drosophila. Development. 130, 4719-4728.

Chavali, G. B., Vijayalakshmi, C. and Salunke, D. M., 2001. Analysis of sequence signature defining functional specificity and structural stability in helix-loop-helix proteins. Proteins. 42, 471-480.

Chawla, S., Hardingham, G. E., Quinn, D. R. and Bading, H., 1998. CBP: a signal-regulated transcriptional coactivator controlled by nuclear calcium and CaM kinase IV. Science. 281, 1505-1509.

Chen, B. and Lim, R. W., 1997. Physical and functional interactions between the transcriptional inhibitors Id3 and ITF-2b. Evidence toward a novel mechanism regulating muscle-specific gene expression. J Biol Chem. 272, 2459-2463.

Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R. and Greenberg, M. E., 2003. Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science. 302, 885-889.

Chiaramello, A., Soosaar, A., Neuman, T. and Zuber, M. X., 1995. Differential expression and distinct DNA-binding specificity of ME1a and ME2 suggest a unique role during differentiation and neuronal plasticity. Brain Res Mol Brain Res. 29, 107-118.

Chiou, G. Y., Cherng, J. Y., Hsu, H. S., Wang, M. L., Tsai, C. M., Lu, K. H., Chien, Y., Hung, S. C., Chen, Y. W., Wong, C. I., Tseng, L. M., Huang, P. I., Yu, C. C., Hsu, W. H. and Chiou, S. H., 2012. Cationic polyurethanes-

short branch PEI-mediated delivery of Mir145 inhibited epithelial-mesenchymal transdifferentiation and cancer stem-like properties and in lung adenocarcinoma. J Control Release. 159, 240-250.

Ch'ng, T. H., Uzgil, B., Lin, P., Avliyakulov, N. K., O'Dell, T. J. and Martin, K. C., 2012. Activity-dependent transport of the transcriptional coactivator CRTC1 from synapse to nucleus. Cell. 150, 207-221.

Chow, C. W. and Davis, R. J., 2000. Integration of calcium and cyclic AMP signaling pathways by 14-3-3. Mol Cell Biol. 20, 702-712.

Cisse, B., Caton, M. L., Lehner, M., Maeda, T., Scheu, S., Locksley, R., Holmberg, D., Zweier, C., den Hollander, N. S., Kant, S. G., Holter, W., Rauch, A., Zhuang, Y. and Reizis, B., 2008. Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. Cell. 135, 37-48.

Cline, T. W., 1989. The affairs of daughterless and the promiscuity of developmental regulators. Cell. 59, 231-234.

Cody, J. D., Carter, E. M., Sebold, C., Heard, P. L. and Hale, D. E., 2009a. A gene dosage map of Chromosome 18: a map with clinical utility. Genet Med. 11, 778-782.

Cody, J. D., Heard, P. L., Crandall, A. C., Carter, E. M., Li, J., Hardies, L. J., Lancaster, J., Perry, B., Stratton, R. F., Sebold, C., Schaub, R. L., Soileau, B., Hill, A., Hasi, M., Fox, P. T. and Hale, D. E., 2009b. Narrowing critical regions and determining penetrance for selected 18q- phenotypes. Am J Med Genet A. 149A, 1421-1430.

Corneliussen, B., Holm, M., Waltersson, Y., Onions, J., Hallberg, B., Thornell, A. and Grundstrom, T., 1994. Calcium/calmodulin inhibition of basic-helix-loop-helix transcription factor domains. Nature. 368, 760-764.

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

Craddock, N. and Forty, L., 2006. Genetics of affective (mood) disorders. Eur J Hum Genet. 14, 660-668.

Craddock, N. and Owen, M. J., 2010. The Kraepelinian dichotomy - going, going... but still not gone. Br J Psychiatry. 196, 92-95.

Cummings, C. A. and Cronmiller, C., 1994. The daughterless gene functions together with Notch and Delta in the control of ovarian follicle development in Drosophila. Development. 120, 381-394.

Davis, R. L. and Turner, D. L., 2001. Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning. Oncogene. 20, 8342-8357.

De Masi, F., Grove, C. A., Vedenko, A., Alibes, A., Gisselbrecht, S. S., Serrano, L., Bulyk, M. L. and Walhout, A. J., 2011. Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. Nucleic Acids Res. 39, 4553-4563.

de Pontual, L., Mathieu, Y., Golzio, C., Rio, M., Malan, V., Boddaert, N., Soufflet, C., Picard, C., Durandy, A., Dobbie, A., Heron, D., Isidor, B., Motte, J.,

Newburry-Ecob, R., Pasquier, L., Tardieu, M., Viot, G., Jaubert, F., Munnich, A., Colleaux, L., Vekemans, M., Etchevers, H., Lyonnet, S. and Amiel, J., 2009. Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. Hum Mutat. 30, 669-676.

de Pooter, R. F. and Kee, B. L., 2010. E proteins and the regulation of early lymphocyte development. Immunol Rev. 238, 93-109.

Deed, R. W., Armitage, S. and Norton, J. D., 1996. Nuclear localization and regulation of Id protein through an E protein-mediated chaperone mechanism. J Biol Chem. 271, 23603-23606.

Del-Favero, J., Gestel, S. V., Borglum, A. D., Muir, W., Ewald, H., Mors, O., Ivezic, S., Oruc, L., Adolfsson, R., Blackwood, D., Kruse, T., Mendlewicz, J., Schalling, M. and Van Broeckhoven, C., 2002. European combined analysis of the CTG18.1 and the ERDA1 CAG/CTG repeats in bipolar disorder. Eur J Hum Genet. 10, 276-280.

Deliri, H., Meller, N., Kadakkal, A., Malhotra, R., Brewster, J., Doran, A. C., Pei, H., Oldham, S. N., Skaflen, M. D., Garmey, J. C. and McNamara, C. A., 2011. Increased 12/15-lipoxygenase enhances cell growth, fibronectin deposition, and neointimal formation in response to carotid injury. Arterioscler Thromb Vasc Biol. 31, 110-116.

Dhulipala, P. D., Lianos, E. A. and Kotlikoff, M. I., 2001. Regulation of human P2X1 promoter activity by beta helix-loop-helix factors in smooth muscle cells. Gene. 269, 167-175.

Dolmetsch, R. E., Pajvani, U., Fife, K., Spotts, J. M. and Greenberg, M. E., 2001. Signaling to the nucleus by an L-type calcium channel-calmodulin complex through the MAP kinase pathway. Science. 294, 333-339.

Dorflinger, U., Pscherer, A., Moser, M., Rummele, P., Schule, R. and Buettner, R., 1999. Activation of somatostatin receptor II expression by transcription factors MIBP1 and SEF-2 in the murine brain. Mol Cell Biol. 19, 3736-3747.

Dunn, L. L., Sekyere, E. O., Suryo Rahmanto, Y. and Richardson, D. R., 2006. The function of melanotransferrin: a role in melanoma cell proliferation and tumorigenesis. Carcinogenesis. 27, 2157-2169.

Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., Berx, G., Cano, A., Beug, H. and Foisner, R., 2005. DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. Oncogene. 24, 2375-2385.

Eghrari, A. O., McGlumphy, E. J., Iliff, B. W., Wang, J., Emmert, D., Riazuddin, S. A., Katsanis, N. and Gottsch, J. D., 2012. Prevalence and Severity of Fuchs Corneal Dystrophy in Tangier Island. Am J Ophthalmol. 153, 1067-1072.

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.

Elhalis, H., Azizi, B. and Jurkunas, U. V., 2010. Fuchs endothelial corneal dystrophy. Ocul Surf. 8, 173-184.

Ellenberger, T., Fass, D., Arnaud, M. and Harrison, S. C., 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev. 8, 970-980.

Fairman, R., Beran-Steed, R. K. and Handel, T. M., 1997. Heteronuclear (1H, 13C, 15N) NMR assignments and secondary structure of the basic region-helix-loop-helix domain of E47. Protein Sci. 6, 175-184.

Feng, Y., Lee, N. and Fearon, E. R., 2003. TIP49 regulates beta-catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. Cancer Res. 63, 8726-8734.

Ferre-D'Amare, A. R., Pognonec, P., Roeder, R. G. and Burley, S. K., 1994. Structure and function of the b/HLH/Z domain of USF. EMBO J. 13, 180-189.

Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B. and Burley, S. K., 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature. 363, 38-45.

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.

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.

Fortune, M. T., Kennedy, J. L. and Vincent, J. B., 2003. Anticipation and CAG*CTG repeat expansion in schizophrenia and bipolar affective disorder. Curr Psychiatry Rep. 5, 145-154.

Furumura, M., Potterf, S. B., Toyofuku, K., Matsunaga, J., Muller, J. and Hearing, V. J., 2001. Involvement of ITF2 in the transcriptional regulation of melanogenic genes. J Biol Chem. 276, 28147-28154.

Gaudilliere, B., Konishi, Y., de la Iglesia, N., Yao, G. and Bonni, A., 2004. A CaMKII-NeuroD signaling pathway specifies dendritic morphogenesis. Neuron. 41, 229-241.

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.

Ghosh, P. S., Freidman, N. R. and Ghosh, D., in press. Pitt-Hopkins Syndrome in a Boy With Charcot Marie Tooth Disease Type 1A: A Rare Co-occurrence of 2 Genetic Disorders. J Child Neurol.

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

Goardon, N., Lambert, J. A., Rodriguez, P., Nissaire, P., Herblot, S., Thibault, P., Dumenil, D., Strouboulis, J., Romeo, P. H. and Hoang, T., 2006. ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. EMBO J. 25, 357-366.

Goldfarb, A. N. and Lewandowska, K., 1994. Nuclear redirection of a cytoplasmic helix-loop-helix protein via heterodimerization with a nuclear localizing partner. Exp Cell Res. 214, 481-485.

Goldfarb, A. N. and Lewandowska, K., 1995. Inhibition of cellular differentiation by the SCL/tal oncoprotein: transcriptional repression by an Id-like mechanism. Blood. 85, 465-471.

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.

Goldfarb, A. N., Lewandowska, K. and Shoham, M., 1996. Determinants of helix-loop-helix dimerization affinity. Random mutational analysis of SCL/tal. J Biol Chem. 271, 2683-2688.

Gonzalez, G. A. and Montminy, M. R., 1989. Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell. 59, 675-680.

Goossens, D., Del-Favero, J. and Van Broeckhoven, C., 2001. Trinucleotide repeat expansions: do they contribute to bipolar disorder? Brain Res Bull. 56, 243-257.

Greer, P. L. and Greenberg, M. E., 2008. From synapse to nucleus: calcium-dependent gene transcription in the control of synapse development and function. Neuron. 59, 846-860.

Grooteclaes, M. L. and Frisch, S. M., 2000. Evidence for a function of CtBP in epithelial gene regulation and anoikis. Oncogene. 19, 3823-3828.

Grove, C. A., De Masi, F., Barrasa, M. I., Newburger, D. E., Alkema, M. J., Bulyk, M. L. and Walhout, A. J., 2009. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. Cell. 138, 314-327.

Guillemot, F., 2007. Spatial and temporal specification of neural fates by transcription factor codes. Development. 134, 3771-3780.

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.

Gustavsson, P., Kimber, E., Wahlstrom, J. and Anneren, G., 1999. Monosomy 18q syndrome and atypical Rett syndrome in a girl with an interstitial deletion (18)(q21.1q22.3). Am J Med Genet. 82, 348-351.

Guy, C. A., Bowen, T., Jones, I., McCandless, F., Owen, M. J., Craddock, N. and O'Donovan, M. C., 1999. CTG18.1 and ERDA-1 CAG/CTG repeat size in bipolar disorder. Neurobiol Dis. 6, 302-307.

Hamdan, F. F., Daoud, H., Patry, L., Dionne-Laporte, A., Spiegelman, D., Dobrzeniecka, S., Rouleau, G. A. and Michaud, J. L., in press. Parent-child exome sequencing identifiesa de novo truncating mutation in TCF4 in non-syndromic intellectual disability. Clin Genet.

Hasi, M., Soileau, B., Sebold, C., Hill, A., Hale, D. E., O'Donnell, L. and Cody, J. D., 2011. The role of the TCF4 gene in the phenotype of individuals with 18q segmental deletions. Hum Genet. 130, 777-787.

Hauser, J., Saarikettu, J. and Grundstrom, T., 2008a. Calcium regulation of myogenesis by differential calmodulin inhibition of basic helix-loop-helix transcription factors. Mol Biol Cell. 19, 2509-2519.

Hauser, J., Sveshnikova, N., Wallenius, A., Baradaran, S., Saarikettu, J. and Grundstrom, T., 2008b. B-cell receptor activation inhibits AID expression through calmodulin inhibition of E-proteins. Proc Natl Acad Sci U S A. 105, 1267-1272.

Hauser, J., Wallenius, A., Sveshnikova, N., Saarikettu, J. and Grundstrom, T., 2010. Calmodulin inhibition of E2A stops expression of surrogate light chains of the pre-B-cell receptor and CD19. Mol Immunol. 47, 1031-1038.

Hauser, J., Verma-Gaur, J., Wallenius, A. and Grundstrom, T., 2009. Initiation of antigen receptor-dependent differentiation into plasma cells by calmodulin inhibition of E2A. J Immunol. 183, 1179-1187.

Hayashi, K., Sueishi, K., Tanaka, K. and Inomata, H., 1986. Immunohistochemical evidence of the origin of human corneal endothelial cells and keratocytes. Graefes Arch Clin Exp Ophthalmol. 224, 452-456.

Heard, P. L., Carter, E. M., Crandall, A. C., Sebold, C., Hale, D. E. and Cody, J. D., 2009. High resolution genomic analysis of 18q- using oligo-microarray comparative genomic hybridization (aCGH). Am J Med Genet A. 149A, 1431-1437.

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

Henthorn, P., McCarrick-Walmsley, R. and Kadesch, T., 1990b. Sequence of the cDNA encoding ITF-2, a positive-acting transcription factor. Nucleic Acids Res. 18, 678.

Herbst, A., Bommer, G. T., Kriegl, L., Jung, A., Behrens, A., Csanadi, E., Gerhard, M., Bolz, C., Riesenberg, R., Zimmermann, W., Dietmaier, W., Wolf, I., Brabletz, T., Goke, B. and Kolligs, F. T., 2009a. ITF-2 is disrupted via allelic loss of chromosome 18q21, and ITF-2B expression is lost at the adenoma-carcinoma transition. Gastroenterology. 137, 639-648, 648 e631-639.

Herbst, A., Helferich, S., Behrens, A., Goke, B. and Kolligs, F. T., 2009b. The transcription factor ITF-2A induces cell cycle arrest via p21(Cip1). Biochem Biophys Res Commun. 387, 736-740.

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.

Hikima, J., Lennard, M. L., Wilson, M. R., Miller, N. W., Clem, L. W. and Warr, G. W., 2005. Evolution of vertebrate E protein transcription factors: comparative analysis of the E protein gene family in Takifugu rubripes and humans. Physiol Genomics. 21, 144-151.

Hirsch, M. R., Tiveron, M. C., Guillemot, F., Brunet, J. F. and Goridis, C., 1998. Control of noradrenergic differentiation and Phox2a expression by MASH1 in the central and peripheral nervous system. Development. 125, 599-608.

Hong, J., Zhou, J., Fu, J., He, T., Qin, J., Wang, L., Liao, L. and Xu, J., 2011. Phosphorylation of serine 68 of Twist1 by MAPKs stabilizes Twist1 protein and promotes breast cancer cell invasiveness. Cancer Res. 71, 3980-3990.

Hosoda, H., Motohashi, J., Kato, H., Masushige, S. and Kida, S., 2004. A BMAL1 mutant with arginine 91 substituted with alanine acts as a dominant negative inhibitor. Gene. 338, 235-241.

Hsu, H. L., Huang, L., Tsan, J. T., Funk, W., Wright, W. E., Hu, J. S., Kingston, R. E. and Baer, R., 1994. Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. Mol Cell Biol. 14, 1256-1265.

Hu, J. S., Olson, E. N. and Kingston, R. E., 1992. HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. Mol Cell Biol. 12, 1031-1042.

Igo, R. P., Jr., Kopplin, L. J., Joseph, P., Truitt, B., Fondran, J., Bardenstein, D., Aldave, A. J., Croasdale, C. R., Price, M. O., Rosenwasser, M., Lass, J. H. and Iyengar, S. K., 2012. Differing Roles for TCF4 and COL8A2 in Central Corneal Thickness and Fuchs Endothelial Corneal Dystrophy. PLoS One. 7, e46742.

Ik Tsen Heng, J. and Tan, S. S., 2003. The role of class I HLH genes in neural development--have they been overlooked? Bioessays. 25, 709-716.

Impey, S., Obrietan, K., Wong, S. T., Poser, S., Yano, S., Wayman, G., Deloulme, J. C., Chan, G. and Storm, D. R., 1998. Cross talk between ERK and PKA is required for Ca2+ stimulation of CREB-dependent transcription and ERK nuclear translocation. Neuron. 21, 869-883.

Ince-Dunn, G., Hall, B. J., Hu, S. C., Ripley, B., Huganir, R. L., Olson, J. M., Tapscott, S. J. and Ghosh, A., 2006. Regulation of thalamocortical patterning and synaptic maturation by NeuroD2. Neuron. 49, 683-695.

Itahana, Y., Piens, M., Fong, S., Singh, J., Sumida, T. and Desprez, P. Y., 2008. Expression of Id and ITF-2 genes in the mammary gland during pregnancy. Biochem Biophys Res Commun. 372, 826-830.

Ivleva, E., Thaker, G. and Tamminga, C. A., 2008. Comparing genes and phenomenology in the major psychoses: schizophrenia and bipolar 1 disorder. Schizophr Bull. 34, 734-742.

Javaux, F., Donda, A., Vassart, G. and Christophe, D., 1991. Cloning and sequence analysis of TFE, a helix-loop-helix transcription factor able to recognize the thyroglobulin gene promoter in vitro. Nucleic Acids Res. 19, 1121-1127.

Jo, C., Cho, S. J. and Jo, S. A., 2011. Mitogen-activated protein kinase kinase 1 (MEK1) stabilizes MyoD through direct phosphorylation at tyrosine 156 during myogenic differentiation. J Biol Chem. 286, 18903-18913.

Jogi, A., Persson, P., Grynfeld, A., Pahlman, S. and Axelson, H., 2002. Modulation of basic helix-loop-helix transcription complex formation by Id proteins during neuronal differentiation. J Biol Chem. 277, 9118-9126.

Jones, S., 2004. An overview of the basic helix-loop-helix proteins. Genome Biol. 5, 226.

Joo, J. K., Kim, S. H., Kim, H. G., Kim, D. Y., Ryu, S. Y., Lee, K. H. and Lee, J. H., 2010. CpG methylation of transcription factor 4 in gastric carcinoma. Ann Surg Oncol. 17, 3344-3353.

Jurkunas, U. V., Bitar, M. S., Funaki, T. and Azizi, B., 2010. Evidence of oxidative stress in the pathogenesis of fuchs endothelial corneal dystrophy. Am J Pathol. 177, 2278-2289.

Kalscheuer, V. M., Feenstra, I., Van Ravenswaaij-Arts, C. M., Smeets, D. F., Menzel, C., Ullmann, R., Musante, L. and Ropers, H. H., 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.

Kannabiran, C., 2009. Genetics of corneal endothelial dystrophies. J Genet. 88, 487-494.

Kato, Z., Morimoto, W., Kimura, T., Matsushima, A. and Kondo, N., 2010. Interstitial deletion of 18q: comparative genomic hybridization array analysis of 46, XX,del(18)(q21.2.q21.33). Birth Defects Res A Clin Mol Teratol. 88, 132-135.

Kawai-Kowase, K., Kumar, M. S., Hoofnagle, M. H., Yoshida, T. and Owens, G. K., 2005. PIAS1 activates the expression of smooth muscle cell differentiation marker genes by interacting with serum response factor and class I basic helix-loop-helix proteins. Mol Cell Biol. 25, 8009-8023.

Kee, B. L., 2009. E and ID proteins branch out. Nat Rev Immunol. 9, 175-184.

Kim, S., Cho, H., Lee, D. and Webster, M. J., 2012. Association between SNPs and gene expression in multiple regions of the human brain. Transl Psychiatry. 2, e113.

Kim, S. K., Jang, H. R., Kim, J. H., Kim, M., Noh, S. M., Song, K. S., Kang, G. H., Kim, H. J., Kim, S. Y., Yoo, H. S. and Kim, Y. S., 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.

Klein, E. S., Simmons, D. M., Swanson, L. W. and Rosenfeld, M. G., 1993. Tissue-specific RNA splicing generates an ankyrin-like domain that affects the dimerization and DNA-binding properties of a bHLH protein. Genes Dev. 7, 55-71.

Klintworth, G. K., 2009. Corneal dystrophies. Orphanet J Rare Dis. 4, 7.

Kolligs, F. T., Nieman, M. T., Winer, I., Hu, G., Van Mater, D., Feng, Y., Smith, I. M., Wu, R., Zhai, Y., Cho, K. R. and Fearon, E. R., 2002. ITF-2, a downstream target of the Wnt/TCF pathway, is activated in human cancers with beta-catenin defects and promotes neoplastic transformation. Cancer Cell. 1, 145-155.

Kotaleski, J. H. and Blackwell, K. T., Modelling the molecular mechanisms of synaptic plasticity using systems biology approaches. Nat Rev Neurosci. 11, 239-251.

Krause, M., Park, M., Zhang, J. M., Yuan, J., Harfe, B., Xu, S. Q., Greenwald, I., Cole, M., Paterson, B. and Fire, A., 1997. A C. elegans E/Daughterless bHLH protein marks neuronal but not striated muscle development. Development. 124, 2179-2189.

Kuipers, H., Schnorfeil, F. M. and Brocker, T., 2010. Differentially expressed microRNAs regulate plasmacytoid vs. conventional dendritic cell development. Mol Immunol. 48, 333-340.

Kumar, M. S., Hendrix, J. A., Johnson, A. D. and Owens, G. K., 2003. Smooth muscle alpha-actin gene requires two E-boxes for proper expression in vivo and is a target of class I basic helix-loop-helix proteins. Circ Res. 92, 840-847.

Kuot, A., Hewitt, A. W., Griggs, K., Klebe, S., Mills, R., Jhanji, V., Craig, J. E., Sharma, S. and Burdon, K. P., 2012. Association of TCF4 and CLU polymorphisms with Fuchs' endothelial dystrophy and implication of CLU and TGFBI proteins in the disease process. Eur J Hum Genet. 20, 632-638.

Kurian, S. M., Le-Niculescu, H., Patel, S. D., Bertram, D., Davis, J., Dike, C., Yehyawi, N., Lysaker, P., Dustin, J., Caligiuri, M., Lohr, J., Lahiri, D. K., Nurnberger, J. I., Jr., Faraone, S. V., Geyer, M. A., Tsuang, M. T., Schork, N. J., Salomon, D. R. and Niculescu, A. B., 2011. Identification of blood biomarkers for psychosis using convergent functional genomics. Mol Psychiatry. 16, 37-58.

Kwon, E., Wang, W. and Tsai, L. H., in press. Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. Mol Psychiatry.

Langlands, K., Yin, X., Anand, G. and Prochownik, E. V., 1997. Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. J Biol Chem. 272, 19785-19793.

Larsson, G., Schleucher, J., Onions, J., Hermann, S., Grundstrom, T. and Wijmenga, S. S., 2001. A novel target recognition revealed by calmodulin in complex with the basic helix--loop--helix transcription factor SEF2-1/E2-2. Protein Sci. 10, 169-186.

Larsson, G., Schleucher, J., Onions, J., Hermann, S., Grundstrom, T. and Wijmenga, S. S., 2005. Backbone dynamics of a symmetric calmodulin dimer in complex with the calmodulin-binding domain of the basic-helix-loop-helix transcription factor SEF2-1/E2-2: a highly dynamic complex. Biophys J. 89, 1214-1226.

Ledent, V., Paquet, O. and Vervoort, M., 2002. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol. 3, RESEARCH0030.

Ledent, V. and Vervoort, M., 2001. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res. 11, 754-770.

Lehalle, D., Williams, C., Siu, V. M. and Clayton-Smith, J., 2011. Fetal pads as a clue to the diagnosis of Pitt-Hopkins syndrome. Am J Med Genet A. 155A, 1685-1689.

Lemieux, M. E., Cheng, Z., Zhou, Q., White, R., Cornell, J., Kung, A. L. and Rebel, V. I., 2011. Inactivation of a single copy of Crebbp selectively alters

pre-mRNA processing in mouse hematopoietic stem cells. PLoS One. 6, e24153.

Le-Niculescu, H., Balaraman, Y., Patel, S., Tan, J., Sidhu, K., Jerome, R. E., Edenberg, H. J., Kuczenski, R., Geyer, M. A., Nurnberger, J. I., Jr., Faraone, S. V., Tsuang, M. T. and Niculescu, A. B., 2007. Towards understanding the schizophrenia code: an expanded convergent functional genomics approach. Am J Med Genet B Neuropsychiatr Genet. 144B, 129-158.

Lennertz, L., Rujescu, D., Wagner, M., Frommann, I., Schulze-Rauschenbach, S., Schuhmacher, A., Landsberg, M. W., Franke, P., Moller, H. J., Wolwer, W., Gaebel, W., Hafner, H., Maier, W. and Mossner, R., 2011. Novel schizophrenia risk gene TCF4 influences verbal learning and memory functioning in schizophrenia patients. Neuropsychobiology. 63, 131-136.

Li, G., Luna, C., Qiu, J., Epstein, D. L. and Gonzalez, P., 2011a. Role of miR-204 in the regulation of apoptosis, endoplasmic reticulum stress response, and inflammation in human trabecular meshwork cells. Invest Ophthalmol Vis Sci. 52, 2999-3007.

Li, H. S., Yang, C. Y., Nallaparaju, K. C., Zhang, H., Liu, Y. J., Goldrath, A. W. and Watowich, S. S., in press. The signal transducers STAT5 and STAT3 control expression of Id2 and E2-2 during dendritic cell development. Blood.

Li, T., Li, Z., Chen, P., Zhao, Q., Wang, T., Huang, K., Li, J., Li, Y., Liu, J., Zeng, Z., Feng, G., He, L. and Shi, Y., 2010. Common variants in major histocompatibility complex region and TCF4 gene are significantly associated with schizophrenia in Han Chinese. Biol Psychiatry. 68, 671-673.

Li, X., Duan, X., Jiang, H., Sun, Y., Tang, Y., Yuan, Z., Guo, J., Liang, W., Chen, L., Yin, J., Ma, H., Wang, J. and Zhang, D., 2006. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. Plant Physiol. 141, 1167-1184.

Li, Y. J., Minear, M. A., Rimmler, J., Zhao, B., Balajonda, E., Hauser, M. A., Allingham, R. R., Eghrari, A. O., Riazuddin, S. A., Katsanis, N., Gottsch, J. D., Gregory, S. G., Klintworth, G. K. and Afshari, N. A., 2011b. Replication of TCF4 through association and linkage studies in late-onset Fuchs endothelial corneal dystrophy. PLoS One. 6, e18044.

Lichtenstein, P., Yip, B. H., Bjork, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F. and Hultman, C. M., 2009. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet. 373, 234-239.

Lin, C. H. and Lee, E. H., 2012. JNK1 inhibits GluR1 expression and GluR1-mediated calcium influx through phosphorylation and stabilization of Hes-1. J Neurosci. 32, 1826-1846.

Lindblad, K., Nylander, P. O., De bruyn, A., Sourey, D., Zander, C., Engstrom, C., Holmgren, G., Hudson, T., Chotai, J., Mendlewicz, J. and et al., 1995. Detection of expanded CAG repeats in bipolar affective disorder using the repeat expansion detection (RED) method. Neurobiol Dis. 2, 55-62.

Lindblad, K., Nylander, P. O., Zander, C., Yuan, Q. P., Stahle, L., Engstrom, C., Balciuniene, J., Pettersson, U., Breschel, T., McInnis, M., Ross, C. A.,

Adolfsson, R. and Schalling, M., 1998. Two commonly expanded CAG/CTG repeat loci: involvement in affective disorders? Mol Psychiatry. 3, 405-410.

Lingbeck, J. M., Trausch-Azar, J. S., Ciechanover, A. and Schwartz, A. L., 2005. E12 and E47 modulate cellular localization and proteasome-mediated degradation of MyoD and Id1. Oncogene. 24, 6376-6384.

Linseman, D. A., Bartley, C. M., Le, S. S., Laessig, T. A., Bouchard, R. J., Meintzer, M. K., Li, M. and Heidenreich, K. A., 2003. Inactivation of the myocyte enhancer factor-2 repressor histone deacetylase-5 by endogenous Ca(2+) //calmodulin-dependent kinase II promotes depolarization-mediated cerebellar granule neuron survival. J Biol Chem. 278, 41472-41481.

Liu, Y., Ray, S. K., Yang, X. Q., Luntz-Leybman, V. and Chiu, I. M., 1998. A splice variant of E2-2 basic helix-loop-helix protein represses the brain-specific fibroblast growth factor 1 promoter through the binding to an imperfect E-box. J Biol Chem. 273, 19269-19276.

Liu, Y. P., Burleigh, D., Durning, M., Hudson, L., Chiu, I. M. and Golos, T. G., 2004. Id2 is a primary partner for the E2-2 basic helix-loop-helix transcription factor in the human placenta. Mol Cell Endocrinol. 222, 83-91.

Lluis, F., Ballestar, E., Suelves, M., Esteller, M. and Munoz-Canoves, P., 2005. E47 phosphorylation by p38 MAPK promotes MyoD/E47 association and muscle-specific gene transcription. EMBO J. 24, 974-984.

Loe-Mie, Y., Lepagnol-Bestel, A. M., Maussion, G., Doron-Faigenboim, A., Imbeaud, S., Delacroix, H., Aggerbeck, L., Pupko, T., Gorwood, P., Simonneau, M. and Moalic, J. M., 2010. SMARCA2 and other genome-wide supported schizophrenia-associated genes: regulation by REST/NRSF, network organization and primate-specific evolution. Hum Mol Genet. 19, 2841-2857.

Longo, A., Guanga, G. P. and Rose, R. B., 2008. Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. Biochemistry. 47, 218-229.

Louttit, M. D., Kopplin, L. J., Igo, R. P., Jr., Fondran, J. R., Tagliaferri, A., Bardenstein, D., Aldave, A. J., Croasdale, C. R., Price, M. O., Rosenwasser, G. O., Lass, J. H. and Iyengar, S. K., 2012. A multicenter study to map genes for Fuchs endothelial corneal dystrophy: baseline characteristics and heritability. Cornea. 31, 26-35.

Loveys, D. A., Streiff, M. B. and Kato, G. J., 1996. E2A basic-helix-loop-helix transcription factors are negatively regulated by serum growth factors and by the Id3 protein. Nucleic Acids Res. 24, 2813-2820.

Lu, R., Bian, F., Zhang, X., Qi, H., Chuang, E. Y., Pflugfelder, S. C. and Li, D. Q., 2011. The beta-catenin/Tcf4/survivin signaling maintains a less differentiated phenotype and high proliferative capacity of human corneal epithelial progenitor cells. Int J Biochem Cell Biol. 43, 751-759.

Lu, R., Qu, Y., Ge, J., Zhang, L., Su, Z., Pflugfelder, S. C. and Li, D. Q., 2012. Transcription factor TCF4 maintains the properties of human corneal epithelial stem cells. Stem Cells. 30, 753-761.

Lu, Y., Sheng, D. Q., Mo, Z. C., Li, H. F., Wu, N. H. and Shen, Y. F., 2005. A negative regulatory element-dependent inhibitory role of ITF2B on IL-2 receptor alpha gene. Biochem Biophys Res Commun. 336, 142-149.

Ma, P. C., Rould, M. A., Weintraub, H. and Pabo, C. O., 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell. 77, 451-459.

Mackey, D. A., Warrington, N. M., Hewitt, A. W., Oates, S. K., Yazar, S., Soloshenko, A., Crawford, G. J., Mountain, J. A. and Pennell, C. E., 2012. Role of the TCF4 gene intronic variant in normal variation of corneal endothelium. Cornea. 31, 162-166.

Maini, I., Cantalupo, G., Turco, E. C., De Paolis, F., Magnani, C., Parrino, L., Terzano, M. G. and Pisani, F., in press. Clinical and Polygraphic Improvement of Breathing Abnormalities After Valproate in a Case of Pitt-Hopkins Syndrome. J Child Neurol.

Marangi, G., Ricciardi, S., Orteschi, D., Lattante, S., Murdolo, M., Dallapiccola, B., Biscione, C., Lecce, R., Chiurazzi, P., Romano, C., Greco, D., Pettinato, R., Sorge, G., Pantaleoni, C., Alfei, E., Toldo, I., Magnani, C., Bonanni, P., Martinez, F., Serra, G., Battaglia, D., Lettori, D., Vasco, G., Baroncini, A., Daolio, C. and Zollino, M., 2011. The Pitt-Hopkins syndrome: report of 16 new patients and clinical diagnostic criteria. Am J Med Genet A. 155A, 1536-1545.

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.

Markus, M., Du, Z. and Benezra, R., 2002. Enhancer-specific modulation of E protein activity. J Biol Chem. 277, 6469-6477.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G. and Sun, Y. E., 2003. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science. 302, 890-893.

Massari, M. E., Grant, P. A., Pray-Grant, M. G., Berger, S. L., Workman, J. L. and Murre, C., 1999. A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. Mol Cell. 4, 63-73.

Massari, M. E., Jennings, P. A. and Murre, C., 1996. The AD1 transactivation domain of E2A contains a highly conserved helix which is required for its activity in both Saccharomyces cerevisiae and mammalian cells. Mol Cell Biol. 16, 121-129.

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

Massari, M. E., Rivera, R. R., Voland, J. R., Quong, M. W., Breit, T. M., van Dongen, J. J., de Smit, O. and Murre, C., 1998. Characterization of ABF-1, a novel basic helix-loop-helix transcription factor expressed in activated B lymphocytes. Mol Cell Biol. 18, 3130-3139.

McInnis, M. G., Swift-Scanlanl, T., Mahoney, A. T., Vincent, J., Verheyen, G., Lan, T. H., Oruc, L., Riess, O., Van Broeckhoven, C., Chen, H., Kennedy, J. L., MacKinnon, D. F., Margolis, R. L., Simpson, S. G., McMahon, F. J., Gershon, E., Nurnberger, J., Reich, T., DePaulo, J. R. and Ross, C. A., 2000. Allelic distribution of CTG18.1 in Caucasian populations: association studies in bipolar disorder, schizophrenia, and ataxia. Mol Psychiatry. 5, 439-442.

McLellan, A. S., Langlands, K. and Kealey, T., 2002. Exhaustive identification of human class II basic helix-loop-helix proteins by virtual library screening. Mech Dev. 119 Suppl 1, S285-291.

Meffert, M. K., Chang, J. M., Wiltgen, B. J., Fanselow, M. S. and Baltimore, D., 2003. NF-kappa B functions in synaptic signaling and behavior. Nat Neurosci. 6, 1072-1078.

Mehta, J. S., Vithana, E. N., Tan, D. T., Yong, V. H., Yam, G. H., Law, R. W., Chong, W. G., Pang, C. P. and Aung, T., 2008. Analysis of the posterior polymorphous corneal dystrophy 3 gene, TCF8, in late-onset Fuchs endothelial corneal dystrophy. Invest Ophthalmol Vis Sci. 49, 184-188.

Meinhardt, G., Husslein, P. and Knofler, M., 2005. Tissue-specific and ubiquitous basic helix-loop-helix transcription factors in human placental trophoblasts. Placenta. 26, 527-539.

Misra, R. P., Bonni, A., Miranti, C. K., Rivera, V. M., Sheng, M. and Greenberg, M. E., 1994. L-type voltage-sensitive calcium channel activation stimulates gene expression by a serum response factor-dependent pathway. J Biol Chem. 269, 25483-25493.

Miyagishi, M., Hatta, M., Ohshima, T., Ishida, J., Fujii, R., Nakajima, T. and Fukamizu, A., 2000. Cell type-dependent transactivation or repression of mesoderm-restricted basic helix-loop-helix protein, POD-1/Capsulin. Mol Cell Biochem. 205, 141-147.

Mologni, L., Dekhil, H., Ceccon, M., Purgante, S., Lan, C., Cleris, L., Magistroni, V., Formelli, F. and Gambacorti-Passerini, C. B., 2010. Colorectal tumors are effectively eradicated by combined inhibition of {beta}-catenin, KRAS, and the oncogenic transcription factor ITF2. Cancer Res. 70, 7253-7263.

Morgan, J. I. and Curran, T., 1986. Role of ion flux in the control of c-fos expression. Nature. 322, 552-555.

Morrow, E. M., Yoo, S. Y., Flavell, S. W., Kim, T. K., Lin, Y., Hill, R. S., Mukaddes, N. M., Balkhy, S., Gascon, G., Hashmi, A., Al-Saad, S., Ware, J., Joseph, R. M., Greenblatt, R., Gleason, D., Ertelt, J. A., Apse, K. A., Bodell, A., Partlow, J. N., Barry, B., Yao, H., Markianos, K., Ferland, R. J., Greenberg, M. E. and Walsh, C. A., 2008. Identifying autism loci and genes by tracing recent shared ancestry. Science. 321, 218-223.

Moskvina, V., Craddock, N., Holmans, P., Nikolov, I., Pahwa, J. S., Green, E., Owen, M. J. and O'Donovan, M. C., 2009. Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. Mol Psychiatry. 14, 252-260.

Mudge, J., Miller, N. A., Khrebtukova, I., Lindquist, I. E., May, G. D., Huntley, J. J., Luo, S., Zhang, L., van Velkinburgh, J. C., Farmer, A. D., Lewis, S., Beavis, W. D., Schilkey, F. D., Virk, S. M., Black, C. F., Myers, M. K., Mader, L. C., Langley, R. J., Utsey, J. P., Kim, R. W., Roberts, R. C., Khalsa, S. K., Garcia, M., Ambriz-Griffith, V., Harlan, R., Czika, W., Martin, S., Wolfinger, R. D., Perrone-Bizzozero, N. I., Schroth, G. P. and Kingsmore, S. F., 2008. Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. PLoS One. 3, e3625.

Muir, T., Sadler-Riggleman, I., Stevens, J. D. and Skinner, M. K., 2006. Role of the basic helix-loop-helix protein ITF2 in the hormonal regulation of Sertoli cell differentiation. Mol Reprod Dev. 73, 491-500.

Muir, T., Wilson-Rawls, J., Stevens, J. D., Rawls, A., Schweitzer, R., Kang, C. and Skinner, M. K., 2008. Integration of CREB and bHLH transcriptional signaling pathways through direct heterodimerization of the proteins: role in muscle and testis development. Mol Reprod Dev. 75, 1637-1652.

Murakami, M., Kataoka, K., Tominaga, J., Nakagawa, O. and Kurihara, H., 2004. Differential cooperation between dHAND and three different E-proteins. Biochem Biophys Res Commun. 323, 168-174.

Murre, C., 2005. Helix-loop-helix proteins and lymphocyte development. Nat Immunol. 6, 1079-1086.

Murre, C., Bain, G., van Dijk, M. A., Engel, I., Furnari, B. A., Massari, M. E., Matthews, J. R., Quong, M. W., Rivera, R. R. and Stuiver, M. H., 1994. Structure and function of helix-loop-helix proteins. Biochim Biophys Acta. 1218, 129-135.

Murre, C., McCaw, P. S. and Baltimore, D., 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell. 56, 777-783.

Nagasawa, M., Schmidlin, H., Hazekamp, M. G., Schotte, R. and Blom, B., 2008. Development of human plasmacytoid dendritic cells depends on the combined action of the basic helix-loop-helix factor E2-2 and the Ets factor Spi-B. Eur J Immunol. 38, 2389-2400.

Navarrete, K., Pedroso, I., De Jong, S., Stefansson, H., Steinberg, S., Stefansson, K., Ophoff, R. A., Schalkwyk, L. C. and Collier, D. A., in press. TCF4 (e2-2; ITF2): A schizophrenia-associated gene with pleiotropic effects on human disease. Am J Med Genet B Neuropsychiatr Genet.

Need, A. C., Shashi, V., Hitomi, Y., Schoch, K., Shianna, K. V., McDonald, M. T., Meisler, M. H. and Goldstein, D. B., 2012. Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet. 49, 353-361.

Neuman, K., Nornes, H. O. and Neuman, T., 1995. Helix-loop-helix transcription factors regulate Id2 gene promoter activity. FEBS Lett. 374, 279-283.

Neuman, T., Keen, A., Knapik, E., Shain, D., Ross, M., Nornes, H. O. and Zuber, M. X., 1993. ME1 and GE1: basic helix-loop-helix transcription factors expressed at high levels in the developing nervous system and in morphogenetically active regions. Eur J Neurosci. 5, 311-318.

Nguyen, P. V. and Woo, N. H., 2003. Regulation of hippocampal synaptic plasticity by cyclic AMP-dependent protein kinases. Prog Neurobiol. 71, 401-437.

O'Donnell, L., Soileau, B., Heard, P., Carter, E., Sebold, C., Gelfond, J., Hale, D. E. and Cody, J. D., 2010. Genetic determinants of autism in individuals with deletions of 18q. Hum Genet. 128, 155-164.

O'Donovan, M., Jones, I. and Craddock, N., 2003. Anticipation and repeat expansion in bipolar disorder. Am J Med Genet C Semin Med Genet. 123C, 10-17.

O'Donovan, M. C., Guy, C., Craddock, N., Murphy, K. C., Cardno, A. G., Jones, L. A., Owen, M. J. and McGuffin, P., 1995. Expanded CAG repeats in schizophrenia and bipolar disorder. Nat Genet. 10, 380-381.

Olson, J. M., Asakura, A., Snider, L., Hawkes, R., Strand, A., Stoeck, J., Hallahan, A., Pritchard, J. and Tapscott, S. J., 2001. NeuroD2 is necessary for development and survival of central nervous system neurons. Dev Biol. 234, 174-187.

Onions, J., Hermann, S. and Grundstrom, T., 1997. Basic helix-loop-helix protein sequences determining differential inhibition by calmodulin and S-100 proteins. J Biol Chem. 272, 23930-23937.

Onions, J., Hermann, S. and Grundstrom, T., 2000. A novel type of calmodulin interaction in the inhibition of basic helix-loop-helix transcription factors. Biochemistry. 39, 4366-4374.

Orrico, A., Galli, L., Zappella, M., Lam, C. W., Bonifacio, S., Torricelli, F. and Hayek, G., 2001. Possible case of Pitt-Hopkins syndrome in sibs. Am J Med Genet. 103, 157-159.

Pace, C. N. and Scholtz, J. M., 1998. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J. 75, 422-427.

Pagliuca, A., Cannada-Bartoli, P. and Lania, L., 1998. A role for Sp and helix-loop-helix transcription factors in the regulation of the human Id4 gene promoter activity. J Biol Chem. 273, 7668-7674.

Pagliuca, A., Gallo, P., De Luca, P. and Lania, L., 2000. Class A helix-loop-helix proteins are positive regulators of several cyclin-dependent kinase inhibitors' promoter activity and negatively affect cell growth. Cancer Res. 60, 1376-1382.

Park, S., Chen, W., Cierpicki, T., Tonelli, M., Cai, X., Speck, N. A. and Bushweller, J. H., 2009. Structure of the AML1-ETO eTAFH domain-HEB peptide complex and its contribution to AML1-ETO activity. Blood. 113, 3558-3567.

Parrinello, S., Lin, C. Q., Murata, K., Itahana, Y., Singh, J., Krtolica, A., Campisi, J. and Desprez, P. Y., 2001. Id-1, ITF-2, and Id-2 comprise a network of helix-loop-helix proteins that regulate mammary epithelial cell proliferation, differentiation, and apoptosis. J Biol Chem. 276, 39213-39219.

Pato, C. N., Macedo, A., Ambrosio, A., Vincent, J. B., Bauer, A., Schindler, K., Xu, J., Coelho, I., Dourado, A., Valente, J., Azevedo, M. H., Kennedy, J. L.

and Pato, M. T., 2000. Detection of expansion regions in Portuguese bipolar families. Am J Med Genet. 96, 854-857.

Peippo, M. M., Simola, K. O., Valanne, L. K., Larsen, A. T., Kahkonen, M., Auranen, M. P. and Ignatius, J., 2006. Pitt-Hopkins syndrome in two patients and further definition of the phenotype. Clin Dysmorphol. 15, 47-54.

Perk, J., Iavarone, A. and Benezra, R., 2005. Id family of helix-loop-helix proteins in cancer. Nat Rev Cancer. 5, 603-614.

Persson, P., Jogi, A., Grynfeld, A., Pahlman, 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.

Petropoulos, H. and Skerjanc, I. S., 2000. Analysis of the inhibition of MyoD activity by ITF-2B and full-length E12/E47. J Biol Chem. 275, 25095-25101.

Phillips, J. M. and Goodman, J. I., 2008. Identification of genes that may play critical roles in phenobarbital (PB)-induced liver tumorigenesis due to altered DNA methylation. Toxicol Sci. 104, 86-99.

Pickard, B. S., Malloy, M. P., Clark, L., Lehellard, S., Ewald, H. L., Mors, O., Porteous, D. J., Blackwood, D. H. and Muir, W. J., 2005. Candidate psychiatric illness genes identified in patients with pericentric inversions of chromosome 18. Psychiatr Genet. 15, 37-44.

Pise-Masison, C. A., Radonovich, M., Dohoney, K., Morris, J. C., O'Mahony, D., Lee, M. J., Trepel, J., Waldmann, T. A., Janik, J. E. and Brady, J. N., 2009. Gene expression profiling of ATL patients: compilation of disease-related genes and evidence for TCF4 involvement in BIRC5 gene expression and cell viability. Blood. 113, 4016-4026.

Pitt, D. and Hopkins, I., 1978. A syndrome of mental retardation, wide mouth and intermittent overbreathing. Aust Paediatr J. 14, 182-184.

Plevin, M. J., Mills, M. M. and Ikura, M., 2005. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends Biochem Sci. 30, 66-69.

Polakis, P., 2000. Wnt signaling and cancer. Genes Dev. 14, 1837-1851.

Portman, D. S. and Emmons, S. W., 2000. The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a C. elegans neuronal sublineage. Development. 127, 5415-5426.

Poser, S. and Storm, D. R., 2001. Role of Ca2+-stimulated adenylyl cyclases in LTP and memory formation. Int J Dev Neurosci. 19, 387-394.

Pscherer, A., Dorflinger, U., Kirfel, J., Gawlas, K., Ruschoff, J., Buettner, R. and Schule, R., 1996. The helix-loop-helix transcription factor SEF-2 regulates the activity of a novel initiator element in the promoter of the human somatostatin receptor II gene. EMBO J. 15, 6680-6690.

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.

Qiu, Y., Sharma, A. and Stein, R., 1998. p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. Mol Cell Biol. 18, 2957-2964.

Quednow, B. B., Brinkmeyer, J., Mobascher, A., Nothnagel, M., Musso, F., Grunder, G., Savary, N., Petrovsky, N., Frommann, I., Lennertz, L., Spreckelmeyer, K. N., Wienker, T. F., Dahmen, N., Thuerauf, N., Clepce, M., Kiefer, F., Majic, T., Mossner, R., Maier, W., Gallinat, J., Diaz-Lacava, A., Toliat, M. R., Thiele, H., Nurnberg, P., Wagner, M. and Winterer, G., 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.

Quednow, B. B., Ettinger, U., Mossner, R., Rujescu, D., Giegling, I., Collier, D. A., Schmechtig, A., Kuhn, K. U., Moller, 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.

Quong, M. W., Massari, M. E., Zwart, R. and Murre, C., 1993. A new transcriptional-activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells. Mol Cell Biol. 13, 792-800.

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.

Reizis, B., 2010. Regulation of plasmacytoid dendritic cell development. Curr Opin Immunol. 22, 206-211.

Riazuddin, S. A., McGlumphy, E. J., Yeo, W. S., Wang, J., Katsanis, N. and Gottsch, J. D., 2011. Replication of the TCF4 intronic variant in late-onset Fuchs corneal dystrophy and evidence of independence from the FCD2 locus. Invest Ophthalmol Vis Sci. 52, 2825-2829.

Riazuddin, S. A., Parker, D. S., McGlumphy, E. J., Oh, E. C., Iliff, B. W., Schmedt, T., Jurkunas, U., Schleif, R., Katsanis, N. and Gottsch, J. D., 2012. Mutations in LOXHD1, a recessive-deafness locus, cause dominant late-onset Fuchs corneal dystrophy. Am J Hum Genet. 90, 533-539.

Riazuddin, S. A., Zaghloul, N. A., Al-Saif, A., Davey, L., Diplas, B. H., Meadows, D. N., Eghrari, A. O., Minear, M. A., Li, Y. J., Klintworth, G. K., Afshari, N., Gregory, S. G., Gottsch, J. D. and Katsanis, N., 2010. Missense mutations in TCF8 cause late-onset Fuchs corneal dystrophy and interact with FCD4 on chromosome 9p. Am J Hum Genet. 86, 45-53.

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., St Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D. H., Pato, C. N., Malhotra, A. K., Purcell, S., Dudbridge, F., Neale, B. M., Rossin, L., Visscher, P. M., Posthuma, D., Ruderfer, D. M., Fanous, A., Stefansson, H., Steinberg, S., Mowry, B. J., Golimbet, V., De Hert, M., Jonsson, E. G., Bitter, I., Pietilainen, O. P., Collier, D. A., Tosato,

S., Agartz, I., Albus, M., Alexander, M., Amdur, R. L., Amin, F., Bass, N., Bergen, S. E., Black, D. W., Borglum, A. D., Brown, M. A., Bruggeman, R., Buccola, N. G., Byerley, W. F., Cahn, W., Cantor, R. M., Carr, V. J., Catts, S. V., Choudhury, K., Cloninger, C. R., Cormican, P., Craddock, N., Danoy, P. A., Datta, S., de Haan, L., Demontis, D., Dikeos, D., Djurovic, S., Donnelly, P., Donohoe, G., Duong, L., Dwyer, S., Fink-Jensen, A., Freedman, R., Freimer, N. B., Friedl, M., Georgieva, L., Giegling, I., Gill, M., Glenthoj, B., Godard, S., Hamshere, M., Hansen, M., Hansen, T., Hartmann, A. M., Henskens, F. A., Hougaard, D. M., Hultman, C. M., Ingason, A., Jablensky, A. V., Jakobsen, K. D., Jay, M., Jurgens, G., Kahn, R. S., Keller, M. C., Kenis, G., Kenny, E., Kim, Y., Kirov, G. K., Konnerth, H., Konte, B., Krabbendam, L., Krasucki, R., Lasseter, V. K., Laurent, C., Lawrence, J., Lencz, T., Lerer, F. B., Liang, K. Y., Lichtenstein, P., Lieberman, J. A., Linszen, D. H., Lonnqvist, J., Loughland, C. M., Maclean, A. W., Maher, B. S., Maier, W., Mallet, J., Malloy, P., Mattheisen, M., Mattingsdal, M., McGhee, K. A., McGrath, J. J., McIntosh, A., McLean, D. E., McQuillin, A., Melle, I., Michie, P. T., Milanova, V., Morris, D. W., Mors, O., Mortensen, P. B., Moskvina, V., Muglia, P., Myin-Germeys, I., Nertney, D. A., Nestadt, G., Nielsen, J., Nikolov, I., Nordentoft, M., Norton, N., Nothen, M. M., O'Dushlaine, C. T., Olincy, A., Olsen, L., O'Neill, F. A., Orntoft, T. F., Owen, M. J., Pantelis, C., Papadimitriou, G., Pato, M. T., Peltonen, L., Petursson, H., Pickard, B., Pimm, J., Pulver, A. E., Puri, V., Quested, D., Quinn, E. M., Rasmussen, H. B., Rethelyi, J. M., Ribble, R., Rietschel, M., Riley, B. P., Ruggeri, M., Schall, U., Schulze, T. G., Schwab, S. G., Scott, R. J., Shi, J., Sigurdsson, E., Silverman, J. M., Spencer, C. C., Stefansson, K., Strange, A., Strengman, E., Stroup, T. S., Suvisaari, J., Terenius, L., Thirumalai, S., Thygesen, J. H., Timm, S., Toncheva, D., van den Oord, E., van Os, J., van Winkel, R., Veldink, J., Walsh, D., Wang, A. G., Wiersma, D., Wildenauer, D. B., Williams, H. J., Williams, N. M., Wormley, B., Zammit, S., Sullivan, P. F., O'Donovan, M. C., Daly, M. J. and Gejman, P. V., 2011. Genome-wide association study identifies five new schizophrenia loci. Nat Genet. 43, 969-976.

Robbins, S. H., Walzer, T., Dembele, D., Thibault, C., Defays, A., Bessou, G., Xu, H., Vivier, E., Sellars, M., Pierre, P., Sharp, F. R., Chan, S., Kastner, P. and Dalod, M., 2008. Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. Genome Biol. 9, R17.

Robinson, K. A. and Lopes, J. M., 2000. SURVEY AND SUMMARY: Saccharomyces cerevisiae basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Res. 28, 1499-1505.

Rosenfeld, J. A., Leppig, K., Ballif, B. C., Thiese, H., Erdie-Lalena, C., Bawle, E., Sastry, S., Spence, J. E., Bandholz, A., Surti, U., Zonana, J., Keller, K., Meschino, W., Bejjani, B. A., Torchia, B. S. and Shaffer, L. G., 2009. Genotype-phenotype analysis of TCF4 mutations causing Pitt-Hopkins

syndrome shows increased seizure activity with missense mutations. Genet Med. 11, 797-805.

Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P. and Vidal, M., 2005. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 437, 1173-1178.

Saarikettu, J., Sveshnikova, N. and Grundstrom, T., 2004. Calcium/calmodulin inhibition of transcriptional activity of E-proteins by prevention of their binding to DNA. J Biol Chem. 279, 41004-41011.

Schmedt, T., Silva, M. M., Ziaei, A. and Jurkunas, U., 2012. Molecular bases of corneal endothelial dystrophies. Exp Eye Res. 95, 24-34.

Schuh, A. H., Tipping, A. J., Clark, A. J., Hamlett, I., Guyot, B., Iborra, F. J., Rodriguez, P., Strouboulis, J., Enver, T., Vyas, P. and Porcher, C., 2005. ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. Mol Cell Biol. 25, 10235-10250.

Scott, I. C., Anson-Cartwright, L., Riley, P., Reda, D. and Cross, J. C., 2000. The HAND1 basic helix-loop-helix transcription factor regulates trophoblast differentiation via multiple mechanisms. Mol Cell Biol. 20, 530-541.

Shain, D. H., Neuman, T. and Zuber, M. X., 1997. Embryonic expression and evolution of duplicated E-protein genes in Xenopus laevis: parallels with ancestral E-protein genes. Genetics. 146, 345-353.

Shain, D. H. and Zuber, M. X., 1995. Identification of non-tissue-specific helix-loop-helix genes in Xenopus laevis. Gene. 165, 319-320.

Sheridan, C. M., Heist, E. K., Beals, C. R., Crabtree, G. R. and Gardner, P., 2002. Protein kinase A negatively modulates the nuclear accumulation of NF-ATc1 by priming for subsequent phosphorylation by glycogen synthase kinase-3. J Biol Chem. 277, 48664-48676.

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.

Sigvardsson, M., 2000. Overlapping expression of early B-cell factor and basic helix-loop-helix proteins as a mechanism to dictate B-lineage-specific activity of the lambda5 promoter. Mol Cell Biol. 20, 3640-3654.

Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B. M. and Vervoort, M., 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol Biol. 7, 33.

Singh, H. A., 1993. Mental retardation, macrostomia and hyperpnoea syndrome. J Paediatr Child Health. 29, 156-157.

Singh, S. and Gramolini, A. O., 2009. Characterization of sequences in human TWIST required for nuclear localization. BMC Cell Biol. 10, 47.

Skerjanc, I. S., Truong, J., Filion, P. and McBurney, M. W., 1996. A splice variant of the ITF-2 transcript encodes a transcription factor that inhibits MyoD activity. J Biol Chem. 271, 3555-3561.

Skinner, M. K., Rawls, A., Wilson-Rawls, J. and Roalson, E. H., 2010. Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature. Differentiation. 80, 1-8.

Sloan, S. R., Shen, C. P., McCarrick-Walmsley, R. and Kadesch, T., 1996. Phosphorylation of E47 as a potential determinant of B-cell-specific activity. Mol Cell Biol. 16, 6900-6908.

Smrt, R. D., Szulwach, K. E., Pfeiffer, R. L., Li, X., Guo, W., Pathania, M., Teng, Z. Q., Luo, Y., Peng, J., Bordey, A., Jin, P. and Zhao, X., 2010. MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. Stem Cells. 28, 1060-1070.

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.

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. Brain Res Mol Brain Res. 25, 176-180.

Szulwach, K. E., Li, X., Smrt, R. D., Li, Y., Luo, Y., Lin, L., Santistevan, N. J., Li, W., Zhao, X. and Jin, P., 2010. Cross talk between microRNA and epigenetic regulation in adult neurogenesis. J Cell Biol. 189, 127-141.

Stamler, J. F., Roos, B. R., Wagoner, M. D., Goins, K. M., Kitzmann, A. S., Riley, J. B., Stone, E. M. and Fingert, J. H., in press. Confirmation of the association between the TCF4 risk allele and Fuchs endothelial corneal dystrophy in patients from the Midwestern United States. Ophthalmic Genet.

Stavropoulos, D. J., MacGregor, D. L. and Yoon, G., 2011. Mosaic microdeletion 18q21 as a cause of mental retardation. Eur J Med Genet. 53, 396-399.

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O. P., Mors, O., Mortensen, P. B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Suvisaari, J., Lonnqvist, J., Paunio, T., Borglum, A. D., Hartmann, A., Fink-Jensen, A., Nordentoft, M., Hougaard, D., Norgaard-Pedersen, B., Bottcher, Y., Olesen, J., Breuer, R., Moller, H. J., Giegling, I., Rasmussen, H. B., Timm, S., Mattheisen, M., Bitter, I., Rethelyi, J. M., Magnusdottir, B. B., Sigmundsson, T., Olason, P., Masson, G., Gulcher, J. R., Haraldsson, M., Fossdal, R., Thorgeirsson, T. E., Thorsteinsdottir, U., Ruggeri, M., Tosato, S., Franke, B., Strengman, E., Kiemeney, L. A., Melle, I., Djurovic, S., Abramova, L., Kaleda, V., Sanjuan, J., de Frutos, R., Bramon, E., Vassos, E., Fraser, G., Ettinger, U., Picchioni, M., Walker, N., Toulopoulou, T., Need, A. C., Ge, D., Yoon, J. L., Shianna, K. V., Freimer, N. B., Cantor, R. M., Murray, R., Kong, A., Golimbet, V., Carracedo, A.,

Arango, C., Costas, J., Jonsson, E. G., Terenius, L., Agartz, I., Petursson, H., Nothen, M. M., Rietschel, M., Matthews, P. M., Muglia, P., Peltonen, L., St Clair, D., Goldstein, D. B., Stefansson, K. and Collier, D. A., 2009. Common variants conferring risk of schizophrenia. Nature. 460, 744-747.

Steinberg, S., de Jong, S., Andreassen, O. A., Werge, T., Borglum, A. D., Mors, O., Mortensen, P. B., Gustafsson, O., Costas, J., Pietilainen, O. P., Demontis, D., Papiol, S., Huttenlocher, J., Mattheisen, M., Breuer, R., Vassos, E., Giegling, I., Fraser, G., Walker, N., Tuulio-Henriksson, A., Suvisaari, J., Lonnqvist, J., Paunio, T., Agartz, I., Melle, I., Djurovic, S., Strengman, E., Jurgens, G., Glenthoj, B., Terenius, L., Hougaard, D. M., Orntoft, T., Wiuf, C., Didriksen, M., Hollegaard, M. V., Nordentoft, M., van Winkel, R., Kenis, G., Abramova, L., Kaleda, V., Arrojo, M., Sanjuan, J., Arango, C., Sperling, S., Rossner, M., Ribolsi, M., Magni, V., Siracusano, A., Christiansen, C., Kiemeney, L. A., Veldink, J., van den Berg, L., Ingason, A., Muglia, P., Murray, R., Nothen, M. M., Sigurdsson, E., Petursson, H., Thorsteinsdottir, U., Kong, A., Rubino, I. A., De Hert, M., Rethelyi, J. M., Bitter, I., Jonsson, E. G., Golimbet, V., Carracedo, A., Ehrenreich, H., Craddock, N., Owen, M. J., O'Donovan, M. C., Ruggeri, M., Tosato, S., Peltonen, L., Ophoff, R. A., Collier, D. A., St Clair, D., Rietschel, M., Cichon, S., Stefansson, H., Rujescu, D. and Stefansson, K., 2011. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. Hum Mol Genet. 20, 4076-4081.

Steinbusch, C. V., van Roozendaal, K. E., Tserpelis, D., Smeets, E. E., Kranenburg-de Koning, T. J., de Waal, K. H., Zweier, C., Rauch, A., Hennekam, R. C., Blok, M. J. and Schrander-Stumpel, C. T., in press. Somatic mosaicism in a mother of two children with Pitt-Hopkins syndrome. Clin Genet.

Stevens, J. D., Roalson, E. H. and Skinner, M. K., 2008. Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: genomic approach to cellular differentiation. Differentiation. 76, 1006-1022.

Sullivan, P. F., Kendler, K. S. and Neale, M. C., 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 60, 1187-1192.

Sun, X. H., 2004. Multitasking of helix-loop-helix proteins in lymphopoiesis. Adv Immunol. 84, 43-77.

Sundin, O. H., Broman, K. W., Chang, H. H., Vito, E. C., 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.

Suzuki, H., Fukunishi, Y., Kagawa, I., Saito, R., Oda, H., Endo, T., Kondo, S., Bono, H., Okazaki, Y. and Hayashizaki, Y., 2001. Protein-protein interaction panel using mouse full-length cDNAs. Genome Res. 11, 1758-1765.

Svec, J., Ergang, P., Mandys, V., Kment, M. and Pacha, J., 2009. Expression profiles of proliferative and antiapoptotic genes in sporadic and colitis-related mouse colon cancer models. Int J Exp Pathol. 91, 44-53.

Svec, J., Musilkova, J., Bryndova, J., Jirasek, T., Mandys, V., Kment, M. and Pacha, J., 2010. Enhanced expression of proproliferative and antiapoptotic genes in ulcerative colitis-associated neoplasia. Inflamm Bowel Dis. 16, 1127-1137.

Taddeucci, G., Bonuccelli, A., Mantellassi, I., Orsini, A. and Tarantino, E., 2010. Pitt-Hopkins syndrome: report of a case with a TCF4 gene mutation. Ital J Pediatr. 36, 12.

Takano, K., Lyons, M., Moyes, C., Jones, J. and Schwartz, C., 2010. Two percent of patients suspected of having Angelman syndrome have TCF4 mutations. Clin Genet. 78, 282-288.

Takano, K., Tan, W. H., Irons, M. B., Jones, J. R. and Schwartz, C. E., 2011. Pitt-Hopkins syndrome should be in the differential diagnosis for males presenting with an ATR-X phenotype. Clin Genet. 80, 600-601.

Takenouchi, T., Yagihashi, T., Tsuchiya, H., Torii, C., Hayashi, K., Kosaki, R., Saitoh, S., Takahashi, T. and Kosaki, K., 2012. Tissue-limited ring chromosome 18 mosaicism as a cause of Pitt-Hopkins syndrome. Am J Med Genet A. 158A, 2621-2623.

Talkowski, M. E., Rosenfeld, J. A., Blumenthal, I., Pillalamarri, V., Chiang, C., Heilbut, A., Ernst, C., Hanscom, C., Rossin, E., Lindgren, A. M., Pereira, S., Ruderfer, D., Kirby, A., Ripke, S., Harris, D. J., Lee, J. H., Ha, K., Kim, H. G., Solomon, B. D., Gropman, A. L., Lucente, D., Sims, K., Ohsumi, T. K., Borowsky, M. L., Loranger, S., Quade, B., Lage, K., Miles, J., Wu, B. L., Shen, Y., Neale, B., Shaffer, L. G., Daly, M. J., Morton, C. C. and Gusella, J. F., 2012. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell. 149, 525-537.

Tanaka, A., Itoh, F., Itoh, S. and Kato, M., 2009. TAL1/SCL relieves the E2-2-mediated repression of VEGFR2 promoter activity. J Biochem. 145, 129-135.

Tanaka, A., Itoh, F., Nishiyama, K., Takezawa, T., Kurihara, H., Itoh, S. and Kato, M., 2010. Inhibition of endothelial cell activation by bHLH protein E2-2 and its impairment of angiogenesis. Blood. 115, 4138-4147.

Thalamuthu, A., Khor, C. C., Venkataraman, D., Koh, L. W., Tan, D. T., Aung, T., Mehta, J. S. and Vithana, E. N., 2011. Association of TCF4 gene polymorphisms with Fuchs' corneal dystrophy in the Chinese. Invest Ophthalmol Vis Sci. 52, 5573-5578.

Thomas, C., Dankesreiter, A., Wolf, H. and Schwarzmann, F., 2003. The BZLF1 promoter of Epstein-Barr virus is controlled by E box-/HI-motif-binding factors during virus latency. J Gen Virol. 84, 959-964.

Uittenbogaard, M. and Chiaramello, A., 1999. Expression of the basic Helix-Loop-Helix ME1 E-protein during development and aging of the murine cerebellum. Neurosci Lett. 274, 191-194.

Uittenbogaard, M. and Chiaramello, A., 2000. Differential expression patterns of the basic helix-loop-helix transcription factors during aging of the murine brain. Neurosci Lett. 280, 95-98.

Van Balkom, I. D., Quartel, S. and Hennekam, R. C., 1998. Mental retardation, "coarse" face, and hyperbreathing: confirmation of the Pitt-Hopkins syndrome. Am J Med Genet. 75, 273-276.

Van Balkom, I. D. C., Vuijk, P. J., Franssens, M., Hoek, H. W. and Hennekam, R. C., 2012. Development, cognition, and behaviour in Pitt-Hopkins syndrome. Dev Med Child Neurol. 54, 925-931.

Verheyen, G. R., Del-Favero, J., Mendlewicz, J., Lindblad, K., Van Zand, K., Aalbregtse, M., Schalling, M., Souery, D. and Van Broeckhoven, C., 1999. Molecular interpretation of expanded RED products in bipolar disorder by CAG/CTG repeats located at chromosomes 17q and 18q. Neurobiol Dis. 6, 424-432.

Verhulst, S. L., De Dooy, J., Ramet, J., Bockaert, N., Van Coster, R., Ceulemans, B. and De Backer, W., 2012. Acetazolamide for severe apnea in Pitt-Hopkins syndrome. Am J Med Genet A. 158A, 932-934.

Verma-Gaur, J., Hauser, J. and Grundstrom, T., 2012. Negative feedback regulation of antigen receptors through calmodulin inhibition of E2A. J Immunol. 188, 6175-6183.

Vicent, S., Luis-Ravelo, D., Anton, I., Garcia-Tunon, I., Borras-Cuesta, F., Dotor, J., De Las Rivas, J. and Lecanda, F., 2008. A novel lung cancer signature mediates metastatic bone colonization by a dual mechanism. Cancer Res. 68, 2275-2285.

Vincent, J. B., Paterson, A. D., Strong, E., Petronis, A. and Kennedy, J. L., 2000. The unstable trinucleotide repeat story of major psychosis. Am J Med Genet. 97, 77-97.

Vincent, J. B., Petronis, A., Strong, E., Parikh, S. V., Meltzer, H. Y., Lieberman, J. and Kennedy, J. L., 1999. Analysis of genome-wide CAG/CTG repeats, and at SEF2-1B and ERDA1 in schizophrenia and bipolar affective disorder. Mol Psychiatry. 4, 229-234.

Voronova, A. and Baltimore, D., 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proc Natl Acad Sci U S A. 87, 4722-4726.

Wang, X., Tang, X., Li, M., Marshall, J. and Mao, Z., 2005. Regulation of neuroprotective activity of myocyte-enhancer factor 2 by cAMP-protein kinase A signaling pathway in neuronal survival. J Biol Chem. 280, 16705-16713.

Watabe, T., 2011. Roles of old players in the suppression of a new player: networks for the transcriptional control of angiogenesis. J Biochem. 149, 117-119.

Wayman, G. A., Tokumitsu, H. and Soderling, T. R., 1997. Inhibitory cross-talk by cAMP kinase on the calmodulin-dependent protein kinase cascade. J Biol Chem. 272, 16073-16076.

West, A. E. and Greenberg, M. E., 2011. Neuronal activity-regulated gene transcription in synapse development and cognitive function. Cold Spring Harb Perspect Biol. 3, a005744.

Whalen, S., Heron, 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., Aftimos, S., Mancini, G., Dias, P., Philip, N., Goldenberg, A., Le Merrer, M., Rio, M., Josifova, D., Van Hagen, J. M., Lacombe, D., Edery, P., Dupuis-Girod, S., Putoux, A., Sanlaville, D., Fischer, R., Drevillon, L., Briand-Suleau, A., Metay, C., Goossens, M., Amiel, J., Jacquette, A. and Giurgea, I., 2012. Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: Clinical score and further delineation of the TCF4 mutational spectrum. Hum Mutat. 33, 64-72.

Wheeler, D. G. and Cooper, E., 2001. Depolarization strongly induces human cytomegalovirus major immediate-early promoter/enhancer activity in neurons. J Biol Chem. 276, 31978-31985.

Wigle, D., Ho, W., Lo, D., Francis, J., Eubanks, J. H. and Wallace, M. C., 1999. Altered expression levels of SEF-2 and p112 in the rat hippocampus after transient cerebral ischemia: identification by mRNA differential display. J Cereb Blood Flow Metab. 19, 435-442.

Wikstrom, I., Forssell, J., Goncalves, M., Colucci, F. and Holmberg, D., 2006. E2-2 regulates the expansion of pro-B cells and follicular versus marginal zone decisions. J Immunol. 177, 6723-6729.

Wikstrom, I., Forssell, J., Penha-Goncalves, M. N., Bergqvist, I. and Holmberg, D., 2008. A role for E2-2 at the DN3 stage of early thymopoiesis. Mol Immunol. 45, 3302-3311.

Wilke, S. A., Hall, B. J., Antonios, J. K., Denardo, L. A., Otto, S., Yuan, B., Chen, F., Robbins, E. M., Tiglio, K., Williams, M. E., Qiu, Z., Biederer, T. and Ghosh, A., 2012. NeuroD2 regulates the development of hippocampal mossy fiber synapses. Neural Dev. 7, 9.

Willemsen, M. H., Rensen, J. H., van Schrojenstein-Lantman de Valk, H. M., Hamel, B. C. and Kleefstra, T., 2012. Adult Phenotypes in Angelman- and Rett-Like Syndromes. Mol Syndromol. 2, 217-234.

Williams, H. J., Moskvina, V., Smith, R. L., Dwyer, S., Russo, G., Owen, M. J. and O'Donovan, M. C., 2011a. Association between TCF4 and schizophrenia does not exert its effect by common nonsynonymous variation or by influencing cis-acting regulation of mRNA expression in adult human brain. Am J Med Genet B Neuropsychiatr Genet. 156B, 781-784.

Williams, H. J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., Georgieva, L., Williams, N. M., Morris, D. W., Quinn, E. M., Giegling, I., Ikeda, M., Wood, J., Lencz, T., Hultman, C., Lichtenstein, P., Thiselton, D., Maher, B. S., Malhotra, A. K., Riley, B., Kendler, K. S., Gill, M., Sullivan, P., Sklar, P., Purcell, S., Nimgaonkar, V. L., Kirov, G., Holmans, P., Corvin, A., Rujescu, D., Craddock, N., Owen, M. J. and O'Donovan, M. C., 2011b. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. Mol Psychiatry. 16, 429-441.

Wirgenes, K. V., Sonderby, 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.

Wong, M. C., Castanon, I. and Baylies, M. K., 2008. Daughterless dictates Twist activity in a context-dependent manner during somatic myogenesis. Dev Biol. 317, 417-429.

Wright, A. F. and Dhillon, B., 2010. Major progress in Fuchs's corneal dystrophy. N Engl J Med. 363, 1072-1075.

Xiang, X., Zhuang, X., Ju, S., Zhang, S., Jiang, H., Mu, J., Zhang, L., Miller, D., Grizzle, W. and Zhang, H. G., 2011. miR-155 promotes macroscopic tumor formation yet inhibits tumor dissemination from mammary fat pads to the lung by preventing EMT. Oncogene. 30, 3440-3453.

Yang, W., Itoh, F., Ohya, H., Kishimoto, F., Tanaka, A., Nakano, N., Itoh, S. and Kato, M., 2011. Interference of E2-2-mediated effect in endothelial cells by FAM96B through its limited expression of E2-2. Cancer Sci. 102, 1808-1814.

Ye, T., Lipska, B. K., Tao, R., Hyde, T. M., Wang, L., Li, C., Choi, K. H., Straub, R. E., Kleinman, J. E. and Weinberger, D. R., 2012. Analysis of copy number variations in brain DNA from patients with schizophrenia and other psychiatric disorders. Biol Psychiatry. 72, 651-654.

Yoder, J. A., Walsh, C. P. and Bestor, T. H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13, 335-340.

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.

Yoshida, S., Ohbo, K., Takakura, A., Takebayashi, H., Okada, T., Abe, K. and Nabeshima, Y., 2001. Sgn1, a basic helix-loop-helix transcription factor delineates the salivary gland duct cell lineage in mice. Dev Biol. 240, 517-530.

Yue, W. H., Wang, H. F., Sun, L. D., Tang, F. L., Liu, Z. H., Zhang, H. X., Li, W. Q., Zhang, Y. L., Zhang, Y., Ma, C. C., Du, B., Wang, L. F., Ren, Y. Q., Yang, Y. F., Hu, X. F., Wang, Y., Deng, W., Tan, L. W., Tan, Y. L., Chen, Q., Xu, G. M., Yang, G. G., Zuo, X. B., Yan, H., Ruan, Y. Y., Lu, T. L., Han, X., Ma, X. H., Cai, L. W., Jin, C., Zhang, H. Y., Yan, J., Mi, W. F., Yin, X. Y., Ma, W. B., Liu, Q., Kang, L., Sun, W., Pan, C. Y., Shuang, M., Yang, F. D., Wang, C. Y., Yang, J. L., Li, K. Q., Ma, X., Li, L. J., Yu, X., Li, Q. Z., Huang, X., Lv, L. X., Li, T., Zhao, G. P., Huang, W., Zhang, X. J. and Zhang, D., 2011. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. Nat Genet. 43, 1228-1231.

Zamudio, N. and Bourc'his, D., 2010. Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity. 105, 92-104.

Zarifi, I., Kiparaki, M., Koumpanakis, K. A., Giagtzoglou, N., Zacharioudaki, E., Alexiadis, T., Livadaras, I. and Delidakis, C., 2012. Essential roles of Da transactivation domains in neurogenesis and in E(spl)-mediated repression. Mol Cell Biol. 32, 4534-4548.

Zhai, Y., Wu, R., Schwartz, D. R., Darrah, D., Reed, H., Kolligs, F. T., Nieman, M. T., Fearon, E. R. and Cho, K. R., 2002. Role of beta-catenin/T-cell factor-regulated genes in ovarian endometrioid adenocarcinomas. Am J Pathol. 160, 1229-1238.

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.

Zhang, T., Bai, Y. Z., Wang da, J., Jia, H. M., Yuan, Z. W. and Wang, W. L., 2009. Spatiotemporal pattern analysis of transcription factor 4 in the developing anorectum of the rat embryo with anorectal malformations. Int J Colorectal Dis. 24, 1039-1047.

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.

Zweier, C., de Jong, E. K., Zweier, M., Orrico, A., Ousager, L. B., Collins, A. L., Bijlsma, E. K., Oortveld, M. A., Ekici, A. B., Reis, A., Schenck, A. and Rauch, A., 2009. CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. Am J Hum Genet. 85, 655-666.

Zweier, C., Peippo, M. M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., Reardon, W., Saraiva, J., Cabral, A., Gohring, I., Devriendt, K., de Ravel, T., Bijlsma, E. K., Hennekam, R. C., Orrico, A., Cohen, M., Dreweke, A., Reis, A., Nurnberg, P. and Rauch, A., 2007. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am J Hum Genet. 80, 994-1001.

Zweier, C., Sticht, H., Bijlsma, E. K., Clayton-Smith, J., Boonen, S. E., Fryer, A., Greally, M. T., Hoffmann, L., den Hollander, N. S., Jongmans, M., Kant, S. G., King, M. D., Lynch, S. A., McKee, S., Midro, A. T., Park, S. M., Ricotti, V., Tarantino, E., Wessels, M., Peippo, M. and Rauch, A., 2008. Further delineation of Pitt-Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. J Med Genet. 45, 738-744.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all the people who have helped, encouraged and supported me during my studies. It took me a while to get to the point of writing this page and during this time I have had an opportunity to learn from and work together with many wonderful people.

First and foremost, I am deeply thankful to my supervisor prof. Tõnis Timmusk for his guidance, patience and ever-lasting optimism. I appreciate the chance he gave me to make my own decisions, mistakes and discoveries.

It is difficult to overstate my gratitude to my previous supervisors Reet Kurg, Aare Abroi and Petri Auvinen for sharing their experience and knowledge. I am also thankful to many other people who have taught me laboratory techniques and ways of critical thinking at the Institute of Molecular and Cell Biology in University of Tartu, at the Institute of Biotechnology in University of Helsinki, at the Department of Gene Technology in Tallinn University of Technology, and at prof. Elena Cattaneo's lab in University of Milan. I am especially thankful to prof. Andres Merits, prof. Mart Saarma, prof. Mart Ustav, Heiti Paves, Mart Speek, Marzia Tartari and Donato Goffredo.

I thank prof. Toomas Neuman, the co-discoverer of *Tcf4* gene, for critically reading the manuscript of this thesis and providing initial constructs when the TCF4 project was started in Tallinn.

I am thankful to all the co-authors and collaborators of my publications.

My special thanks go to all former and present members of Neuro-lab for the great time we have spent together in and out of lab. To Indrek Koppel, Priit Pruunsild, Epp Väli, Marko Piirsoo, Richard Tamme, Kristi Luberg, Hanna Vihma, Kaur Jaanson, Ingrid Lekk, Kati Taal, Maila Rähn, Tamara Aid-Pavlidis, Pavlos Pavlidis, Mari Palgi, Karin Tamm, Mirjam Luhakooder, Elina Aleksejeva, Madis Jaagura, Rahel Park, Jürgen Tuvikene, Angela Pärn, Risto Jamul, Eli-Eelika Esvald, Anne Kaljas, Anna Kazantseva, Anželika Sarapuu, Ester Orav, Martin Kask, Grete Rullinkov and Helen Verev. My very special thanks to Ave Eesmaa, Kaja Kannike, Mari Urb and Jaan Toots for keeping me motivated.

Thanks also to the people from the neighbouring labs, especially Alla Piirsoo, Lagle Kasak, Piret Laht, Pille Pata, Jekaterina Kazantseva, Teet Velling, Kaia Palm and Kairit Tints. Particular thanks to Katrina Laks and Tiina Kirsipuu for making the office-hours most enjoyable. I appreciate the help of Ants Uus, Tatjana Tanvel, Epp Ints, Annely Rätsep and Tiiu Tammsalu in keeping things running.

I am grateful to my coursemates Pirjo Spuul, Nele Tamberg and Kersti Jääger for the inspiring "critical" period when we first started doing science in lab. I cordially thank all my friends who have stood by me physically and emotionally throughout the years. I am indebted to Kaie and Toomas for providing me a friendly place to stay when I needed it the most.

My greatest gratitude to my parents Milvi and Margus, to my brothers Sander and Silver and to my sister Katrina for their care, understanding and unconditional belief in me.

And last but certainly not least, my warmest thanks to Priit. *Tuhat tänu!*

KOKKUVÕTE

Transkriptsioonitegur TCF4 (tuntud ka kui ITF2, SEF2 ja E2-2) on laialdaselt avalduv aluselise heeliks-ling-heeliks perekonna valk, mis seondub CANNTG DNA elemendile homo- või heterodimeerina. TCF4 osaleb paljude erinevate rakutüüpide arengus ja talitluses, aga hiljutistest uuringutest lähtuvalt on tal eriti tähtis roll täita närvisüsteemis. Varieeruvus *TCF4* geenis on seotud skisofreeniaga ning deletsioonid, lugemisraami kärpivad ja missens-mutatsioonid *TCF4* geenis põhjustavad Pitt-Hopkinsi sündroomi (PTHS), mida iseloomustab sügav vaimne ja motoorne alaareng, tüüpilised näojooned ja hingamishälbed.

Käesoleva töö eesmärgiks oli (i) kirjeldada inimese *TCF4* geeni struktuuri ja avaldumist mRNA ja valgu tasemel, (ii) uurida TCF4 valgu isovormide talitlemist mitteneuraalsetes rakkudes ja närvirakkudes ning (iii) teha kindlaks, millist mõju avaldavad PTHSi põhjustavad mutatsioonid TCF4 valgu talitlemisele.

Antud töö tulemused näitavad, et kuigi *TCF4* geenile on iseloomulik lai avaldumismuster, on *TCF4* transkriptide tasemed kõige kõrgemad närvisüsteemis. Arvukate 5' eksonite olemasolu *TCF4* geenis võimaldab toota 18 erineva aminoterminusega TCF4 valgu isovorme. Käesoleva töö tulemuste järgi on TCF4 isovormide rakusisene paiknemine erinevalt reguleeritud. Nimelt sisaldavad osad isovormid tuuma paiknemise signaali (NLS), samas kui teised peavad tuuma sisenemiseks heterodimeriseeruma NLSi sisaldavate partnervalkudega. Lisaks sellele on TCF4 isovormide võime aktiveerida transkriptsiooni varieeruv ning sõltuv raku tüübist. Kui mitteneuraalsetes rakkudes on TCF4 vaikimisi aktiivne, siis närvirakkudes on TCF4-sõltuvaks transkriptsiooniks vajalik kaltsiumi sissevool rakku läbi L-tüüpi pingest sõltuvate kaltsiumi kanalite ning valgu kinaasi A (PKA) poolne signaalimine. Antud uurimistööst lähtuvalt on TCF4 otsene PKA sihtmärk ning PKA fosforüleerimiskohad TCF4 valgus vajalikud tõhusaks TCF4-sõltuvaks transkriptsiooniks närvirakkudes. Lisaks, endogeenset TCF4 valku modifitseeritakse vastusena membraani depolariseerimisele närvirakkudes. PTHSiga seotud mutatsioonide analüüsist selgus, et kõik deletsioonid ja lugemisraami kärpivad mutatsioonid ei häiri muteerunud *TCF4* alleeli talitlemist täielikult. Missens ja lugemisraami pikendavate mutatsioonide mõju varieerub vaevu märgatavatest puudustest kuni dominant-negatiivsete efektideni. Viimati nimetatud mutatsioonide hulgast mõned takistavad TCF4 DNAle seondumist ja transkriptsiooni aktiveerimist viisil, mis sõltub spetsiifilisest dimeerist, mõned muudavad valgu ebastabiilseks ja mõned ei põhjusta märgatavaid talitlushäireid.

Kokkuvõttes kirjeldab käesolev töö *TCF4* geeni avaldumise varieeruvust inimese erinevates kudedes ja annab tunnistust TCF4 isovormide talitlemise erinevustest. Lisaks näitavad antud töö tulemused, et närvirakkudes talitleb TCF4 kui närvitalitluse poolt reguleeritud transkriptsioonitegur ning et PTHSiga seotud mutatsioonid mõjutavad *TCF4* talitlemist erinevatel viisidel, mis ulatuvad hüpomorfsetest kuni dominant-negatiivsete efektideni. Need tulemused annavad uusi teadmisi TCF4 talitlemise kohta ning heidavad valgust mehhanismidele, mis võivad olla aluseks *TCF4* geeni seosele vaimse arengu hälbeid põhjustavate haigustega.

PAPER I

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. 2011;6(7):e22138.

OPEN ACCESS Freely available online

PLoS one

Functional Diversity of Human Basic Helix-Loop-Helix Transcription Factor TCF4 Isoforms Generated by Alternative 5′ Exon Usage and Splicing

Mari Sepp, Kaja Kannike, Ave Eesmaa, Mari Urb, Tõnis Timmusk*

Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia

Abstract

Background: Transcription factor 4 (TCF4 alias ITF2, E2-2, ME2 or SEF2) is a ubiquitous class A basic helix-loop-helix protein that binds to E-box DNA sequences (CANNTG). While involved in the development and functioning of many different cell types, recent studies point to important roles for TCF4 in the nervous system. Specifically, human *TCF4* gene is implicated in susceptibility to schizophrenia and *TCF4* haploinsufficiency is the cause of the Pitt-Hopkins mental retardation syndrome. However, the structure, expression and coding potential of the human *TCF4* gene have not been described in detail.

Principal Findings: In the present study we used human tissue samples to characterize human *TCF4* gene structure and *TCF4* expression at mRNA and protein level. We report that although widely expressed, human *TCF4* mRNA expression is particularly high in the brain. We demonstrate that usage of numerous 5′ exons of the human *TCF4* gene potentially yields in TCF4 protein isoforms with 18 different N-termini. In addition, the diversity of isoforms is increased by alternative splicing of several internal exons. For functional characterization of TCF4 isoforms, we overexpressed individual isoforms in cultured human cells. Our analysis revealed that subcellular distribution of TCF4 isoforms is differentially regulated: Some isoforms contain a bipartite nuclear localization signal and are exclusively nuclear, whereas distribution of other isoforms relies on heterodimerization partners. Furthermore, the ability of different TCF4 isoforms to regulate E-box controlled reporter gene transcription is varied depending on whether one or both of the two TCF4 transcription activation domains are present in the protein. Both TCF4 activation domains are able to activate transcription independently, but act synergistically in combination.

Conclusions: Altogether, in this study we have described the inter-tissue variability of TCF4 expression in human and provided evidence about the functional diversity of the alternative TCF4 protein isoforms.

Citation: Sepp M, Kannike K, Eesmaa A, Urb M, Timmusk T (2011) Functional Diversity of Human Basic Helix-Loop-Helix Transcription Factor TCF4 Isoforms Generated by Alternative 5′ Exon Usage and Splicing. PLoS ONE 6(7): e22138. doi:10.1371/journal.pone.0022138

Editor: Fatah Kashanchi, George Mason University, United States of America

Received February 28, 2011; **Accepted** June 16, 2011; **Published** July 15, 2011

Funding: This work was supported by Estonian Ministry of Education and Research (Grant 0140143) and Estonian Science Foundation (Grants 7257 and 8844). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tonis.timmusk@ttu.ee

Introduction

TCF4 (Gene 6925), alias ITF2 (immunoglobulin transcription factor 2), SEF2 (leukemia virus SL3-3 enhancer factor 2), E2-2 and ME2 (mouse E2), is one of the widely expressed class A basic helix-loop-helix (bHLH) transcription factors (TFs) that are homologous to *Drosophila melanogaster* protein daughterless (Gene 34413) [1,2]. The bHLH factor TCF4 discussed here should not be confused with the high mobility group box transcription factor 7-like 2 (TCF7L2; Gene 6934) that is a downstream effector of the β-catenin signaling pathway and is also known as TCF4 (T-cell specific factor 4).

Class A bHLH factors in mammals include TCF4, HEB (TCF12; Gene 6938) and E2A (TCF3, ITF1; Gene 6929) alternative isoforms E12 and E47 [3]. These proteins are referred to as E-proteins since they bind to Ephrussi box (E-box) sequence (CANNTG) as homodimers or as heterodimers with tissue-specific bHLH factors [3,4]. Dimerization is mediated by the C-terminal HLH motif that together with the preceding stretch of basic amino acids is required also for DNA binding. Structurally related Id proteins (inhibitors of differentiation) hinder DNA binding of E-proteins by heterodimerization, whereas Ca^{2+}-calmodulin specifically inhibits DNA binding of E-protein homodimers [5–7]. Three amino-terminally distinct TCF4 isoforms have been described – TCF4-A, TCF4-B and TCF4-D [2]. All these isoforms contain the bHLH domain and a transcription activation domain (AD2) [8]. TCF4-B has an additional transcription activation domain in its N-terminus (AD1) [9].

In Drosophila the only E-protein, daughterless, is involved in sex determination and neurogenesis [10,11]. In mammals, substantial functional overlap among E-proteins has hampered deciphering their exact roles. However, it is known that TCF4 is required for postnatal survival in mice [12,13] and has many cell lineage specific functions. For instance, TCF4 regulates development of B-, T- and plasmacytoid dendritic cells [13–15], development of Sertoli cells [16] and pontine nucleus neurons [17], myogenesis [18], melanogenesis [19] and epithelial-mesenchymal transition [20]. The importance of TCF4 in human nervous system development is underscored by the association of a TCF4 allele with schizophrenia [21] and identification of *TCF4*

haploinsufficiency as the cause for Pitt-Hopkins syndrome (OMIM 610954), a rare disease featuring mental retardation, hyperventilation and seizures [22–24].

In this study we show that *TCF4* is widely, but not equally expressed and its levels are particularly high in the nervous system. We demonstrate that usage of alternative 5′ exons for transcribing the human *TCF4* gene potentially yields in numerous TCF4 protein isoforms that differ in their subcellular localization and capacity to activate transcription.

Results

TCF4 gene contains many mutually exclusive 5′ exons

To describe the structure and alternative splicing of the human *TCF4* gene we performed bioinformatic analysis of mRNA and expressed sequence tag (EST) sequences available in public databases and sequences of RT-PCR products from this study. In estimation of transcription start sites we relied on publicly available data from sequencing of oligo-cap, cap-trapping and SMART cDNA libraries [25–28]. *TCF4* gene is located on chromosome 18q21.2 and spans 437 kbs. It has 41 exons of which 21 are alternative 5′ exons situated at various positions throughout the gene (Figure 1). In this study *TCF4* exons are named as follows: initial 5′ exons are designated with a lowercase letter preceded by a number that shows the following internal exon in the gene; internal exons are numbered from 1 to 20; exon 21 is the only terminal 3′ exon. *TCF4* transcripts are named according to the initial exon they contain. Initial exons 1a and 1b are located upstream of internal exon 1. Exons 3a–3d precede and 4a–4c follow internal exon 3. 4c is a 5′ extension of internal exon 4. There are three 5′ exons (5a–5c) in front of exon 5; two (7a, 7b) in front of exon 7; four (8a–8d) in front of exon 8; and three (10a–10c) in front of exon 10. Generally, 5′ exons are spliced together with the next internal exon in the gene, apart from the cases when the following internal exon is a cassette exon and can be skipped (exons 1 or 2, exon 3, exons 8–9; see below). As an exception 5′ exon 1a is never used together with internal exons 1 or 2 and is always joined to internal exon 3 or 4. We identified several alternative splice donor sites in the 5′ exons: two (I–II) in exons 5a, 8b and 8c; three (I–III) in exons 4a and 7a. In case of exon 7b the intron between 7b-I and internal exon 7 can be retained giving rise to 7b-II transcripts. Different usage of splice sites sometimes affects the coding potential of a transcript (Figure 1C). In addition, the reading frame shifts with alternative splicing of internal cassette exons 1 or 2 and exon 3. Examination of open reading frames and search for possible translation start codons demonstrated that, altogether, *TCF4* transcripts potentially code for 18 N-terminally distinct protein isoforms named TCF4-A – TCF4-R (Figures 1C and S1) of which only isoforms TCF4-A, TCF4-B and TCF4-D have been described previously as SEF2-1A or ITF2-A, SEF2-1B or ITF2-B, and SEF2-1D, respectively [2,18]. Most of the TCF4 isoforms have a stretch of unique amino acids in their N-termini and the whole aminoterminal transactivation domain AD1 (coded by exons 3–6) is present only in isoforms with the longest N-termini such as TCF4-J, -K, -L and TCF4-B (Figures 1C and S1). Other TCF4 isoforms contain parts of AD1 or are completely devoid of it. More precisely, isoforms TCF4-C, -E, -M, -O and -P contain a region of AD1 coded by exons 4–6, isoforms TCF4-F, -N and -R contain a region coded by exons 5–6, isoform TCF4-Q has only a short sequence coded by exon 6, and the rest of the isoforms (TCF4-D, -G, -A, -H and -I) lack AD1.

Apart from differences in the N-termini, the number of TCF4 isoforms is increased by in-frame alternative splicing of internal exons (Figures 1C and S1). Firstly, simultaneous skipping of internal cassette exons 8 and 9 gives rise to TCF4 Δ isoforms, as opposed to full-length isoforms that contain the amino acids coded by exons 8–9. Secondly, at exon 18 there are two alternative splice donor sites that enable to splice in or out a 12 bps sequence encoding amino acid sequence RSRS present in + isoforms and absent in − isoforms. Thirdly, exons 8 and 15 contain two alternative splice acceptor sites that lead to optional inclusion of the first three nucleotides (CAG) of an exon in mRNA and a glutamine or alanine residue, respectively, in the corresponding position in protein sequence (Figure S1). All TCF4 isoforms, regardless of their N-terminal or internal differences, contain transactivation domain AD2 (coded by exons 14–16) and the bHLH domain (coded by exon 19).

The overall structure of *TCF4* gene is conserved in the mouse genome. The identity between human and mouse sequences for internal exons 3–20 and 3′ exon 21 is 92%. Out of the 21 *TCF4* 5′ exons in human genome at least 13 are also transcribed in mouse as assessed by the presence of respective ESTs in public databases (Table S1). Sequences of most human 5′ exons align to the respective region in the mouse genome with approximately 70–99% identity (Figure S2). Exons 1b, 3c, 5b and 5c are more divergent and for exons 1a, 1 and 2 no alignment between human and mouse genes was obtained. Five out of the seven non-conserved human *TCF4* exons indicated above originate from exonization of various transposable elements. Namely, exon 1a overlaps with two LTR repeats, exon 1b overlaps with DNA transposon MER5B, exon 1 consists of SINE (Alu) and LINE repeat sequences, exon 2 consists of Alu repeat sequence and exon 5c consists of SINE (MIR) element sequence. In addition, 5′ exon 3a immediately follows a SINE (MIR) element in the genome (Figure S2).

In order to determine the relative abundance of mRNAs initiated at different positions within the *TCF4* gene, we carried out ribonuclease protection assays with the probe spanning internal exons 3–11 (Figure 2A). Transcripts containing different number of internal exons were detected in human cerebellum and muscle (Figure 2B). The longest protected fragment, containing exons 3–11, corresponds to transcripts initiated at 5′ exons 1a, 1b and 3a–3d. The fragment comprising exons 4–11 represents the sum of transcripts containing the above mentioned 5′ exons in case of exon 3 skipping and transcripts initiated at 4a–4c. The fragments comprising exons 5–11, 7–11 and 8–11 rise from transcripts initiated at 5a–5c, 7a–7b and 8a–8d, correspondingly. The fragments comprising exons 3–7, 4–7 and 5–7 represent transcripts that are initiated at the same sites as fragments 3–11, 4–11 and 5–11, respectively, but that lack the cassette exons 8–9 as a result of alternative splicing. All Δ8–9 transcripts additionally give rise to the fragment containing exons 10–11 and this fragment also includes transcripts initiated at exons 10a–10c. Densitometric quantification of the protected fragments showed that the levels of *TCF4* transcripts spanning different number of internal exons were comparable in both human cerebellum and muscle (Figure 2C). From these data we concluded that transcription is initiated at relatively similar levels from alternative sites within the *TCF4* gene.

TCF4 mRNAs are ubiquitously but not equally expressed

We studied the usage of different 5′ exons in a variety of human tissues and brain regions by reverse transcription polymerase chain reaction (RT-PCR). Our results showed that although the majority of the alternative *TCF4* transcripts were present in most tissues analyzed, there were a few that had a more limited expression pattern (Figure 3A). For instance we detected transcripts containing exon 1a only in testis, prostate and placenta. The use of 5′ exons 1b, 3a and 5c was restricted to testis, prostate and trachea. Furthermore, there were several transcripts, most remarkably 8d transcripts, that

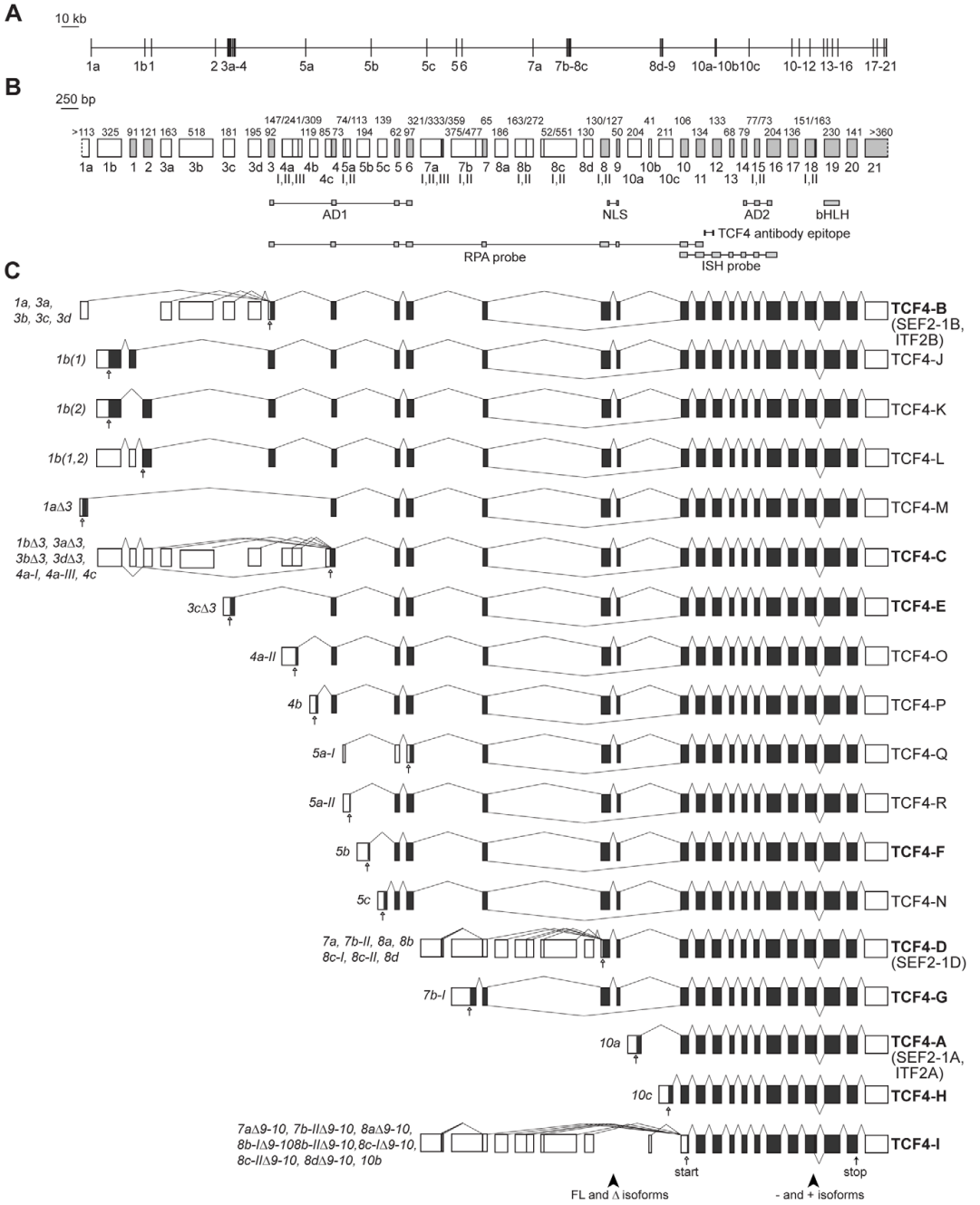


Figure 1. Structure and alternative splicing of the human *TCF4* gene. *TCF4* genomic organization with (**A**) introns drawn in scale or (**B**) exons drawn in scale. White boxes mark 5′ exons and light grey boxes represent internal or 3′ exons. Exon names are shown below the boxes. Roman numerals designate alternative splice donor or acceptor sites. Numbers above the exons indicate their sizes in bps. The regions encoding the respective domains of TCF4, the NLS identified in this study and the epitope of the used TCF4 antibody are indicated below the gene structure. Locations of RPA and ISH riboprobes used in this study are also shown. AD, transcription activation domain; bHLH, basic helix-loop-helix domain; NLS, nuclear localization signal; RPA,

ribonuclease protection assay; ISH, *in situ* hybridization. (**C**) *TCF4* alternative transcripts grouped together according to the encoded TCF4 protein isoform. Translated and untranslated regions are indicated as dark grey and white boxes, respectively. Transcripts are designated with the name of the 5′ exon and, if needed, with the number of the splice site used in the 5′ exon. Excluded internal exons are shown with the symbol Δ and included internal exons in parentheses, if necessary. The names of the protein isoforms are shown at the right. The isoforms cloned in this study are brought in bold. The position of the first in-frame start codon for each transcript and stop codon are shown with empty and filled arrows, respectively. Arrowheads at the bottom of the panel point to the regions of alternative splicing giving rise to full-length (FL) and Δ, − and + isoforms.
doi:10.1371/journal.pone.0022138.g001

featured considerably higher expression levels in the nervous system than in other tissues. RT-PCR analysis also provided information about the occurrence of alternative splicing. The skipping of exon 3 (Δ3) was a minor event in case of transcripts 1b, 3a, 3c, and 3d, whereas comparable levels of full-length and Δ3 mRNAs were present in case of transcripts 1a and 3b. Several general observations were made concerning the usage of alternative splice donor sites at 5′ exons. Firstly, the prevalence of splice site utilization at exon 7a decreased in the row of II, III, I. Secondly, the levels of 7b-I transcripts were higher than those of 7b-II transcripts in most of the tissues analyzed, except in the nervous system where nearly equal levels of 7b-I and 7b-II transcripts were detected. Thirdly, 8b-I transcripts were present only in the cerebellum, whereas in all other tissues and brain regions splice donor site II of exon 8b was exclusively used. Fourthly, splice site II was predominantly used at exon 8c. The levels of 5a and 8c-I transcripts were too low for reliable

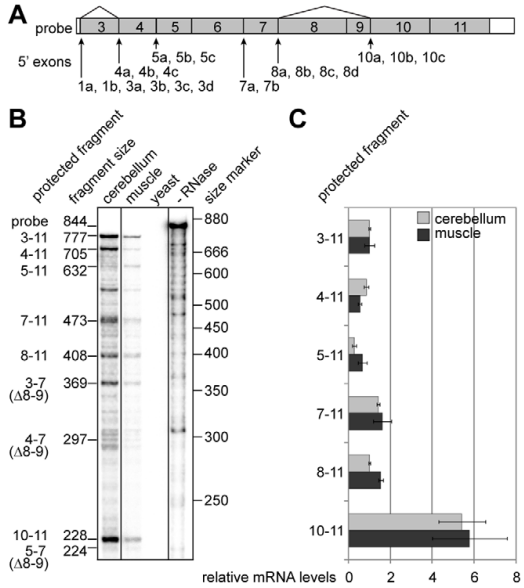


Figure 2. Initiation of transcription from alternative sites within the *TCF4* gene. (**A**) Schematic representation of the ribonuclease protection assay probe complementary to *TCF4* exons 3–11. Location of the *TCF4* 5′ exons relative to the probe is shown with arrows and the sites of alternative splicing with lines. (**B**) Autoradiograph of the probe fragments protected by human cerebellum or muscle RNA and fragments obtained from control reactions with yeast RNA or without RNase treatment. The expected sizes of the protected fragments in bps and the exons they span are shown at the left and the location of the size markers at the right. (**C**) Densitometric quantification of the protected fragments in B from two assays. The values are given in relation to the levels of the fragment spanning exons 3–11 for both tissues. Error bars indicate standard deviations.
doi:10.1371/journal.pone.0022138.g002

expression analysis; nevertheless, the respective PCR amplification products were consistently detected in the brain samples (data not shown). Our analysis did not reveal the usage of splice donor site II at exon 4a in any of the tissues studied.

Next, we examined in-frame alternative splicing of *TCF4* internal exons (Figure 3B). To monitor the splicing of cassette exons 8–9, we performed PCR with forward and reverse primers in exon 5 and 11, respectively. To evaluate the usage of two alternative splice donor sites at exon 18, we amplified the region spanning exons 10–20 and analyzed the PCR product by restriction with BglII, that has a unique recognition site in the 12 bp region present only in transcripts coding for the + isoforms. As shown in Figure 3B inclusion of exons 8–9 was prevalent and skipping of exons 8–9 a rare event in most tissues analyzed. Comparable amounts of full-length and Δ8–9 transcripts were present only in the *corpus callosum*. The levels of transcripts containing or lacking the extra 12 bps of exon 18 were roughly equal in most tissues analyzed (Figure 3B).

To compare *TCF4* expression levels in different human tissues more precisely, we performed quantitative PCR with three pairs of primers designed to amplify all *TCF4* transcripts (products spanning exons 10–11, 17–18 or 19–20). The obtained relative values were normalized to the expression levels of four housekeeping genes as described in Materials and Methods. We found that *TCF4* mRNA levels were considerably elevated in fetal brain and adult cerebellum – approximately 200 and 100 times above the levels measured in colon, respectively (Figure 4A). Compared to the levels in colon, about 40-fold higher levels were detected in cerebral cortex and spleen and more than 10-fold higher levels were seen in uterus, lung, thymus and placenta. The lowest quantities of *TCF4* transcripts were present in fetal liver, pancreas and colon. From these results we concluded that although the expression of *TCF4* is ubiquitous, its levels vary considerably between tissues. Particularly, we turned our attention to high expression levels of *TCF4* in the nervous system and carried out *in situ* hybridization experiments on sections from human hippocampus and cerebellum to characterize *TCF4* expression at cellular level. As shown in Figure 4B *TCF4* mRNA was detected in hippocampal neurons in dentate gyrus and CA1-CA3 regions, neurons of subiculum and parahippocampal gyrus of the cortex, and cerebellar granule neurons.

Several TCF4 protein isoforms are expressed in human tissues

We next asked whether and which of the different TCF4 protein isoforms are translated *in vivo*. To address this question we monitored the expression of endogenous TCF4 isoforms in different human tissue extracts by western blotting with TCF4-specific antibodies that recognize an epitope present in all described TCF4 isoforms. As shown in Figure 5A multiple TCF4 isoforms were present in human lung, liver, kidney, muscle and testis. In these tissues we detected three prominent bands that were assigned as high, medium and low molecular weight (Mw) TCF4. In case of frontal cortex, hippocampus and cerebellum low molecular weight TCF4 band was predominant and the signals of medium and high Mw TCF4 were very low. The levels of high or

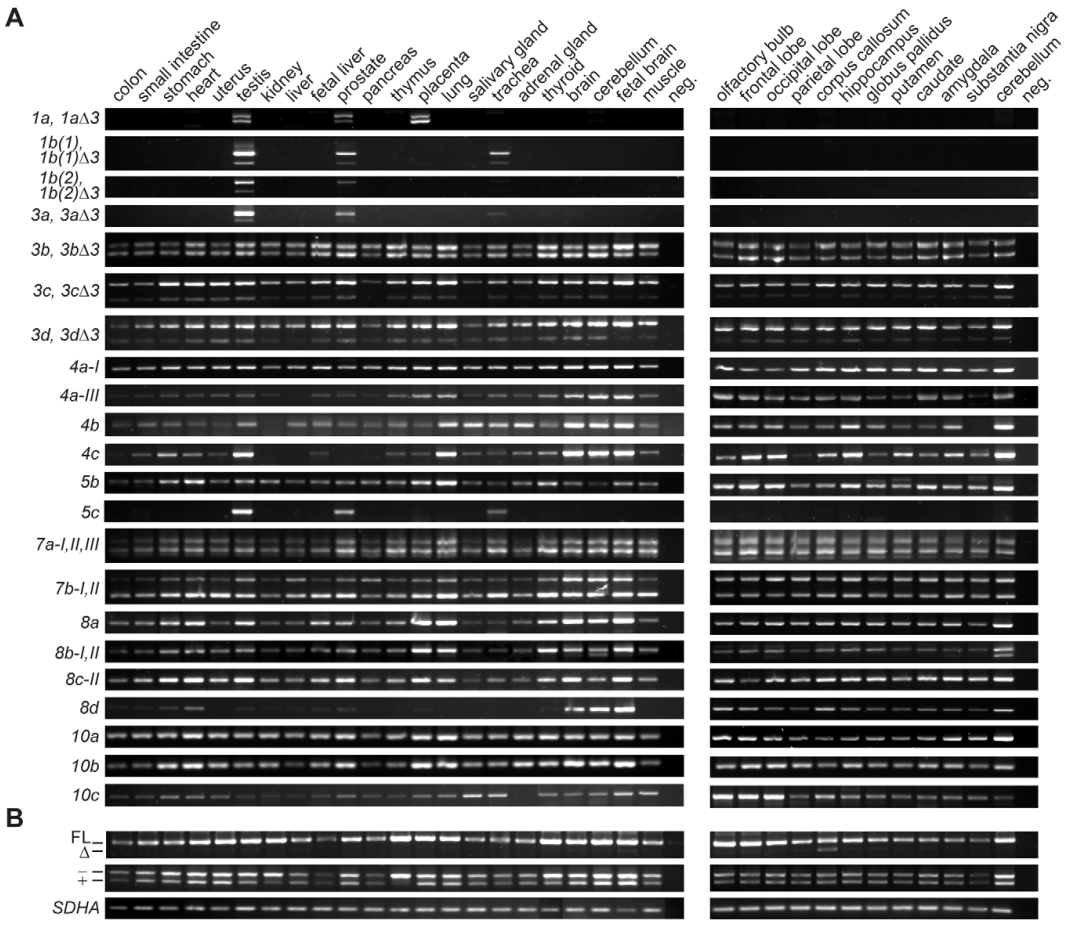


Figure 3. Expression of alternative *TCF4* mRNAs in human tissues and brain regions. (**A**) RT-PCR analysis of *TCF4* transcripts with different 5′ exons and (**B**) with alternative internal splicing. Transcripts are designated as in Figure 1C. The positions of bands respective to the transcripts encoding the full-length (FL) and Δ isoforms, − and + isoforms are indicated at the left on panel B. mRNAs with longer exon 18 (+) give rise to RT-PCR product that has a unique BglII restriction site enabling discrimination from RT-PCR products amplified from mRNAs not containing the 12 bps insert (−). House-keeping gene *SDHA* mRNA expression is shown at the bottom of the panel. PCR with no template was performed as a negative control (neg) with each primer pair.
doi:10.1371/journal.pone.0022138.g003

medium Mw TCF4 were elevated relative to other forms of TCF4 in testis or lung, respectively. To validate the specificity of the TCF4 antibodies, we implemented a RNAi based approach in Neuro2A mouse neuroblastoma cells. Similarly to the human tissue extracts, high, medium and low Mw bands were detected in the lysates of Neuro2A cells by western blotting with the TCF4 antibodies. Compared to mock and scrambled siRNA transfected cells the levels of all three forms were reduced in cells transfected with three different siRNAs targeting *TCF4* exon 12 or 20 that are present in all described *TCF4* transcripts (Figure 5B), thus verifying that the TCF4 antibodies used in the current study specifically recognize TCF4 proteins.

In order to define the nature of the three bands, we cloned the coding regions of isoforms TCF4-A – TCF4-I into pCDNA mammalian expression vector. In addition, we cloned variants coding for + and − isoforms, full-length and Δ isoforms. We transfected the obtained plasmids into HEK293 cells and detected the overexpressed proteins by western blotting with TCF4-specific antibodies. The different TCF4 isoforms were expressed at variable levels (Figure 5C). The levels of isoforms TCF4-B$^+$, -B$^-$, -BΔ$^+$, -BΔ$^-$, -CΔ$^-$, -A$^+$, -A$^-$, -H$^-$ and -I$^-$ were high, those of TCF4-C$^-$ and -D$^-$ medium, and those of TCF4-E$^-$ and -F$^-$ low. Expression of TCF4-G$^-$ was not detected by western blotting. Although it was not possible to distinguish all isoforms from each other by size, we compared side by side the bands detected in human muscle and testis to different TCF4 isoforms translated *in vitro* (Figure 5D) and made the following conclusions about the nature of the three bands in human tissue extracts: the high Mw TCF4 fractionates similarly to TCF4-B and -C, the medium Mw TCF4 similarly to TCF4-D, and low Mw TCF4 to TCF4-A and TCF4-H. In addition, a few proteins with smaller molecular weight than any of the cloned TCF4 isoforms were detected in

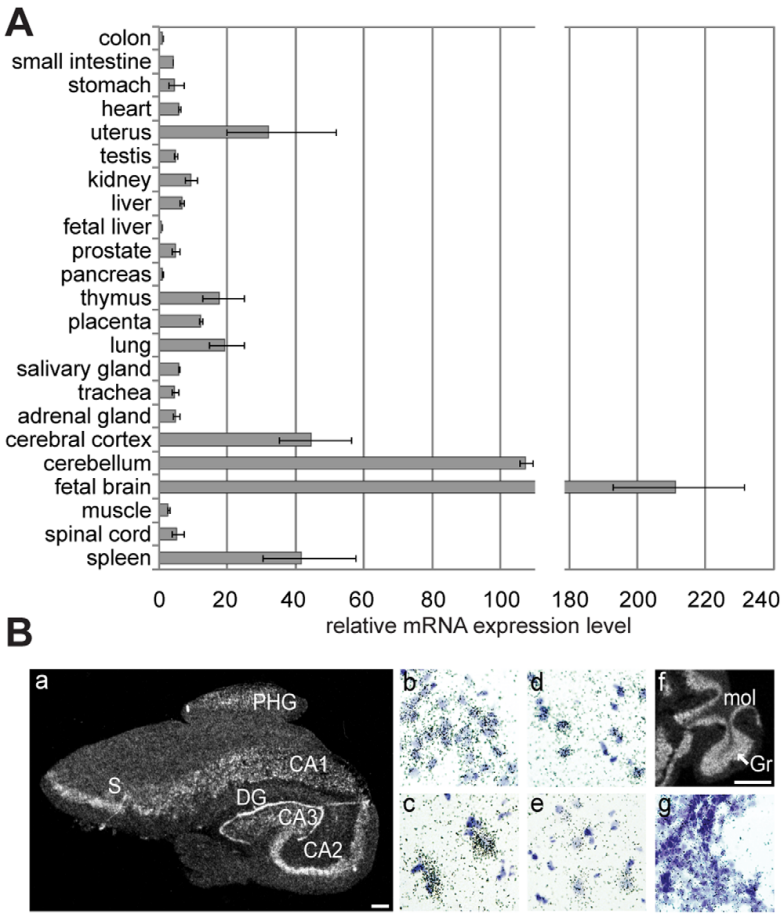


Figure 4. Inter-tissue variability of *TCF4* mRNA levels. (**A**) Quantitative RT-PCR analysis of *TCF4* expression in human tissues. Levels of *TCF4* transcripts were determined using three different primer pairs and the results were normalized to the expression levels of four house-keeping genes (*SDHA*, *HMBS*, *GAPDH* and *UBC*). Shown are the means relative to the TCF4 expression level measured in colon that was arbitrarily set as 1. Error bars indicate standard deviations. (**B**) *In situ* hybridization analysis of *TCF4* expression in human hippocampus and cerebellum. Autoradiographs from (**a**) a coronal section of the hippocampus and (**f**) a sagittal section of the cerebellum are shown. Scale bar 1 mm. Bright-field higher magnification images of emulsion-dipped and hematoxylin-stained sections of (**b**) the granular cell layer of the dentate gyrus, (**c**) pyramidal cell layer of the CA2 region, (**d**) subiculum and (**e**) PHG region of the cortex. CA1, CA2, CA3, respective regions of the hippocampus; DG, dentate gyrus; Gr, granular cell layer of the cerebellum; mol, molecular layer of the cerebellum; PHG, parahippocampal gyrus; S, subiculum.
doi:10.1371/journal.pone.0022138.g004

several tissues (Figure 5A). These could be non-specific signals, represent TCF4 proteolytic fragments or undescribed TCF4 isoforms. In sum, the above described *TCF4* gene structure and expression data demonstrate that multiple TCF4 isoforms are present in human tissues and suggest that the isoforms could potentially differ in their functional properties.

TCF4 is imported to the nucleus due to its NLS or via a piggy-back mechanism

To gain insight into possible functional variation amongst the TCF4 isoforms we first investigated, by indirect immunofluorescence, the intracellular localization of TCF4 isoforms overexpressed in HEK293 cells (Figure 6A). We observed two patterns of distribution – exclusively nuclear and nuclear plus cytoplasmic. Full-length TCF4 isoforms that have longer N-termini (TCF4-B^+, -B^-, -C^-, -D^-, -E^-, -F^- and -G^-) were restricted to cell nucleus whereas isoforms with shorter N-termini (TCF4-A^-, -H^- and -I^-) and Δ isoforms (TCF4-$B\Delta^+$, -$B\Delta^-$ and -$C\Delta^-$) localized to the nucleus and, in addition, to the cytoplasm.

Since all isoforms with exclusively nuclear distribution contain the amino acids coded by exons 8–9 and the isoforms with broader intracellular localization are devoid of these amino acids, we searched for possible nuclear localization signal (NLS) in this region. Using PSORT software we identified a potential bipartite NLS that contains two clusters of basic amino acids (RRR and KKVRK) separated by a linker of 11 amino acids. The two clusters are conserved in TCF4 of *Mus musculus*, *Xenopus laevis* and

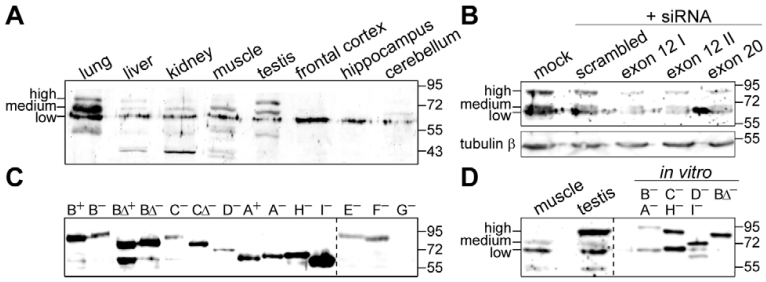


Figure 5. Expression of TCF4 protein isoforms. (**A**) Western blot analysis of different human tissues and brain regions with TCF4 antibodies. (**B**) The effect of TCF4 targeting siRNAs on the levels of proteins detected by TCF4 antibodies in extracts of Neuro2A cells. Three different siRNAs specific for *TCF4* exon 12 or 20 were transfected into Neuro2A cells; mock and scrambled siRNA transfections were performed in control. Tubulin β levels were determined to demonstrate equal loading. (**C**) Western blot analysis of TCF4 isoforms overexpressed in HEK293 cells using TCF4 antibodies. (**D**) Comparison of fractionation of proteins recognized by TCF4 antibodies in human muscle and testis extracts to *in vitro* translated selected human TCF4 isoforms. The localization of endogenous TCF4 with high, medium or low molecular weight is indicated at the left in A, B and D; the dashed line in C and D separates different exposures; molecular mass (in kDa) marker bands are shown at the right on all panels.
doi:10.1371/journal.pone.0022138.g005

Danio rerio, and also in two other E-proteins E2A and HEB (Figure 6B). To validate the NLS we performed site-directed mutagenesis in the context of TCF4-B$^-$ isoform and replaced the basic amino acids in cluster 1, 2 or both with alanines. As shown in Figure 6C mutagenesis of either cluster 1 (M1) or 2 (M2) produced a protein that is partly cytoplasmic, but even when both clusters were mutated (M1+2) a substantial portion of the protein remained nuclear, similarly to the genuine TCF4 isoforms not containing the region coded by exons 8–9. As a next step we fused the NLS to the C-terminus of EGFP and monitored its localization in HEK293 cells. EGFP-NLS fusion protein was confined to the nucleus whereas control EGFP was distributed diffusely all over the cell (Figure 6D). From these results we concluded that the identified NLS is functional, but additional region(s) responsible for nuclear import of isoforms lacking the amino acids coded by exons 8–9 exist in TCF4 protein.

Another region with high basic charge density is located in the DNA-binding domain of TCF4. To test whether this part of the protein is able to mediate nuclear import we constructed different fusion proteins where the entire bHLH domain or its N-terminal or C-terminal part is fused with EGFP. The fusion proteins were overexpressed in HEK293 cells and their localization was studied by direct EGFP fluorescence (Figure 6D). EGFP-bHLH was present in the nucleus and cytoplasm indicating that the bHLH domain was to some extent able to direct the protein to the nucleus, since in contrast to EGFP (29 kDa), the EGFP-bHLH (43 kDa) fusion protein cannot be transported to the nucleus by passive mechanisms due to its higher molecular mass. At the same time EGFP-bHLH-N (36 kDa) and EGFP-bHLH-C (37 kDa) were detected only in the cytoplasm, meaning that neither part alone was able to mediate nuclear import. These data led us to hypothesize that structural integrity of the bHLH domain is necessary for EGFP-bHLH nuclear localization mediated by heterodimerization with endogenous NLS bearing HLH-proteins. To test this assumption, we co-expressed EGFP and the described fusion proteins together with known TCF4 dimerization partners: E2-tagged NeuroD2 (Gene 18013), that features nuclear localization, and mCherry-fused Id2 (Gene 15902), that bears a nuclear export signal (NES) and is similarly to the described EGFP-Id2 fusion protein [29] predominantly cytoplasmic. The results of this assay showed that NeuroD2-E2 did not influence the distribution of EGFP and EGFP-bHLH-N, but was able to direct EGFP-bHLH and to lesser extent also EGFP-bHLH-C to the nucleus. Co-expression of mCherry-Id2 altered only the localization of EGFP-bHLH by excluding it from the nucleus (Figure 6D). We additionally monitored the effects of NeuroD2-E2 and mCherry-Id2 co-expression on the localization of genuine NLS-lacking TCF4 isoform A$^-$. Similarly to EGFP-bHLH, the distribution of TCF4-A$^-$ was guided by its heterodimerization partner (Figure 6E). Altogether these results demonstrate that TCF4 can be directed to the nucleus by two mechanisms. Firstly, isoforms that bear the NLS independently translocate to the nucleus. Secondly, all TCF4 isoforms can be transported to the nucleus by a piggy-back mechanism through heterodimerization with NLS containing bHLH partners. Additionally, isoforms lacking NLS can be exported from the nucleus by heterodimerization with NES containing partner proteins.

Subsequently, we examined the intracellular distribution of endogenous TCF4. For this we chose to study human hippocampus and cerebellum because of high TCF4 expression levels in these brain regions. We performed immunohistochemical staining of the sections using TCF4 and neuronal marker NeuN or glial marker GFAP specific antibodies, followed by confocal microscopy. As shown in Figure 6F, TCF4 signal was found mainly in the neuronal nuclei but also in the cytoplasm of neurons in different hippocampal regions (CA, dentate gyrus, hilus) and cerebellar granular layer. Of note, the TCF4 antibody additionally marked processes of GFAP-positive cells in tissues and rat primary cultures, but to our knowledge this glial cytoplasmic signal is non-specific as no such staining was observed when tagged TCF4 isoforms were overexpressed in cultured rat glial cells and stained with tag-specific antibodies (data not shown).

TCF4 isoforms activate transcription differentially

As a next step in functional characterization of TCF4 isoforms we performed reporter assays with a construct carrying 12 μE5 (CACCTG) E-boxes [1] in front of a minimal promoter controlling the expression of firefly luciferase gene *luc2P* (Figure 7A). The reporter construct was transfected into HEK293 cells together with expression plasmids encoding different TCF4 isoforms and a vector with the minimal promoter in front of *Renilla* luciferase *hRlucP* gene. Compared to empty vector transfected cells the normalized luciferase activity was approximately 250 or 150 times higher in cells expressing TCF4-

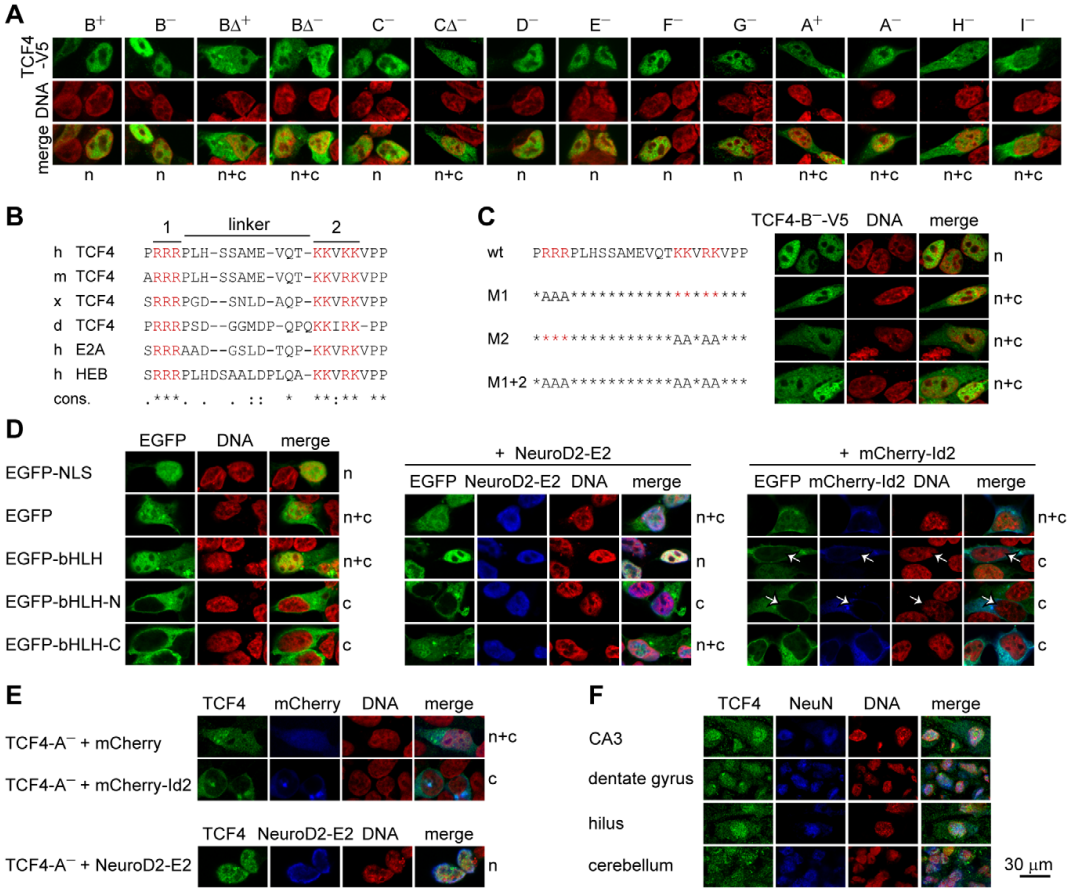


Figure 6. TCF4 intracellular localization. (**A**) Immunocytochemical analysis of V5-tagged TCF4 isoforms overexpressed in HEK293 cells. Isoforms analyzed are indicated at the top and localization pattern at the bottom of the panel. n, nuclear; c, cytoplasmic; n+c, nuclear and cytoplasmic. DNA was counterstained with DAPI (pseudocoloured red) to visualize nuclei. (**B**) Alignment of the identified bipartite NLS in TCF4 of *Homo sapiens* (h), *Mus musculus* (m), *Xenopus laevis* (x) and *Danio rerio* (d); in E2A and HEB of *Homo sapiens*. Two clusters of basic amino acids (in red, 1 and 2) and the linker are indicated with lines at the top. Conservation is indicated at the bottom. '*', identity; ':', conserved; '.', semi-conserved substitution. (**C**) The effect of site-directed mutagenesis of the NLS on the localization of TCF4 in HEK293 cells. Basic amino acids in cluster 1 (M1), 2 (M2) or both (M1+2) were replaced with alanines in the context of TCF4-B$^-$-V5 and the localization of the proteins was monitored by immunocytochemistry. (**D**) Localization of EGFP fusion proteins with TCF4 NLS, bHLH, N-terminal (N) or C-terminal (C) half of TCF4 bHLH domain in HEK293 cells. NeuroD2-E2 or mCherry-Id2 encoding plasmid was cotransfected when indicated. White arrows indicate cells expressing mCherry-Id2. (**E**) Effect of mCherry-Id2 or NeuroD2-E2 co-expression on subcellular distribution of TCF4-A$^-$. The localization patterns of overexpressed TCF4 proteins are indicated at the right of the panels in C, D and E. (**F**) Immunohistochemical analysis of endogenous TCF4 in human hippocampal and cerebellar sections. NeuN staining was used to identify neurons.

doi:10.1371/journal.pone.0022138.g006

B$^+$ or -B$^-$, respectively. More than 650-fold increase was observed with TCF4-B Δ$^+$ and -B Δ$^-$ isoforms. TCF4-C$^-$ activated reporter gene transcription over 10 and TCF4-C Δ$^-$ almost 50 times. The increased luciferase activity in Δ isoforms expressing cells compared to the respective full-length isoform expressing cells could reflect higher levels of Δ isoforms in HEK293 cells (Figure 5C). The two isoforms with low expression level, TCF4-E$^-$ and -F$^-$, activated reporter gene transcription about 7- and 21-fold, correspondingly. Isoforms completely lacking AD1 elevated luciferase activity approximately 20 (TCF4-D$^-$, -A$^+$ and -A$^-$) or 70 times (TCF4-H$^-$ and -I$^-$). Search for correlations revealed a strong positive relationship between the presence of full-length AD1 and isoform's ability to transactivate in HEK293 cells (point biserial correlation coefficient $r = 0.87$, $p < 0.00011$). There were no significant correlations between the isoform's capacity to activate transcription and the presence of NLS, presence of only the C-terminal part of AD1 or presence of the extra four amino acids in the + isoforms. These results indicate that, although to a different extent, all TCF4 isoforms analyzed are able to activate transcription controlled by μE5 E-boxes in HEK293 cells.

To study further the individual role of TCF4 transcription activation domains AD1 and AD2, we took two approaches.

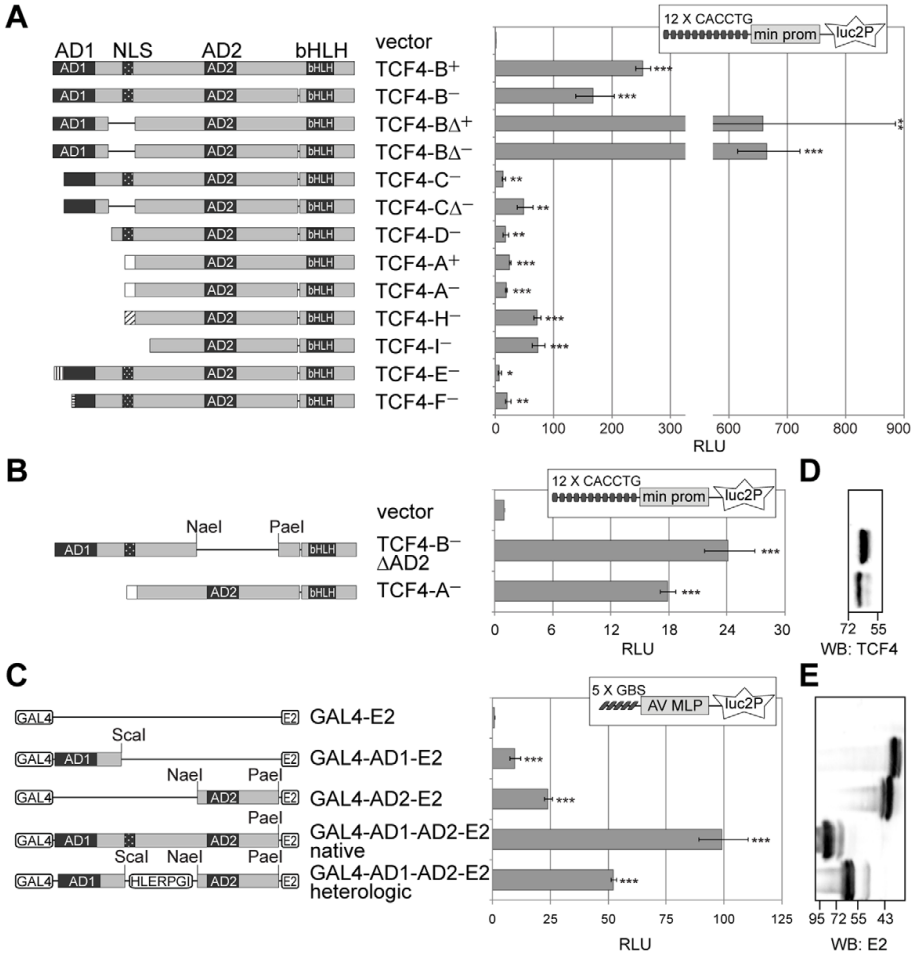


Figure 7. Transcription activation by alternative TCF4 isoforms. (**A**) Reporter assay with HEK293 cells transfected with firefly luciferase construct carrying 12 μE5 E-boxes in front of a minimal (min) promoter along with the indicated TCF4 isoform encoding plasmid or an empty vector. (**B**) Reporter assay with HEK293 cells transfected with luciferase constructs and TCF4-B$^-$ΔAD2 or TCF4-A$^-$ encoding plasmid or empty vector as indicated. (**C**) Reporter assay with HEK293 cells transfected with firefly luciferase construct carrying 5 GAL4 binding sites (GBS) in front of adenovirus major late promoter (AV MLP) along with the indicated E2-tagged GAL4 fusion proteins. (**A**, **B** and **C**) For normalization *Renilla* luciferase construct with minimal promoter was cotransfected. Luciferase activities were measured and data are presented as fold induced levels above the signals obtained from empty vector transfected cells. Shown are the mean results from at least three independent experiments performed in duplicates, error bars indicate standard deviations. Statistical significance shown with asterisks is relative to the luciferase activity measured from empty vector transfected HEK293 cells (*, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; t-test). RLU, relative luciferase units. Schematic representation of the expressed TCF4 proteins with the locations of restriction enzymes used for the generation of the respective plasmids is shown at the left. (**D**) Western blot analysis of TCF4-B$^-$ΔAD2 and TCF4-A$^-$ expressed in HEK293 cells. (**E**) Western blot analysis of E2-tagged GAL4 fusion proteins expressed in HEK293 cells. (**D** and **E**) Localization of molecular mass (in kDa) marker bands is indicated at the bottom and the order of samples is as in B and C, respectively.
doi:10.1371/journal.pone.0022138.g007

Firstly, we compared the abilities of artificial TCF4-B$^-$ without AD2 (ΔAD2) and native TCF4-A$^-$ to regulate E-box controlled transcription. TCF4-B$^-$ ΔAD2 contains the AD1 and TCF4-A$^-$ the AD2 domain, both have the bHLH domain. Secondly, we used heterologic constructs where AD1 or AD2 is fused with GAL4 DNA binding domain and E2 epitope-tag. We assessed the ability of the GAL4 fusion proteins to activate reporter transcription from pG5luc vector that carries GAL4 binding sites in front of firefly luciferase *luc* gene. As shown in Figures 7B and 7C similar reporter gene activation levels were achieved with TCF4-B$^-$ ΔAD2 compared to TCF4-A$^-$, and GAL4-AD1-E2 compared to GAL4-AD2-E2 in HEK293 cells. When joined to native bHLH DNA binding domain, 24 or 18 fold upregulation of reporter gene transcription were seen with AD1 or AD2 containing proteins, respectively (Figure 7B) When fused with heterologic GAL4 DNA binding domain, AD1 activated transcription about 10 times and AD2 about 24 times (Figure 7C). We noted that when both activation domains are present in a single

protein as is the case for TCF4-B isoforms, the activation of transcription exceeds the additive effect of AD1 and AD2, suggesting that the two domains may act synergistically in HEK293 cells. To test this we studied the transactivation capacity of GAL4 fusion proteins that contain both TCF4 activation domains joined by native or heterologic amino acids. As shown in Figure 7C these proteins activated reporter transcription approximately 100 and 50 times, respectively. In both cases the activation fold exceeded the additive effect of single TCF4 activation domain containing GAL4 fusion proteins. All the compared proteins were expressed at similar levels in HEK293 cells as determined by western blotting with TCF4 or E2-specific antibodies (Figures 7D and 7E). Altogether these data show that the two activation domains are capable of mediating transactivation to a similar extent in HEK293 cells. Additionally, the results indicate that when both activation domains are present in a single protein, the two domains act synergistically in HEK293 cells.

Discussion

Large-scale human transcriptome analyses have revealed that more than 90% of protein-coding genes undergo alternative splicing and around half have two or more alternative promoters [26,30–32]. Additionally, over-representation of alternative promoters has been attributed to genes involved in development and regulation of transcription [32]. Here, we demonstrate that the human gene for the transcription factor TCF4 is transcribed using 21 mutually exclusive 5′ initial exons, many of which contain more than one transcription start sites and/or splice donor sites. These 5′ exons are located at various positions in the gene interspersed with internal exons 1–9, followed by constitutive internal exons 10–20 and 3′ exon 21. Therefore, *TCF4* transcripts containing different number of internal exons are generated and, according to our data, these are expressed at comparable levels in human tissues and potentially encode for TCF4 protein isoforms with 18 different N-termini. We named the isoforms TCF4-A – TCF4-R in agreement with the previous studies that have described three N-terminally distinct isoforms, i.e. TCF4-A, -B and -D [2,18]. Additionally, several studies have described TCF4 isoforms differing by the presence or absence of four amino acids (RSRS) N-terminal to the bHLH domain [2,18,33,34], which are denoted here as + and − isoforms, respectively. These isoforms result from alternative splice donor site selection at exon 18 and their mRNAs are present at comparable levels in most human tissues analyzed here. We show that in human *TCF4* gene there is another site of alternative splicing at internal exons 8–9. Transcripts lacking these exons are expressed at low levels in human tissues and code for protein isoforms indicated with Δ in this study.

One by one analysis of alternative 5′ exons containing *TCF4* transcripts' expression and quantitative analysis of overall *TCF4* expression in human tissues corroborated the concept of TCF4 being a ubiquitous transcription factor. Most alternative 5′ exons containing transcripts were broadly expressed. Nevertheless, expression of transcripts containing 5′ exons 1a, 1b, 3a and 5c was detected only in a few tissues including testis and prostate. These exons originate from exonization of different transposable elements (TEs) or are located immediately behind a TE in the human genome. Incorporation of TE-derived sequences into promoters or UTR or coding regions of genes has been documented by many studies [35–37]. In accordance with the restricted expression of TE-dependent *TCF4* transcripts, TEs are known to be transcriptionally silenced in most mammalian tissues as a defense mechanism against potentially deleterious effects of their activity [38,39]. TE-dependent *TCF4* transcripts code for 5 unique TCF4 protein isoforms with long N-termini (TCF4-J – TCF4-N). Since it was not possible to distinguish all TCF4 isoforms from each other by SDS-PAGE fractionation, we were not able to conclusively determine which protein isoforms are expressed in human tissues; nevertheless, we detected three prominent TCF4 forms of different molecular weight. Notably, high Mw TCF4 protein was relatively more abundant in testis, medium Mw TCF4 in lung and low Mw TCF4 was the major form in the brain tissue, insinuating that the ratio of different TCF4 isoforms with distinct N-termini is varied between human tissues.

Our report reveals that although broadly expressed, *TCF4* transcript levels differ greatly between tissues. The highest levels are present in fetal brain, but expression remains elevated also in adult brain where we detected *TCF4* mRNA and protein in neurons. These findings are consistent with earlier studies that have demonstrated TCF4 expression in human or rodent nervous system [18,40–44] and further support the important role for TCF4 in the development and functioning of the nervous system as exemplified by involvement of TCF4 in the pontine nucleus development in mice, association of *TCF4* with schizophrenia and identification of *TCF4* haploinsufficiency as the cause of Pitt-Hopkins mental retardation syndrome [17,21–24]. Interestingly, translocation that results in a *TCF4* allele without at least 6 upper 5′ exons has been described in a patient with mild mental retardation [45]. Possibly, in this patient, initiation of *TCF4* transcription at downstream 5′ exons still takes place, explaining the less severe phenotype than in classical Pitt-Hopkins syndrome patients. Nevertheless, it seems that for production of sufficient amounts of TCF4 protein and normal development, the presence of all transcription initiation sites is critical. Based on experiments with *HEB* and *E2A* knockout mice, it has been suggested that the overall E-protein dosage is more crucial in the development of nervous system than family member identity [46]. However, since even slight disturbances in *TCF4* expression in human cause a neurodevelopmental disorder and *TCF4* knockout mice display disrupted pontine nucleus development [17], it is apparent that other E-proteins are not able to compensate for the loss of TCF4 in all aspects of the nervous system development. Additionally, transgenic mice with mild overexpression of *TCF4* in forebrain display deficits in contextual and cued fear conditioning and sensorimotor gating [47], indicating that precise regulation of *TCF4* expression is crucial for correct brain function and deviations in either way result in cognitive disturbances. All this substantiates the complex structure and the high number of 5′ exons in *TCF4* gene, as these probably enable proper amounts of TCF4 protein to be synthesized. It is noteworthy that according to the *TCF4* gene structure described here, the SNP rs9960767 associated with susceptibility to schizophrenia [21] is located in *TCF4* intronic sequence between 5′ exons 5b and 5c. However, whether and how this marker is coupled to dysregulation of *TCF4* expression remains to be elucidated.

The variety of TCF4 isoforms raises a question – are they produced only for sufficient amount of TCF4 proteins or are the functions of different isoforms divergent? In the present study we have shown that the alternative TCF4 isoforms differ in two aspects: first, their intracellular distribution is differentially regulated depending on the presence or absence of a nuclear localization signal; second, their ability to activate E-box controlled transcription is varied depending mainly on whether they contain one or two functional transcription activation domains. We discovered in the region coded by exons 8–9, upstream of AD2, a bipartite NLS that is responsible for nuclear localization of TCF4 isoforms with longer N-termini. This NLS is conserved among E-proteins and mutating the second half of the NLS has been demonstrated to reduce nuclear import of E2A [48].

TCF4 isoforms with shorter N-termini (TCF4-A, -H and -I) and Δ isoforms do not contain the NLS. Our analysis shows that these proteins are transported to the nucleus via heterodimerization with NLS-bearing partners and directed to the cytoplasm by forming dimers with NES-containing proteins. As substantial amount of NLS-lacking TCF4 isoforms overexpressed in HEK293 cells is localized to the nuclei, there must be abundance of NLS-containing heterodimerization partners expressed endogenously. This is not surprising since TCF4 is able to form dimers with a variety of class B bHLH factors [3]. In addition, we noticed that NLS-bearing TCF4-B isoform was not able to mediate nuclear redirection of NLS-lacking TCF4-A isoform (data not shown), suggesting that TCF4 homodimers are not efficiently formed in HEK293 cells.

E-proteins function as transcription activators or repressors [49]. In our reporter assays all TCF4 isoforms were able to activate transcription from a promoter containing μE5 E-boxes in HEK293 cells whereas TCF4-B isoforms were more potent transactivators than other isoforms. All the studied isoforms contain the AD2 domain, but only TCF4-B isoforms contain full-length AD1, including exon 3 encoded LDFS motif that is known to be required for AD1 mediated transactivation by recruiting histone acetyltransferases [50,51]. This motif is present also in some of the TE-dependent isoforms (TCF4-J, -K, and -L) that were not included in our reporter assays. We suggest that the AD1 acts synergistically with the AD2, since the additive effect of individual capacities of these domains to activate transcription is surpassed by proteins that contain both TCF4 activation domains. Similar synergism between two activation domains has been described in bHLH-zipper protein TFE3 and POU homeodomain protein Oct-2 [52,53]. In contrast to a study on *FGF-1* promoter that proposed differential roles for TCF4 + and − isoforms [33], we saw no differences between the + and − isoforms in ability to activate transcription from μE5 E-boxes controlled promoter. However, we noticed higher reporter gene transcription in the presence of Δ isoforms than the respective full-length isoforms, probably due to higher expression levels of Δ isoforms compared to full-length isoforms. This indicates that NLS coding region absent in Δ isoforms could affect the stability of TCF4 proteins. In sum, we show that differences in the functioning of alternative TCF4 isoforms do exist. In the light of the knowledge that an isoform of HEB, homologous to TCF4-A, is specifically required for the generation of T-cell precursors *in vivo* and this function cannot be carried out by the HEB isoform homologous to TCF4–B [54], it would be of importance to study the distinct functions of TCF4 isoforms by rescue experiments with different TCF4 isoforms in *TCF4* knockout background.

Materials and Methods

Ethics statement

All experiments with human postmortem tissues were approved by the ethics committee of medical studies at National Institute for Health Development of Estonia (Permit Number: 402). The protocols involving animals were approved by the ethics committee of animal experiments at Ministry of Agriculture of Estonia (Permit Number: 45).

Bioinformatic analyses

Human *TCF4* gene structure and mRNAs were identified by analyzing genomic, mRNA and expressed sequence tag (EST) databases using tools available at http://www.ncbi.nlm.nih.gov and http://genome.ucsc.edu. The locations of transposable elements were determined by RepeatMasker track in UCSC Genome browser. Nuclear localization signal (NLS) was predicted using software at http://wolfpsort.org. Sequence alignments were prepared with tools available at http://www.ebi.ac.uk/Tools. The nucleotide sequences have been deposited in the EMBL Nucleotide Sequence Database under Accession Numbers FR748202-FR748223.

Constructs

Standard methods of recombinant DNA technology were used for generation of all constructs. Full-length coding regions of TCF4 isoforms were PCR-amplified from human brain cDNA and cloned into pcDNA3.1 (Invitrogen). Both, constructs coding for native TCF4 and C-terminally V5/His-tagged TCF4 isoforms, were created. TCF4-B$^-$ amino acid sequence is used as a reference for the description of following TCF4 constructs. pEGFP-C1-C3 (Clontech) were used for generation of pEGFP-NLS, pEGFP-bHLH, pEGFP-bHLH-N and pEGFP-bHLH-C that code for EGFP fusion proteins containing TCF4 amino acids P156-P178, I541-M667, I541-K585 and E586-M667, respectively. Mutagenesis of TCF4 NLS was performed using complementary primers against the target sequence containing the respective mutation using Phusion High-Fidelity DNA Polymerase (Finnzymes). pcDNA-TCF4B$^-$ΔAD2 codes for TCF4-B$^-$ without amino acids G316-M497. GAL4 DNA-binding domain was obtained from pBind vector (Promega) and inserted into pQM-CMV-E2-C vector (Icosagen). TCF4 activation domains AD1 (M1-Y148), AD2 (G316-G496), AD1–AD2 (M1-G469) and AD1 plus AD2 with heterologic linker (M1-Y148, HLERPGI, G316–G496) were cloned in-frame between GAL4 DNA-binding domain and E2-tag. Full-length coding regions of NeuroD2 and Id2 were PCR-amplified from mouse brain cDNA and inserted into pQM-CMV-E2-C (Icosagen) in front of E2 tag and pmCherry-C (Clontech) behind mCherry sequence, respectively. For E-box reporter vector pGL4.29[luc2P/12 μE5/Hygro], the CRE binding-site in pGL4.29[luc2P/CRE/Hygro] (Promega) was replaced with 12 μE5 E-boxes by tandem insertion of annealed oligonucleotides. For pGL4[hRlucP/min/Hygro], the 12 μE5 E-boxes were removed from pGL4.29[luc2P/12 μE5/Hygro] and firefly luciferase encoding *luc2P* was replaced with *Renilla* luciferase encoding *hRlucP* gene from pGL4.83[hRlucP/Puro] (Promega). For pBluescript-TCF4(3F-11R) and pSC-A-TCF4(10F-16R) used for cRNA probe synthesis, PCR amplified TCF4 cDNA fragments spanning exons 3–11 and 10–16 were ligated into EcoRV linearized pBluescriptKS+ vector or cloned into pSC-A vector (Stratagene), respectively. Sequences of all oligonucleotides used are listed in Supporting Table S2.

Ribonuclease protection assay

EcoRI linearized pBluescript-TCF4(3F-11R) was subjected to *in vitro* transcription using MAXIscript Kit and T3 polymerase (Ambion). The concentration of limiting nucleotide (UTP) was 10 μM of which 1/6 was [α-^{32}P]UTP (specific activity 3000 Ci/mmol; Hartmann analytics). 10 μg of total RNA and 2.5×10^5 CPM of radiolabeled probe were used for hybridization and the assay was performed with the RPA III Kit (Ambion) as suggested by the manufacturer. The protected fragments were separated in 5% acrylamid urea gel, visualized by autoradiography and quantified with ImageQuant T4 software (Amersham Biosciences).

RNA isolation and RT-PCR

Total RNAs from postmortem adult human brain regions and muscle were purified using RNAwiz reagent (Ambion) and treated with TURBO DNase (Ambion). Other human tissue RNAs were obtained from Clontech. First-strand cDNAs were synthesized from 5 μg of total RNA with Superscript III reverse transcriptase

(Invitrogen) with oligo(dT) primers according to manufacturer's recommendations. PCR amplification was performed using HotFire polymerase (Solis Biodyne). For quantitative PCR, LightCycler 2.0 engine (Roche), qPCR Core kit for SYBR R Green I No ROX (Eurogentec) and polycarbonate qPCR capillaries (Bioron) were used. The reactions were carried out in a volume of 10 µl containing 1/80 of reverse transcription reaction as a template. In control, PCR with primers specific for the ubiquitously expressed hypoxanthine-guanine phosphoribosyltransferase (HPRT), succinate dehydrogenase complex subunit A (SDHA), ubiquitin C (UBC), hydroxymethylbilane synthase (HMBS) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were performed. For normalization of quantitative PCR data the geometric mean of four selected housekeeping genes (SDHA, UBC, HMBS and GAPDH) with high expression stability ($M<1.5$) was calculated using geNorm software [55]. For quantification of *TCF4* mRNA expression, three different primer pairs detecting all *TCF4* transcripts were designed (products spanning exons 10–11, 17–18 and 19–20). The results obtained with each primer pair were normalized with the geNorm calculated factor, log-transformed and standardized as described [56]. Means and standard deviations (SD) were calculated and the data were back-transformed to the original scale for graphical representation. The bars represent geometric means and error bars represent upper and lower limits back-transformed as mean+SD and mean–SD, respectively. All products from the RT-PCR reactions were verified by sequencing. The primers together with the used annealing temperatures, cycle numbers and product sizes are listed in Supporting Table S2. When indicated PCR amplified DNA was diluted three times and subjected to restriction with BglII (Fermentas).

In situ hybridization

cRNA probes were synthesized from BamHI linearized pSC-A-TCF4(10F-16R) with MAXIScript *in vitro* Transcription Kit and T3 polymerase (Ambion), using [α-^{35}S]UTP (Amersham Biosciences) for labeling. Serial coronal sections (16 µm) from fresh-frozen adult male human hippocampus and cerebellum were subjected to *in situ* hybridization following the protocol described earlier [57]. Emulsion-dipped sections were developed after 3 weeks using D-19 developer (Eastman Kodak), fixed (sodium fixer; Kodak), and counterstained with hematoxylin (Shandon).

Cell culture and transfection

Human embryonic kidney HEK293 cells and mouse neuroblastoma Neuro2A cells were grown in MEM (Minimum Essential Medium Eagle; PAA) or DMEM (Dulbecco's modified Eagle's medium; PAA), respectively, supplemented with 10% fetal bovine serum (PAA), 100 U/ml penicillin (PAA) and 0.1 mg/ml streptomycin (PAA) at 37°C in 5% CO_2. For transfection of DNA constructs 0.375 µg DNA and 0.75 µl of LipoD293 reagent (SignaGen) were used per well of a 48-well plate or scaled up accordingly. In case of cotransfections, equal amounts of all plasmids were used. For transfection of siRNAs 24 pmol siRNA and 4 µl of Lipofectamine RNAiMAX (Invitrogen) were used per well of a 6-well plate. siRNAs were ordered from Ambion and their sequences are brought in Supporting Table S2.

Protein extracts and Western blotting

In vitro translation was performed using TnT Quick Coupled Transcription/Translation System (Promega) according to manufacturer's instructions. Cell and tissue extracts were prepared in RIPA buffer (50 mM Tris HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% Na-DOC, 0.1% SDS, 1 mM DTT, 1 mM PMSF, protease inhibitors cocktail Complete (Roche)). Protein concentrations were determined using BCA assay (Pierce). Equal amounts of proteins were separated in 8% SDS-PAGE and transferred to PVDF membrane (Biorad). For western blotting the antibodies were diluted in 2% skim milk and 0.1% Tween 20 in PBS as following: rabbit polyclone anti TCF4/ITF2 (CeMines) 1:1000, mouse monoclone anti E2 (Icosagen; 5E11) 1:5000, mouse monoclone anti tubulin β (Developmental Studies Hybridoma Bank) 20 ng/ml, HRP-conjugated goat anti mouse/rabbit IgG (Pierce) 1:5000. Chemiluminescent signal was detected using SuperSignal West Femto Chemiluminescent Substrate (Pierce).

Cyto- and histochemical immunostaining

For cytochemistry, cells were grown on poly-L-lysine (Sigma) coated coverslips and fixed with 4% paraformaldehyde for 15 minutes, treated with 50 mM NH_4Cl in PBS for 10 minutes and permeabilized in 0.5% Triton X-100 in PBS for 15 minutes. Cells were blocked with 2% bovine serum albumin (BSA) in PBS. The reactions with primary and secondary antibodies were carried out in 0.2% BSA and 0.1% Tween 20 in PBS at room temperature. For histochemistry, 16 µm coronal sections from fresh-frozen adult male human hippocampus and cerebellum were fixed with 4% paraformaldehyde for 30 minutes and blocked in 0.25% Triton X-100, 0.5% Tween 20, 3% goat serum in PBS. Primary antibody reactions were carried out in blocking buffer overnight at 4°C. The antibodies were diluted as following: rabbit polyclone anti TCF4/ITF2 (CeMines) 1:200, mouse monoclone anti V5 (Invitrogen) 1:500, mouse monoclone anti E2 (Icosagen, 5E11) 1:500, mouse monoclone anti NeuN 1:100 (Chemicon), mouse monoclone anti GFAP 1:800 (Chemicon), Alexa 488 or Alexa 568 conjugated goat anti mouse/rabbit IgG (Molecular Probes) 1:2000. The samples were mounted in ProLong Gold antifade reagent with DAPI (Molecular Probes) and analyzed by confocal microscopy (LSM Duo, Zeiss).

Luciferase assays

Cells on 48-well plates were lysed 24 hours post-transfection in 50 µl Passive Lysis Buffer (Promega). Dual-Glo Luciferase assay (Promega) was performed following manufacturer's instructions and luminescence was measured with GENios pro (Tecan) plate reader. For data analysis background signals from untransfected cells were subtracted and firefly luciferase signal values were normalized to *Renilla* luciferase signals. The obtained data were log-transformed, autoscaled, means and standard deviations (SD) were calculated and t-tests for analyses of statistical significance were performed. For graphical representation, the data were back-transformed to the original scale. Error bars represent upper and lower limits back-transformed as mean+SD and mean–SD, respectively. Correlation analysis was performed with tools at http://faculty.vassar.edu/lowry/pbcorr.html.

Supporting Information

Figure S1 Alignment of TCF4 isoforms. Amino acids different from the consensus are in blue. Localization of functional domains is indicated with lines above the sequence. Amino acid(s) in parentheses are absent from (1) isoforms coded by transcripts spliced at acceptor II of exon 8, (2) Δ isoforms coded by transcripts without exons 8–9, (3) isoforms coded by transcripts spliced at acceptor II of exon 15, (4) − isoforms (as opposed to + isoforms) coded by transcripts spliced at donor I of exon 18. AD, activation domain; NLS, nuclear localization signal; bHLH, basic helix-loop-helix domain.
(PDF)

Figure S2 Multiple alignment of *Homo sapiens* (h) genomic DNA regions containing *TCF4* 5′ exons and internal exons 1–2 with the respective regions in *Pan troglodytes* (c), *Mus musculus* (m), *Rattus*

norvegicus (r) and/or *Macaca mulatta* genomes (rh). The exons' names are given above each aligned region. The nucleotides in the aligned regions are numbered according to human genome assembly March 2006 NCBI36/hg18, chimp genome assembly March 2006 CGSC 2.1/panTro2, mouse genome assembly July 2007 NCBI37/mm9, rat genome assembly November 2004 Baylor 3.4/rn4 and rhesus genome assembly January 2006 MGSC merged 1.0/rheMac2. Alignments were produced with ClustalW and the percentages of identity were calculated between human TCF4 5′ exon sequence and the respective mouse sequence using Needleman-Wunsch global alignment. The exon sequences are in bold case, internal exons are in blue and sequences of primers used for expression analysis are underlined. Dotted blue lines above the sequences indicate transposable elements. Arabic numerals above exonic sequences indicate the number of human ESTs starting at the respective position and obtained from oligo-cap, cap-trapping and SMART libraries available in public databases through UCSC genome browser as of 3rd of November 2010. Possible in-frame translation start codons are shaded in gray. For each in-frame ATG codon NetStart translation start score is shown above the start codon. When needed the locations of alternative splice donor sites are indicated with blue roman numerals above the first intronic nucleotides (or above the first internal exon nucleotides in case of exon 7b-II).
(PDF)

Table S1 Accession numbers of representative mRNA or EST sequences for alternative *TCF4* transcripts and complete coding sequences of TCF4 isoforms cloned in full-length in this study.
(PDF)

Table S2 List of oligonucleotides used in this study.
(PDF)

Acknowledgments

We thank Enn Jõeste from North Estonian Regional Hospital, Tallinn, for help in collecting tissue samples, Epp Väli for technical assistance, Heiti Paves for advice on confocal microscopy and Priit Pruunsild for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: MS TT. Performed the experiments: MS KK AE MU. Analyzed the data: MS KK AE MU TT. Contributed reagents/materials/analysis tools: MS KK AE TT. Wrote the paper: MS TT.

References

1. Henthorn P, Kiledjian M, Kadesch T (1990) Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. Science 247: 467–470.
2. Corneliussen B, Thornell A, Hallberg B, Grundstrom T (1991) Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. J Virol 65: 6084–6093.
3. Massari ME, Murre C (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol 20: 429–440.
4. Ephrussi A, Church GM, Tonegawa S, Gilbert W (1985) B lineage–specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science 227: 134–140.
5. Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell 61: 49–59.
6. Saarikettu J, Sveshnikova N, Grundstrom T (2004) Calcium/calmodulin inhibition of transcriptional activity of E-proteins by prevention of their binding to DNA. J Biol Chem 279: 41004–41011.
7. Hauser J, Sveshnikova N, Wallenius A, Baradaran S, Saarikettu J, et al. (2008) B-cell receptor activation inhibits AID expression through calmodulin inhibition of E-proteins. Proc Natl Acad Sci U S A 105: 1267–1272.
8. Quong MW, Massari ME, Zwart R, Murre C (1993) A new transcriptional-activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells. Mol Cell Biol 13: 792–800.
9. Massari ME, Jennings PA, Murre C (1996) The AD1 transactivation domain of E2A contains a highly conserved helix which is required for its activity in both Saccharomyces cerevisiae and mammalian cells. Mol Cell Biol 16: 121–129.
10. Cronmiller C, Cline TW (1986) The relationship of relative gene dose to the complex phenotype of the daughterless locus in Drosophila. Dev Genet 7: 205–221.
11. Caudy M, Grell EH, Dambly-Chaudiere C, Ghysen A, Jan LY, et al. (1988) The maternal sex determination gene daughterless has zygotic activity necessary for the formation of peripheral neurons in Drosophila. Genes Dev 2: 843–852.
12. Zhuang Y, Cheng P, Weintraub H (1996) B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. Mol Cell Biol 16: 2898–2905.
13. Bergqvist I, Eriksson M, Saarikettu J, Eriksson B, Corneliussen B, et al. (2000) The basic helix-loop-helix transcription factor E2-2 is involved in T lymphocyte development. Eur J Immunol 30: 2857–2863.
14. Cisse B, Caton ML, Lehner M, Maeda T, Scheu S, et al. (2008) Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. Cell 135: 37–48.
15. Nagasawa M, Schmidlin H, Hazekamp MG, Schotte R, Blom B (2008) Development of human plasmacytoid dendritic cells depends on the combined action of the basic helix-loop-helix factor E2-2 and the Ets factor Spi-B. Eur J Immunol 38: 2389–2400.
16. Muir T, Sadler-Riggleman I, Stevens JD, Skinner MK (2006) Role of the basic helix-loop-helix protein ITF2 in the hormonal regulation of Sertoli cell differentiation. Mol Reprod Dev 73: 491–500.
17. Flora A, Garcia JJ, Thaller C, Zoghbi HY (2007) The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. Proc Natl Acad Sci U S A 104: 15382–15387.
18. Skerjanc IS, Truong J, Filion P, McBurney MW (1996) A splice variant of the ITF-2 transcript encodes a transcription factor that inhibits MyoD activity. J Biol Chem 271: 3555–3561.
19. Furumura M, Potterf SB, Toyofuku K, Matsunaga J, Muller J, et al. (2001) Involvement of ITF2 in the transcriptional regulation of melanogenic genes. J Biol Chem 276: 28147–28154.
20. Sobrado VR, Moreno-Bueno G, Cubillo E, Holt LJ, Nieto MA, et al. (2009) The class I bHLH factors E2-2A and E2-2B regulate EMT. J Cell Sci 122: 1014–1024.
21. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, et al. (2009) Common variants conferring risk of schizophrenia. Nature 460: 744–747.
22. Zweier C, Peippo MM, Hoyer J, Sousa S, Bottani A, et al. (2007) Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am J Hum Genet 80: 994–1001.
23. Brockschmidt A, Todt U, Ryu S, Hoischen A, Landwehr C, 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.
24. Amiel J, Rio M, de Pontual L, Redon R, Malan V, 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.
25. Maruyama K, Sugano S (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene 138: 171–174.
26. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, et al. (2006) Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res 16: 55–65.
27. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, et al. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38: 626–635.
28. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques 30: 892–897.
29. Kurooka H, Yokota Y (2005) Nucleo-cytoplasmic shuttling of Id2, a negative regulator of basic helix-loop-helix transcription factors. J Biol Chem 280: 4313–4320.
30. Gustincich S, Sandelin A, Plessy C, Katayama S, Simone R, et al. (2006) The complexity of the mammalian transcriptome. J Physiol 575: 321–332.
31. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456: 470–476.
32. Baek D, Davis C, Ewing B, Gordon D, Green P (2007) Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. Genome Res 17: 145–155.
33. Liu Y, Ray SK, Yang XQ, Luntz-Leybman V, Chiu IM (1998) A splice variant of E2-2 basic helix-loop-helix protein represses the brain-specific fibroblast growth factor 1 promoter through the binding to an imperfect E-box. J Biol Chem 273: 19269–19276.

34. Yoon SO, Chikaraishi DM (1994) Isolation of two E-box binding factors that interact with the rat tyrosine hydroxylase enhancer. J Biol Chem 269: 18453–18462.
35. Wu M, Li L, Sun Z (2007) Transposable element fragments in protein-coding regions and their contributions to human functional proteins. Gene 401: 165–171.
36. Jordan IK, Rogozin IB, Glazko GV, Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19: 68–72.
37. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 19: 530–536.
38. Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13: 335–340.
39. Zamudio N, Bourc'his D Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity 105: 92–104.
40. Pscherer A, Dorflinger U, Kirfel J, Gawlas K, Ruschoff J, 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.
41. Uittenbogaard M, Chiaramello A (2000) Differential expression patterns of the basic helix-loop-helix transcription factors during aging of the murine brain. Neurosci Lett 280: 95–98.
42. Soosaar A, Chiaramello A, Zuber MX, Neuman T (1994) Expression of basic-helix-loop-helix transcription factor ME2 during brain development and in the regions of neuronal plasticity in the adult brain. Brain Res Mol Brain Res 25: 176–180.
43. Chiaramello A, Soosaar A, Neuman T, Zuber MX (1995) Differential expression and distinct DNA-binding specificity of ME1a and ME2 suggest a unique role during differentiation and neuronal plasticity. Brain Res Mol Brain Res 29: 107–118.
44. de Pontual L, Mathieu Y, Golzio C, Rio M, Malan V, et al. (2009) Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. Hum Mutat 30: 669–676.
45. Kalscheuer VM, Feenstra I, Van Ravenswaaij-Arts CM, Smeets DF, Menzel C, 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.
46. Ravanpay AC, Olson JM (2008) E protein dosage influences brain development more than family member identity. J Neurosci Res 86: 1472–1481.
47. Brzozka MM, Radyushkin K, Wichert SP, Ehrenreich H, Rossner MJ (2010) Cognitive and sensorimotor gating impairments in transgenic mice overexpressing the schizophrenia susceptibility gene Tcf4 in the brain. Biol Psychiatry 68: 33–40.
48. Lingbeck JM, Trausch-Azar JS, Ciechanover A, Schwartz AL (2005) E12 and E47 modulate cellular localization and proteasome-mediated degradation of MyoD and Id1. Oncogene 24: 6376–6384.
49. Murre C (2005) Helix-loop-helix proteins and lymphocyte development. Nat Immunol 6: 1079–1086.
50. Qiu Y, Sharma A, Stein R (1998) p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. Mol Cell Biol 18: 2957–2964.
51. Massari ME, Grant PA, Pray-Grant MG, Berger SL, Workman JL, et al. (1999) A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. Mol Cell 4: 63–73.
52. Artandi SE, Merrell K, Avitahl N, Wong KK, Calame K (1995) TFE3 contains two activation domains, one acidic and the other proline-rich, that synergistically activate transcription. Nucleic Acids Res 23: 3865–3871.
53. Tanaka M, Clouston WM, Herr W (1994) The Oct-2 glutamine-rich and proline-rich activation domains can synergize with each other or duplicates of themselves to activate transcription. Mol Cell Biol 14: 6046–6055.
54. Wang D, Claus CL, Vaccarelli G, Braunstein M, Schmitt TM, et al. (2006) The basic helix-loop-helix transcription factor HEBAlt is expressed in pro-T cells and enhances the generation of T cell precursors. J Immunol 177: 109–119.
55. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: research0034.0031–research0034.0011.
56. Willems E, Leyns L, Vandesompele J (2008) Standardization of real-time PCR gene expression data from independent biological replicates. Anal Biochem 379: 127–129.
57. Timmusk T, Palm K, Metsis M, Reintam T, Paalme V, et al. (1993) Multiple promoters direct tissue-specific expression of the rat BDNF gene. Neuron 10: 475–489.

Supporting Figure S1. Alignment of TCF4 isoforms. Amino acids different from the consensus are in blue, localization of functional domains is indicated with lines above the sequence. Amino acid(s) in parentheses are absent from (1) isoforms coded by transcripts spliced at acceptor II of exon 8, (2) Δ isoforms coded by transcripts without exons 8-9, (3) isoforms coded by transcripts spliced at acceptor II of exon 15, (4) – isoforms (as opposed to + isoforms) coded by transcripts spliced at donor I of exon 18. AD, activation domain; NLS, nuclear localization signal; bHLH, basic helix-loop-helix domain.

Supporting Figure S2. Multiple alignment of *Homo sapiens* (h) genomic DNA regions containing *TCF4* 5' exons and internal exons 1-2 with the respective regions in *Pan troglodytes* (c), *Mus musculus* (m), *Rattus norvegicus* (r) and/or *Macaca mulatta* genomes (rh).

- The exons' names are given above each aligned region.
- The nucleotides in the aligned regions are numbered according to human genome assembly March 2006 NCBI36/hg18, chimp genome assembly March 2006 CGSC 2.1/panTro2, mouse genome assembly July 2007 NCBI37/mm9, rat genome assembly November 2004 Baylor 3.4/rn4 and rhesus genome assembly January 2006 MGSC merged 1.0 / rheMac2.
- The multiple alignments were produced with ClustalW program at http://www.ebi.ac.uk/Tools/clustalw2/index.html
- The percentages of identity were calculated between human TCF4 5' exon sequence and the respective mouse sequence using Needleman-Wunsch global alignment algorithm at http://www.ebi.ac.uk/Tools/emboss/align/.
- The **exon** sequences are in bold case.
- The **internal exons** are in blue.
- The sequences of primers used for expression analysis are underlined.
- Dotted blue lines above the sequences indicate the transposable elements as determined by RepeatMasker track in UCSC genome browser.
- The arabic numerals above exonic sequences indicate the number of human ESTs starting at the respective position and obtained from oligo-cap, cap-trapping and SMART libraries available in public databases through UCSC genome browser as of 3rd of November 2010.
- Possible in-frame translation start codons are shaded in gray. For each in-frame ATG codon translation start score calculated at http://www.cbs.dtu.dk/services/NetStart is shown above the start codon.
- When needed the locations of alternative splice donor sites are indicated with blue roman numerals (**I**, **II**) above the first intronic nucleotides (or above the first internal exon nucleotides in case of exon 7b-II).

TCF4 exon 1a

Aligned regions

h chr18:51482817-51483103
c chr18:52061786-52062072
rh chr18:48945911-48946209

note: exon 1a is located in ERV1 family LTR repeats containing region that is not conserved in mouse and rat; the position of both LTR12c and MER51a transposons is conserved in chimp, whereas only MER51a is present at this location in the rhesus macaque genome.

```
h               CAGGCTGAGG--GAGCCGGCTCCCGCCTCGGCCATCCCAGGAAGGGGCTCCCACAGTGCC
c               CAGGCTGAGG--GAGCCGGCTCCCGCCTCGGCCATCCCAGGAAGGGGCTCCCACAGTGCC
rh              CAGTCTGTGGTCAGTCTGATTAGTTGTTGGAGGGGGCCAATCAGAGGCTGAAACAAAGTT
                *** *** **     * *  *      * *      ***   ** ****   ***  *

h               GCGGCG----GGCTGAAGGGCTT--CTCAAGCGCCGCCAGAATTGGCA---CGGACGCCG
c               GCGGCG----GGCTGAAGGGCTT--CTCAAGCGCCGCCAGAATTGGCA---CGGACGCCG
rh              ACAAAGTTATACCCTATGCAATCAACTGATTGGTTGTGGAAAGTAACCAATCAGAGGCTG
                 *   *      *  * *   *   ** *   *  *    ** * *   * ** ** *
                       0.626                                          ← LTR12c
h               AGGAGGCATGAGAGCGAGCAAGGGCTCTGAGGGCTGCCAGCATGCTGTCACCT--CTCAA
c               AGGAGGCATTAGAGCGAGCAAGGGCTCTGAGGACTGCCAGCATGCTGTCACCT--CTCAA
rh              AAGTGAAGTTACAAAGTTATACTCCTATAAAAATGAAGACTTGGCCCACAACCAGCCTGA
                * * *   * * *  *    *   ** * *        *    **   ** *   *   *
                          MER51A ←
h               TTGGTTACAGGAGCAGGAGGGGACCAAACAGAGGTACTTTCAATTTTTCATCTGCCTCAC
c               TTGGTTACAGGAGCAGGAGGGGACCAAACAGAGGTACTTTCAATTTTTCATCTGCCTCAC
rh              TTGGTTACAGGAGCAGGAGGGGAAAAAACAGAGGTACTTTCAAGTTTTCATCTGCTGCAC
                ***********************  ****************** ***********  ***

h               AGAAAAGGGGGAGGTTTCAAAGGAAGTAGCCTCTGGTCCTTTTGTTACGTGGGCTTGGAA
c               AGAAAAGGGGGAGGTTTCAAAGGAAGTAGCCTCTGGTCCTTTTGTTACGTGGGCTTGGAA
rh              AGAAAAGGGGGAGGTT-CAAAGGGAATAGCTTCTGGTCCTTCTGTTACATGGGCTTGGGA
                **************** ****** * **** ********** ****** ********* *
```

TCF4 exon 1b

h chr18:51453879-51454330
c chr18:52029997-52030448
m chr18:69475444-69476010
r chr18:65989132-65989715

exon 1b sequence identity between h and m: 49%
note: exon 1b contains DNA transposon (MER5b) element.

```
h               ATTAATTCTCTTGGCAAGAATCTTTCTCTTATCTTGTTTGTTTACAACAATGCCAACAGC
c               ATTAATTCTCTTGGCAAGAATCTTTCTCTTATCTTGTTTGTTTACAACAATGCCAACAGC
m               GCTAACTCTCTTGCTAACAATCTGTCCCCAAACCTGACTGTTTTTGACAA-GCTACCAAA
r               ACTAAGTCTTTTGCTACCAATCTGTTCCCAAACCTGACT--TTATAACAA-GCTATCAAA
                  *** *** ***  *  ***** *  *  * * **  *  **   **** ** * **
                          1                      0.635
h               ATGCTGGGTTACACAGTAGACTCTTCCAGCTTTGCCATGGTAACACTTGTGCA----GCA
c               ATGCTGGGTTACACAGTAGACTCTTCCAGCTTTGCCATGGTAACACTTGTGCA----GCA
m               ATGCTAGGTTGTATC---------------TTCACAATGGTTGTGGTTGTGAAAGAAGCC
r               ATGCTAGCCTGTATCGTAGACTCTTTGAACTTCACGGTGGTTATAGACGCAAC----GCC
                ***** *  *  *                **  *  ****       *        **
                          1                3  1   1 1
h               TTCCTTTTCGAACAGTGCCCATTCTCTTGATGTCTACAGTATCACCTTTCTTGTAGATTG
c               TTCCTTTTCGAACAGTGCCCATTCTCTTGATGTCTACAGTATCACCTTTCTTGTAGATTG
m               TTCCTATTTGAAAAGTGCTCATTTCTTTGCTG-CTATAGGATTGTCTTTGTTGGAGAGCT
r               TTCCTATTTGAAAACTGCCCACTTCTTTGCTGGCCATAGGACTGCTTTTGTTGGAGAGCT
                ***** ** *** * *** ** *   *** ** * * ** *     *** *** ***
                                       0.676 1
h               GCACTTATGCGGCCAAAGGAACAACACCATGTTTTCTAAAAGGCTGGAGAAAATAGCACA
c               GCACTTATGTGGCCAAAGGAACAACATCATGTTTTCTAAAAGGCTGGAGAAAATACCACA
m               GTATGTGTAGAGCCGTTGGAATGAGTCCATGGTGTTTGACAGGCCAAA-AGAACATTATT
r               GTATACATAGAGCTAATGGAATGAGCCCATGATTTTTGATAGGCCAG------CATTATT
                * *    *   **    ****  *   **** * * * * ****         *  *
```

```
h                GGTGCCTCTCCTCTTTCCCTTCATCTTCATCATTTTGACAAAT--TACTCA---------
c                GGTGCCTCTCCTCTTTCCCTTCATCTTCATCATTTTGACAAAT--TACTCA---------
m                GGTACCTTACCTCTTTCCCTTTGTTCCCATCATTCAGAGCAGTGGTCCTCA-TCCTGTGG
r                GGTACTTTACCTCTTTCCCTTTTCTCCCATCATTCAGAGAAGTGGTTCTCAACCCCATGG
                 *** * *  ************      *******  **  * *  * ****

                                0.386                          MER5B →
h                --------------AAGATGGAGGG----------------TGCAG---TTGAAAGCCAG
c                --------------AAGATGGAGGT----------------TGCAG---TTGAAAGCCAG
m                G---------CTCTGGGGTAGGGGGGGTTGGATGAGTCTTTTGCAGGAGTGGCATATCAG
r                GTCATGACTGCTTTGGGAGGAAGGGGATTGAATGAGTCTTTCACAGGGGTGGCATATCAG
                               *     **                  ***   * * *   ***

h                CCAT---CATT----TTT----------TAAAACTTC--TCAGGACATTGTAACATG---
c                GCAT---CATT----TTT----------CAAAACTTC--TCAGGACATTGCAACATG---
m                ATATTTACATTATGATTCACAACAGTAGCAAAACTACAGTTAGGAAGTAGCAACAAAATC-
r                ATATTTACATCCCAACTTGTAACAGTAGTAAAGATACAGATATAAAGTAGCAACAAG-TA
                   **   ***      *            ***  * *    *  *  * * ****

h                -CACTTGGGTTGAGA---ACTGCTACTCGA---GCT-TCTCCAGGAG-----GCCCTTGG
c                -CATCGGGGTTGAGA---ACTGCTACTCGA---GCT-TCTCCAGGAG-----GCCCTTGG
m                ATTTTATAGTTGGGAGTCACCATGGCCTGAGGAGCTGTATGAAAGGGC---AGCCTTTGG
r                ATATTATAGCTGGGAGTCACCGTAGCATGA--AGCTATATTAAAGGGCTGCAGCATTTGG
                         * ** **   **     *  **   *** * *  * * *     **  ****

h                AGG--------------------------TAGGAAC----------CAG-----------
c                AGG--------------------------TAAGAAC----------CAG-----------
m                AAGGTTGA-GGCCGCTGATTATGAGGCGCTAAGAAAGTGGAGCATACAGATGAAAATAAA
r                AAGGTTAATGATTGCAGGTTAAGAGGAGCTAAGAAGGTGGTGGGTGCAGATGAAAATAAA
                 * *                          ** ***          ***

h                -----------------CATTTGGGCTACTTTAGCATCTTAGTAGAAACTGAAACCAA-A
c                -----------------CATTTGGGCTACTTTAGCATCTTAGTAGAAACTGAAACCAA-A
m                GTA-CCGGTTTGAAAACCTCTCAGGATACTTTAATGTCTAGCCAGAAGACAGAGCTATCA
r                GTATCCTTTTCAAAAACCTCTCAGGATATTTTAATGTCAAGCCAGAAGACAGAACTATCA
                                 *  *  ** ** ****   **     ****     * * *  *
```

TCF4 exon 1

Aligned regions

```
h        chr18:51449467-51449706
c        chr18:52025588-52025827
rh       chr18:48911757-48911996
```

note: internal exon 1 is located in SINE (AluSx) and LINE repeats containing region; this exon is used only together with 5’ exon 1b.

```
h                CATTTCTTCCCCAGACTACCAGTACCATCTTCTAACTTGTCTACCTGCCTTCATTCTAAC
c                CATTTCTTCCCCAGACTACCAGTACCATCTTCTAACTTGTCTACCTGCCTTCATTCTAAC
rh               CATTTCTTCCCCAGACTACTAGTACCATCTTCTAACTTGTCTACCTGCCTTCATTCTAAC
                 ****************** ****************************************

h                ATGTCCACCATCTTTTCTCCACCCTGCAGCAAATGTTTTGTAAACACCAATCTAAGAACA
c                ATGTCCACCATCTTTTCTCCACCCTGCAGCAAATGTTTTGTAAACACCAATCTAAGAACG
rh               AAGTCCACCATCTTTTCTCCACGCTGCAGCGAATGTTTTCTAAACACCAATCTAAGAACA
                 * ******************** ******* ******** ******************

                 → Line2B  AluSx →
h                TTATCTCTTGGACGGGCATGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAG
c                TTATCTCTTGGACGGGCATGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAG
rh               TTATCTCTTGGACGGGCATGGTGGCTCACGCCCGCAATTCCAGCACTTTGGGAGGCCAAG
                 ***************************** ** * *** ***********************

h                GTAGGTGGATCACCTGAGGTCAGGTGTTCAAGACAGCCTGGCCAACATGATGAAATCCTG
c                GTAGGTGGATCACCTGAGGTCAGGTGTTCAAGACAGCCTGGCCAACATGATGAAATCCTG
rh               GTAGGTGGATCACCTGAAGTCAGGTGTTCAAGACAGCCTGGCAAACATGATGAAATCCTG
                 ***************** ************************ *****************
```

TCF4 exon 2

Aligned regions
h chr18:53263228-53263515
c chr18:51988836-51989124
rh chr18:48870958-48871256
note: internal exon 2 is located in SINE (AluSz) repeat containing region.
this exon is used only together with 5′ exon 1b.

```
                                                                  AluSz ←
h               CATAGCTTTAAGTGACAGCCAATGCAATGAAGATTTCCATGTGCTGATTTTTTTCC----
c               CATAGCTTTAAGTGACAGCCAATGCAGTGAAGATTTCCATGTGCTGATTTTTTTCC----
rh              CATAGCTTTATGTGACAGCCAATGCAGAGAAGATTTGCATGTGCTGATTTTTTTTTTTT
                ********** ***************  ******** *****************

                        0.537
h               -------AGATGGAGTCTCATTCTGTCATCCAGACCGGAGTGCAGTGGCACCATCTCAGC
c               -------AGATGGAGTCTCATTCTGTCATCCAGACCGGAGTGCAGTGGCACCATCTCAGC
rh              TTTTTTGAGATGGAGTCTTGTTCTGTCATCCAGACTGGAGTGCAGTGGCACCATCTCGCC
                       ***********  *************** ********************  *

h               TCACTGCAACCTCCACCTCCCAGGTTCAAGCTATCCTCCTGCCTCAGCCTCCTGAGTACC
c               TCACTGCAACCTCCACCTCCCGGGTTCAAGCTATCCTCCCGCCTCAGCCTCCTGAGTACC
rh              TCACTGCAACCTCCACCTGCAGGGTTCAAGCTATTCTCCCGTCTCAGCCTCCTGAGTACC
                ****************** *  ************ **** * ******************

h               CGGGATAGGCGTGAGTCATCACTCGAGGCTAATTTTTGTATTTTTAGTAGAGACGGGGG-T
c               CGGGATAGGCGTGTGTCATCACTCGAGGCTAATTTTTGTATTTTTAGTAGAGACGGGGGT
rh              TGGGATCGGCATGTGCCATCACTCCAAGCTAATTTTCATATTTTTAGTAGACACGGGG-T
                 ***** *** ** * ******** * *********  ************* ****** *

h               TTCACCACGTTGGCCAGGCTGGTCACGAACTCCTGACCTCAAGTGATCCACCCACCTCGG
c               TTCACCACGTTGGCCAGTCTGGTCACGAACTCCTGACCTCAAGTGATCCACCCACCTCGG
rh              TTCACCATGTTGGCCAGGCTGGTCACAAACTCCTGACCTCAAGTGATCCACCCACCTCGG
                ******* ********* ******** **********************************
```

TCF4 exon 3a

Aligned regions
h chr18:51409263-51409608
c chr18:51983884-51984229
m chr18:69502792-69503123
r chr18:66016648-66016980
exon 3a sequence identity between h and m: 70%
note: a SINE (MIR3) element is located upstream of this exon

```
h               CTAGAGGTAGAGCCAGTGGAAT---TTCTCCTAACCCAACGTTTGTGATCTTTTCATTAC
c               CTAGAGGTAGAGCCAGTGGAAT---TTCTCCTAACCCAACGTTTGTGATCTTTTCATTAC
m               TTGGAGGTAGAGCCAATGGAATATTTTTTCCTACGCCAATGCTTGTGATCTTTTCATTAC
r               TTGGAGGTAGAGCCGATGGAATATTTTTTTCCAAGCCAATGTTTTTGATCTTTTCATTAC
                 * ***********  ******   ** * * *  **** * ** ***************

                      ← MIR3                                          1
h               A--------TCATCTAAGCAGAACCTGT-AGGGGTTAACTCTGAGAATTGGGGATGATGC
c               A--------TCATCTAAGCAGAACCTGT-AGGGGTTAACTCTGAGAATTGGGGATGATGC
m               A-GTGATTTTTTTCCCCCCAGAGCCTGTCGAGGGCTTACTCTGAGAACCAGAGATAATGT
r               ATGATTTTTTTTTTTTCCCAGAGCATGTCAAGGGCTTACTCTGAGAACCAGAGATAATGT
                *        *  *     **** * ***   *** * **********   * *** ***

h               TTGTGATAATGCAGCAAAAAGTAATGTCTAAATGTAAGATATTATTTTGAGATGAGAACA
c               TTGTGATAATGCAGCAAAAAGTAATGTCTAAATGTAAGATATTATTTTGAGATGAGAACA
m               CTGTGATAATGCACTATAAAGAGGTGGCTAAATGTAAAATATTATTTTAGGATGAAAATG
r               CTGCGGTAATGCAGCATAAAGAGGTGGCTAAATGTAAAATATTATTTTAGCATGCAAATG
                 ** * *******  * ****   ** ********** **********   ***  **

h               CCGAATGGGG-ATGGAAGACAGGAATCTTCACCATGTGTGAATGTACAGGAAACAACCAA
c               CCGAATGGGG-ATGGAAGACAGGAATCTTCACCATGTGTGAATGTACAGGAAACAACCAA
m               TGGAATGGGGGATAAAACACTGGAGTCCTCATGATGTGTGAA----------ATAGCCCA
r               TGGAATGGGGGATAAAACACTGGAATCTTCCTGATGTGTGAA----------ATAGCCCA
                  ******** **  ** ** *** ** **   *********          * * ** *
```

```
h               TAAATGCCTTAATTCTTGTTCTGTAAAAACTTACATGTAAGTAAAAGATCGCCGAAATTG
c               TAAATGCCTTAATTCTTGTTCTGTAAAAACTTACATGTAAGTAAAAGATCGCCGAAATTG
m               TAAATGTCCCAATTCTTGTCCTGTAAAA----ACTCTCAAGTAAAACATCTC--------
r               TAAATGTCCCAATTCTTGTCCTGTAAAAG---ACTCTCAAGTAAAACATCTC--------
                ****** *  ********* ********    **    ******** *** *

h               ATCTGAACTCTGTTACTG-CCACAAACTTTTATCAGCAAATGGAGATTAGGATGTTCCCA
c               ATCTGAACTCTGTTACTG-CCACAAACTTTTATCAGCAAATGGAGATTAGGATGTTTCCA
m               -----AATTCTGTCATGGGCCACGAGTTAAAATCAGAAGATAGAGATGAGGACGTTCTCA
r               -----AATTCTGTCACTG-CAGTGAATTAAAATCAGAAAATAGAGATGAGTACGTTCTCA
                     ** ***** *  * *    *  *   ***** * ** ***** ** * ***  **
```

TCF4 exon 3b

Aligned regions

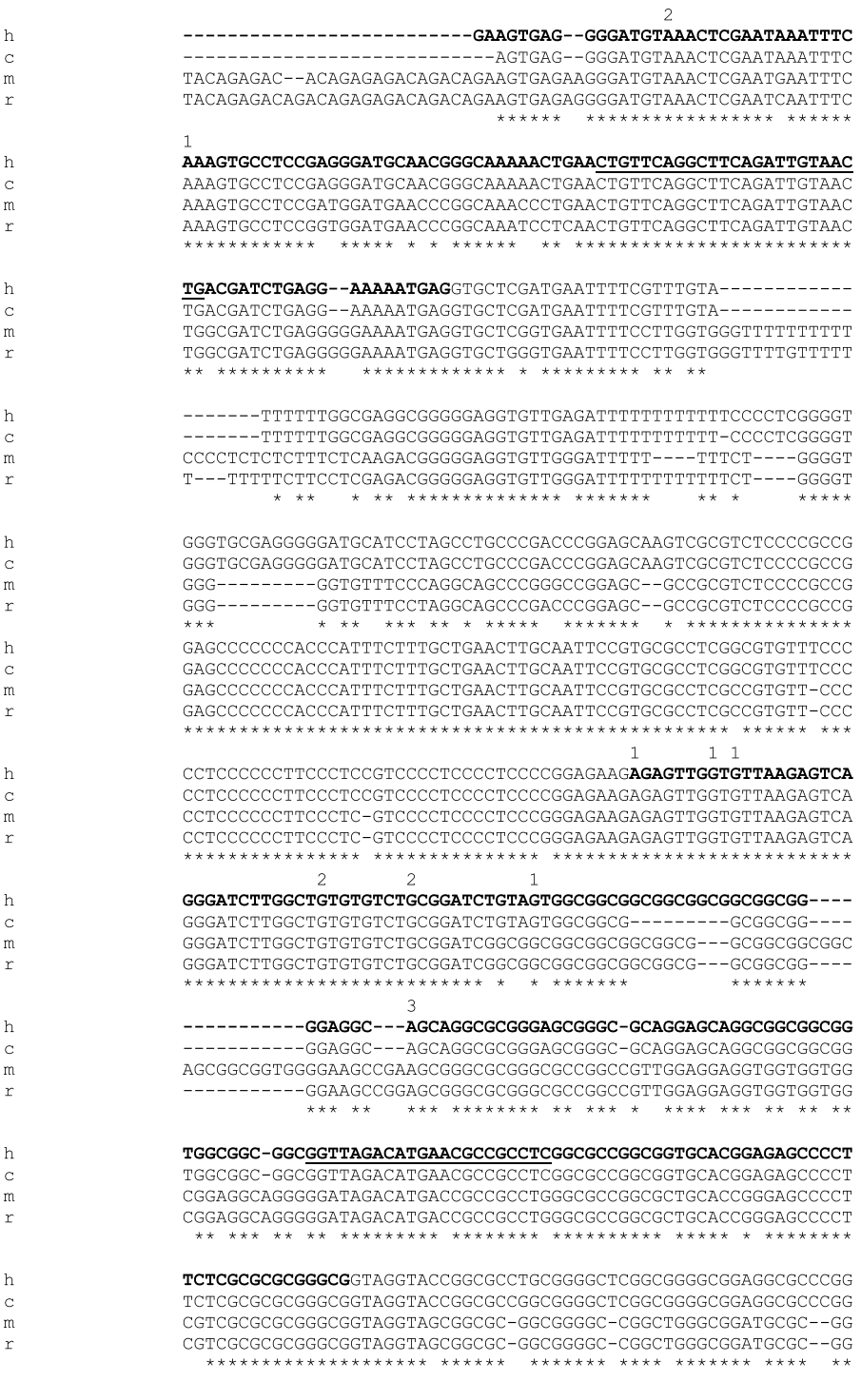
```
h       chr18:51407362-51408184
c       chr18:51981981-51982803
m       chr18:69504022-69504842
r       chr18:66017864-66018680
exon 3b sequence identity between h and m: 87%

h               CCGC------TCCCGGAGGAGGCGTGGTCGACCAGCACCGCCATCTTGGCCGCCTCTCTA
c               CCGC------TCCCGGAGGAGGCGTGGTCGACCAGCACCGCCATCTTGGCCGCCTCTCTA
m               CCGCCCTGGTTCCCGGAGGAGGCGTGGCCTAGA-GCGCCGCCATCTTGGCGGCCTCTCTA
r               CCGCCCCGGTTCCCGGAGGAGGCGTGGCTTAGC-GCGCCGCCATCTTGGCGGCCTCTCTA
                ****      ***************   *   ** ************ **********
                                                                        11
h               GGCGCTCTGTTTACCACTCTATGGTCGCCTGCTCCCGCCCCGCCCCTTCCCAGTC
c               GGCGCTCTGTTTACCACTCTATGGTCGCCCGCTCCTGCCCCGCCCCTTCCCAGTCATTGT
m               GGCGCTCTGTTTACCACTCTATGGTCGCGTGCTCCCGCCCCGCCCCTCCCCAGTCATTGT
r               GGCGCCGTGTTTACCACTCTATGGTCGCGCGCTCCCGCCCCGCCCCTCCCCAGTCATTGT
                *****  ********************  ***** *********** ************
                                          1 2  11   2614
h               CTGGAGGAGCAGCCGCGGCCGCAGCTCCTTCTCTTTATAAGCCCGCAGTTCCCGGATGTG
c               CTGGAGGAGCAGCCGCGGCCGCAGCTCCTTCTCTTTATAAGCCCGCAGTTCCCGGATGTG
m               CTGGTGGAGCAGCCGCGGCCGCAGCGCCTTCTCTTTATAAGCC-GCAGTGCCCGGATGTG
r               CTGGTGGAGCAGCCGCGGCCGCAGCGCCTTCTCTTTATAAGCC-GCAGTGCCCGGATGTG
                **** ******************** ***************** ***** **********
                   4211                                3 1   1
h               AATGGATTACAATGTATCTTTCAGGGAAACCTATTATTATCAATGTGACTCCACGGGGGA
c               AATGGATTACAATGTATCTTTCAGGGAAACCTATTATTATCAATGTGACTCCACGGGGGA
m               AATGGATTACAATGTATCTTTCAGGGAAACCTATTATTATCAATGTGACTCCTCGGGGGA
r               AATGGATTACAATGTATCTTTCAGGGAAACCTATTATTATCAATGTGACTCCTCGGGGGA
                ************************************************** *******
                          1
h               GTCCATGGTGATGATGATGAGGAGGAGGATGATGATGATGAGACACCTCTAAACTTGGAA
c               GTCCATGGTGATGATGATGAGGAGGAGGATGATGATGATGAGACACCTCTAAACTTGGAA
m               GTCAATGATGGTGTTGG------GGAGGAGGATGATGATGAGACGCCTCTAAACTTGGAA
r               GTCAATGATGGTGTTGA------GGATGATGATGATGATGAGACGCCTCTAAACTTGGAA
                *** *** ** ** **       *** ** ************** ***************

h               CAAGTTTAAGACTTTATGAGAGAAGA-----AAAAAAATCACCAACAAGAATTGTTTGAG
c               CAAGTTTAAGACTTTATGAGAGAAGA-----AAAAAAATCACCAACAAGAATTGTTTGAG
m               CAAGTTTAGGACTTTGAAAGAGAAGAGAAAAAAAAAATACAACCAACAAGA-----CCGAA
r               CAAGTTTAGGACTTTGAAAGAGAAGA----AAAAAATACAACCAACAAGA-----CCGAA
                ******** ******   ********     ***** *  **********       **

h               GAAAAATTATAACTATCCTGTGTTCATTTTTTTTTTTATAAACAATAAGAAAAAGTTGTT
c               GAAAAATTATAACTATCCTGTGTTGATTTTTTTT--TATAAACAATAAGAAAAAGTTGTT
m               GAACAATTATAACTATCCAGTGTTGATTATTTTT---ATAAACAATACGAAAAAGTTGTC
r               GAACAATTATAACTATCCTGTGTTGATTATTTTTT--ATAAACAATAAGAAAAAGTTGTC
                *** ************** ***** *** *****   ********** ***********
                                          1
h               GGATTTTTTTTT--AATGATTTCTTTTTTGGGGGAGGGAATTTTGTTGCAGTTTTATGGT
c               GGATTTTTTTTTTTAATGATTTCTTTTTTGGGGGAGGGAATTTTGTTGCAGTTTTATGGT
m               GGATTTTTTTTTTTAATGATTACTTTTTGGGGGGAGGGAATTTTGTTACAGTTTGATGAT
r               GGATTTTTTTTT---ATGATTACTTTTTTGGGGGAGGGGATTTTGTTTCAGTTTTATGAT
                ************   ****** ****** ********* ******** ****** *** *
```

```
h                GGAAAATGCAAAAACCAGAGCCAGGTGCATAATCTTGTAATCTGTGGATATCCCTGGAGC
c                GGAAAATGCAAAAACCAGAGCCAGGTGCATAATCTTGTAATCTGTGGATATCCCTGGAGC
m                GGAAAATGCAAAAACC-GAGCCAGGTGCATAATCTTGTAATCTGTGGCTAACCCTGGAAC
r                GGAAAATGCAAAAACCAGAGCCAAGTGCATAATCTTGTAATCTGTGGCTAACCCTGGAAC
                 ********************************************** ** ******* *
                                                                          1
h                AGGACTGAGTACCAGTTAAAATACTTTTT---GGGGATACACATGTGAGATACTAAGTAC
c                AGGACTGAGTACCAGTTAAAATACTTTTT---GGGGATACACATGTGAGATACTAAGTAC
m                AGGACTGACTTCTATTTAAAATACTCTTTTGGGGGGAACACTCATGTGAGACACTAAGTTC
r                AGGACTGACTACTATTTAAAATACTCTTTTGGGGGAATACACATGTGAGATACTAAGTAC
                 ******** * * * ********** ***   *** * ** ********* ******* *
                           1
h                TTGCAGAAGATTTTTGTCTCTCTTTTTAAAGTCTCTTTCCTTGGAATATTGTGAGAATAT
c                TTGCAGAAGATTTTTGTCTCTCTTTTTAAAGTCTCTTTCCTTGGAATATTGTGAGAATAT
m                TTGCAGAAGATTTTTGTCTCTCTTTTTAAAGTCTCTTTCCTTGGAATATTGTGAGCATAT
r                TTGCAGAAGATTTTTGTCTCTCTTTTTAAAGTCTCTTTCCTTGGAATATTGTGAGCATAT
                 ************************************************** ****

h                TTGTGGCCATTTAAGGTAACGTTTCAATTTGCCGTCAGAGTAAACTGTTTGTAATTGAAT
c                TTGTGGCCATTTAAGGTAACGTTTCAATTTGCCGTCAGAGTAAACTGTTTGTAATTGAAT
m                TTGTGGCCATTGAAGGTAACGTTACAATTTGCTGTCAGAGTAAACTGTTTGTAATTGAAC
r                TTGTGGCCATTGAAGGTAACGTTACAATTTGCTGTCAGAGTAAACTGTTTGTAATTGAAC
                 *********** *********** ******** **************************

h                TTAATTTTTAAAATGTCGAT-CCCGATGTTTTATTAAAACAAAAGGGACAACCTATTAAA
c                TTAATTTTTAAAATGTCGAT-CCCGATGTTTTATTAAAACAAAAGGGACAACCTATTAAA
m                TGAATTCTTAGAATGTTGAG-CCCAATGTTTTATTCAAACAAAAGGAACAACCTATTAAA
r                TGAATTTTAAGAATGTTGATTCCCAATGTTTTATTCAAACAAAAGGAACAACCGATTAAA
                 * **** * * ***** **  *** ********** ********** ****** ******

h                TAATTGCCAATGGGAAAAAGTGTGTGCATGCATTTTGTGTATGTACTCGCACATCTAGCA
c                TAATTGCCAATGAGAAAAAGTGTGTGCATGCATTTTGTGTATGTACTCGCACATCTAGCA
m                TAATTGCCAGTGGGAAAACGTGTGTGCATGCATTTCC-GTACGTACTCGTACATTTAGCA
r                TAATTGCCAGTGGGCAAACGTGTGTGCATGCATTTCG-GTGCGTACTCGCACATTTAGCA
                 ********* ** * *** ****************   **  ******* **** *****
```

TCF4 exons 3c, 3d

Aligned regions

```
h        chr18:51406110-51407003
c        chr18:51980740-51981619
m        chr18:69505196-69506161
r        chr18:66019053-66020010
exon 3c sequence identity between h and m: 63%
exon 3d sequence identity between h and m: 80%

h                TCG-CCCCCAACGCAAAACTACCGAAACGAAA-ATCTCCC-CAGGAAAA--TGTGGCTTG
c                TCG-CCCCCAACGCAAAACTACCGAAACGAAA-ATCTCCC-CAGGAAAA--TGTGGCTTG
m                TCGGCCCCAAACGCAAAACTATCCAAACAAAATATCCCCC-TAGTAAAAA-TGTGGATGG
r                TCGGCCCCAAACGCAAAACTATCCAAACAAAATATCCCCCCTAGTAAAAAATGTGGATGG
                 *** **** ************ * **** *** *** ***  ** ****  ***** * *

h                ATTTATAGACGAAAAAGAAAAATATTGTTCCTCTTGGCAAAGTCTTTTTGCATCACGTGT
c                ATTTATAGACGAAAAAGAAAAATATTGTTCCTTTTGGCAAAGTCTTTTTGCATCACGTGT
m                GTTTCTGTATGGAAAAGAAAAATATTGTTCGTCTTGGCAAAGTCTTTATGCATCACGTGT
r                CTTTCTGTATGGAAAAGAAAAATATTATTCGTCTTGGCAAAGTCTTTATGCATCACGTGT
                  *** *  * * ************** *** * ************** ************
                         1                    2          1
h                ATTTGCAGCCTAGTATAAACTTGCTTTGCCGTGTGTGGATGTGTG----------------
c                ATTTGCAGCCTAGTATAAACTTGCTTTGCCGTGTGTGGATGTGTG----------------
m                ATTTGCAGCAAAGCA-AAACTTGCTTTGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
r                ATTTGCAGCAAAGCA-AAACTTGCTTTGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
                 *********  ** * ************* *******  ******

h                ---------------------------AGTGAGAGGGAACGAGAGTAAGAGAAAGAAA--
c                ---------------------------AGTGAGAGGGAACGAGAGTAAGAGAAAGA----
m                GTGTGTGTGTGTGTGTGA--GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACA
r                GTGTGTGTGTGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT
                                     ** ***** **  *****  ***** ***
```

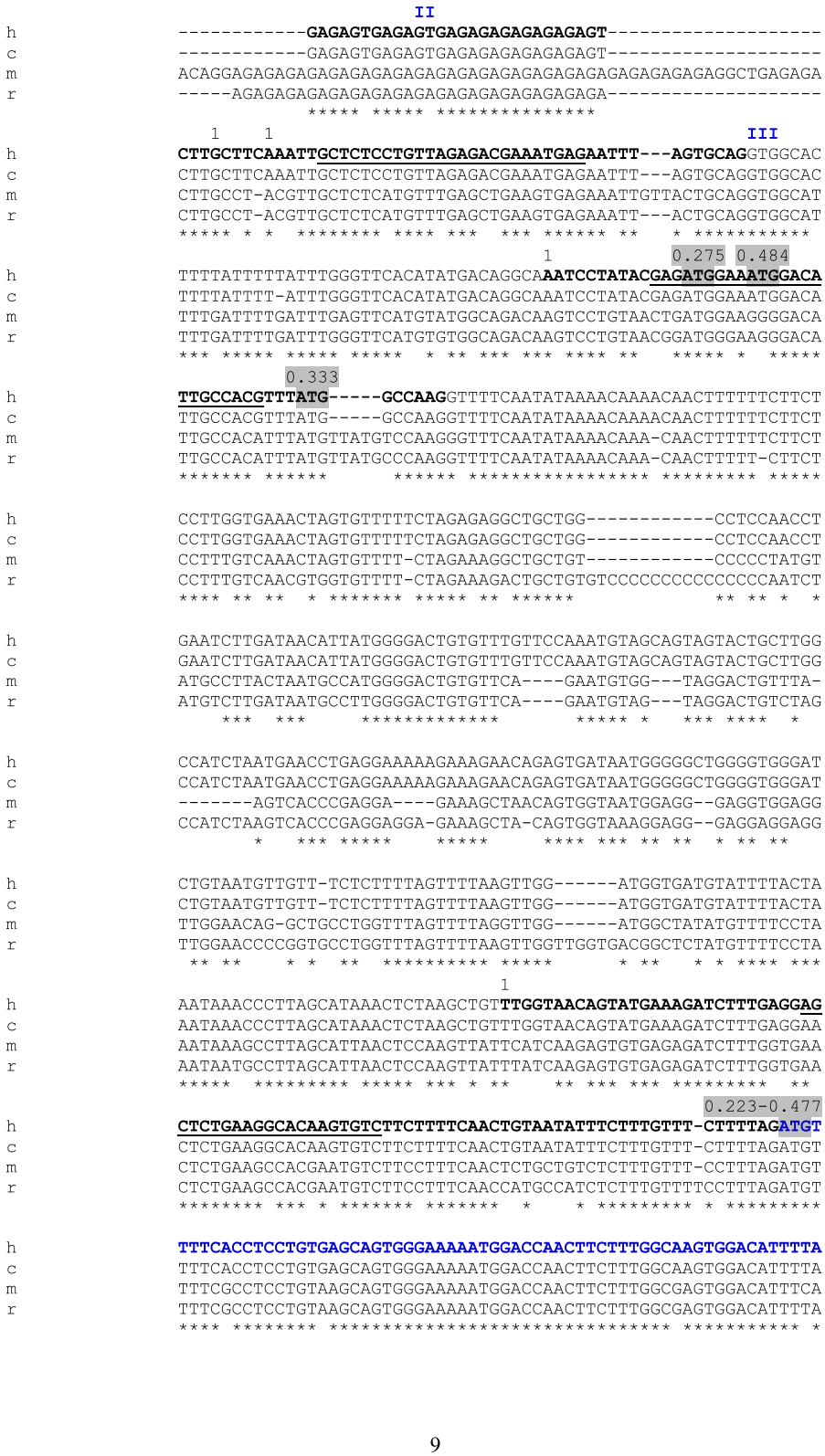


```
h               CGGCGCGGGGTT----CGGGCTCGGCGGCCCCGCACGCGGCTCCGCGCCTCCCG--CGCC
c               CGGCGCGGGGCT----CGGGCTCGGCGGCCCCGCACGCGGCTCCGCGCCTCCCG--CGCC
m               CGGGGCCCGGCCGGGCCGGGCTGGGCGATCGCGCTCG-GGCTCGGGGCCCCGTGGTCCCC
r               CGGGGCCGGGCCGGGCCGGGCTGGGCGGCCGCACACG-GGCTCGGGGCCCCGTGGTCCCG
                *** **  **      ****** ****  * * * ** ***** * *** *  *  * *

h               GCGGGCTCC--CGGCGCCCGGCGCTCCCAGAAGAGACACCCCTTCCCCTCCCGCCGCTTC
c               GCGGGCTCC--CGGCGCCCGGCGCTCCCAGAAGAGACACCCCTTCCCCTCCCGCCGCTTC
m               GCTGTCCCCGCCGAGGTTCGG-GAGCTCGGGAGCGACG-----------------GCGTC
r               GCTGTCCCCGCCGGGGTTCGG-GAGCTGGGGAGCGGCG-----------------GCGTC
                ** * * **  **  *  *** *  *   * ** * *                  ** **
```

TCF4 exons 4a (I,II,III), 4b, 4c, 4

Aligned regions

```
h       chr18:51403400-51404497
c       chr18:51978044-51979134
m       chr18:69507820-69508908
r       chr18:66021490-66022567
exon 4a sequence identity between h and m: 67%
exon 4b sequence identity between h and m: 73%
exon 4c sequence identity between h and m: 77%

h               AAATAAAATCCAAACCGCCTTCCAAGTGGGGCTCTTTCATGCTGCTGCTGCTGCTGCTGC
c               AAATAAAATCCAAACCGCCTTCCAAGTGGGGCTCTTTCATGCTGCTGCTGCTGCTGCTGC
m               AAATAAAATCCAAACCACCTTCGAAGTGGGGCTCCCCCACGCTGCTCCTGCTCC-ACTGC
r               AAATAAAATCCACACCACCTTCTAAGTGGGGCTCCCCCACTCTGCTCCCGCTCCTGCTGC
                ************ *** ***** ***********   **  ***** * *** *  ****

h               TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCCTCC
c               TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCCTG------CTCC
m               --CTGCCGCTGCTG--GCTGCCACCACTGCTGCT--------------------------
r               --CTGCCGAAGCTG--GCTGCCACCTCTGCTGCT--------------------------
                  **** *  ****  *****  *  ********
                                          1          3  21
h               TCCTCCTCCTCCTTCTCCTCCTCCTCCTCCTCTTCTAGACCTTCTTTTGGAGAAATGGCT
c               TCCTCCTCCTCCTTCTCCTCCTCCTCCTCCTCTTCTAGACCTTCTTTTGGAGAAATGGCT
m               ----------------------------CCTCTGGCAGACCCTGT---GGAGAAGTGGTT
r               ----------------------------CCTCCCGCAGACCCCGT---GGAGAAGTGGTT
                                           ****    *****   *   ****** *** *
                    1    12
h               TTCGG--AAGTTTTGCCAGGAAAC-GTAGCCCTAGGCAGGCAGCTTTGCAGCCCCCTTTC
c               TTCGG--AAGTTTTGCCAGGAAAC-GTAGCCCTAGGCAGGCAGCTTTGCAGCCCCCTTTC
m               CTCCGGAAAGTTTTGGCAGGAAGCAGTAGCCC---ATAGGCGGCTTTGCAGCCCCCTTTC
r               CTCTG-AAAGTTTTGGCAGGAAGCAGTGGCCC---ATAGGCGGCTTTGCAGCCCCCTTTC
                 ** *  ******** ****** * ** ****     **** ******************
                          1
h               TGCTTGTTGCACTTTCTCCATTCGTTCCTTTGCTTTTT-GCAGGCTCTGACTCAGGGAAG
c               TGCTTGTTGCACTTTCTCCCTTCGTTCCTTTGCTTTTT-GCAGGCTCTGACTCAGGGAAG
m               TGCTTGTTGCACTTTCTCCATTCGCTCTTTTGCTTTTTTGCAGGCTCTGACTCAGGGAAG
r               TGCTTGTTGCACTTTCTCCATTCGCTCCTTTGCTTTTTTGCAGGCTCTGACTCGGGGAAG
                ****************** **** ** ********** ************** ******
                I   1               1
h               GTGTGCATTATC-CACTAGATACGTCGAAGAAGAGGGAAACCAATTAGGGTCGAAATAAA
c               GTGTGCATTATC-CACTAGATACGTCGAAGAAGAGGGAAACCAATTAGGGTCGAAATAAA
m               GTGTGCGAAAACGCTTTAGAC-TGGAGGAGAAGAGGGAAATCAGTTAGGGCGGAAATAGA
r               GTGTGTGAAA-CGCATTAGAC-TGGAGGAGAAGAGGGAAACCAGTTAGGGCGGAAATAGA
                *****    * * *  ****   *  * ************ ** ******  ****** *
                0.223
h               TGCTGG------AGAGAGAGGGAGTGAAAGA----------------------------
c               TGCTGG------AGAGAGAGGGAGTGAAAGA----------------------------
m               TGCTGGAGAGTGAGAGAGAGGGAGTGAAAGATTGAGAGAGAGAGAGAGAGGGTGAGAGAG
r               TGCTGGAGAGAGAGAGAGAGGGAGTGAAAGATTGAGAGAGAGTGAGAGAG----------
                ******      ******************
```


II
III
0.275
0.484
0.333
0.223-0.477

```
h               CTGGCTCAAGTATGTAAATTCTTTTCTTAGTGT-TAAGTTAAATGTTCTGGGCTGTTTAT
c               CTGGCTCAAGTATGTAAATTCTTTTCTTAGTGT-TAAGTTAAATGTTCTGGGCTGTTTAT
m               CTGGCTCAAGTACGTGAATTCTTTTCTTGGTGTTGAAGTTCAATGGTCTCAGTTGTTTAT
r               CGGGCTCAAGTACGTGAATTCTTTTCTTGGTGT-GAAGTTCAATGTTCTGGGTTGTTTAT
                * ********** ** ************ ****  ***** **** ***  * *******

h               TTACTCAATTGTATGTTTTTGGCTCTGTTGAAGAGTTCTGAAA-AAGTCATTGAATTTTA
c               TTACTCAATTGTATGTTTTTGGCTCTGTTGAAGAGTTCTGAAA-AAGTCATTGAATTTTA
m               TTGCTCT-TTGTGTATTTTTGGCTCTGTTGGAGTGTTCTGAAAGAAGTCATTGAGTTTTA
r               TTGCTCT-TTGTGTATTTTTGGCTCTGTTGGAGTGTTCTGAAAGAAGTCATTGAGTTTTA
                ** ***  **** * *************** ** ********* ********** *****
```

TCF4 exon 5a (I,II)

Aligned regions

```
h       chr18:51364943-51365292
c       chr18:51938781-51939130
m       chr18:69544313-69544647
r       chr18:66056297-66056645
exon 5a sequence identity between h and m: 79%

h               GAATCTGTCTGGAATCGGCGCGGAACTTTGGAGAATGTTTTCTCCTCAGGAGCTGGAACT
c               GAATCTGTCTGGAATCGGCGCGGAACTTTGGAGAATGTTTTCTCCTCAGGAGCTGGAACT
m               GAATCTGTCTGCAACCAGCAGGGTGCACTGGGTG-CG-TTGGTCCTCAGGAGCTGGGTCT
r               GAATCTGTCTGGAACCAGCGGGGCGCATTGGGTACCAATTGCTCCTCAGGAGCTGGGCCT
                *********** ** * **  **  *  ***       **  **************  **

h               CAACTCTTCCTCA-GCCCACA-CAATGCAGGATCTGTGGTTTTCTGTCCCTGTTTCTTTT
c               CAACTCTTCCTCA-GCCCACA-CAATGCAGGATCTGTGGTTTTCTGTCCCTGTTTCTTTT
m               CCACACACCC----ACCCTCC---GCGCAGGATCCCAGAT---------CTCCTTTCTCT
r               CCACCCCTCCCGACACCCACCTCACTGCTGGATCTGAGAT---------CTAGTTTCTTT
                * ** *  **     *** *      ** *****   * *         **  **  * *
                          1
h               TCTCCATTTTTCTCCCCTTCACTTACTCTGATTGATGGAAACACGTGCT--GGGAGCTTT
c               TCTCCATTTTTCTCCCCTTCACTTACTCTGATTGATGGAAACACGTGCT--GGGAGCTTT
m               TCTC--TTTCCCCTCCCTTCCCTGGCTCTGATTGATGGAGACACGTGCT-GGAGAGCTCT
r               TCTCTCTTTCCCCTCCCTTCACTTGCTCTGATTGATGGAAACACGTGCCCGGGGAGCTTT
                ****  ***  *  ****** **  ************** ********   * ***** *
                                        I                                0.260
h               CCCTGACATTCTTTCCAAAGACTGAAAGTAAGAAATGTTATGGTGGTG------CTATGT
c               CCCTGACATTCTTTCCAAAGACTGAAAGTAAGAAATGTTATGGTGGTG------CTATGT
m               TCCTGACAATCTTCCCAAACACGGAGAGAGGGAGATGCTATGGTGGTGGTG---CTGGCT
r               CCCTGACAATCTTCCCAAACGCAGAAAGAAGGAGATGCTATGGTGGTGGTGGTGCTGGGT
                 ******* **** *****  * ** **   ** *** **********      **   *
                         II
h               CTATTTTTGAAAGTAAGTCTGTTAACGTAAAGCCTGTGTTGATATATTGGGATGAATTAT
c               CTATTTTTGAAAGTAAGTCTGTTAACGTAAAGCCTGTGTTGATATATTGGGATGAATTAT
m               CTATTTTTGAAAAGAAGACTGTTAATGTAAAGCCGGTGTTGATATATTCGGGTGAATTAT
r               CTATTTTTGAAAAGAAGGCTGTTAATGTAAAGCCGGTGTTGATATATTCGG-TGAATTAT
                ************  *** ******* ******** ************* ** ********

h               AATGGTGATATTATTACAGCCTTCAAAGCTGAAAATTTATTATTTTTCTCACCTACTTTT
c               AATGGTGATATTATTACAGCCTTCAAAGCTGAAAATTTATTATTTTTCTCACCTACTTTT
m               AATGGGGATATTATTACAGCCTTCCAAGCTGAAAATGTATTATTTTTCTCACCTACTTT-
r               AATGGGGTTATTATTACAGCCTTTCAAGCTGAAAATGTATTATTTTTCTCACCTACTTT-
                ***** * ***************  *********** **********************
```

TCF4 exon 5b

```
        Aligned regions
h       chr18:51328689-51329127
c       chr18:51902034-51902472
m       chr18:69575418-69575840
r       chr18:66088428-66088869
exon 5b sequence identity between h and m: 53%

h               TATTTTATTTTTGTTTCAGAAAAGGTCTGCTATTGTTCAAATTTCTAAT--GTAGCTCAT
c               TATTTTATTTTTGTTTCAGAAAAGGTCTGCTATTGTTCAAATTTCTAAT--GTAGCTCAT
m               TAGTTTATTTTTGTTGCGGTAAAGGTCTTCTATTGTTCTGATTTCTAACTCGCGGCTCAC
r               TATTTTATTTTTGTTTCGGTAAAGGTCTGCTATTGTTCAGATTTCTAAC--GCGGCTCAT
                ** ************ * * ******** *********  ********   *  *****

h               AAACACACAGCCAGTATCATTTCCCCTTCTACTGACTGATATATACAAATAAATTAAAAT
c               AAACACACAGCCAGTATCATTTCCCCTTCTACTGACTGATATATACAAATAAATTAAAAT
m               AAACACACAGTCCATATCCTCTCCCCTTCTCCTGACTGATGTGTAC-AATAAATGAAAAC
r               AAACACACAGCCAATATCCTCTCCCCTTCTCGTGACTGATGTATACCAACAAATGAAAAT
                ********** *  **** * *********  ******** * *** ** **** ****
                                  1  2  3    1
h               CTACGG------------TCATCATTAACTCATGTTCTGTTTCCATGGAGCACAGGAGAG
c               CTACGG------------TCATCATTAGCTCATGTTCTGTTTCCATGGAGCACAGGAGAG
m               CTCCAG---------TCATCGTCATTAGTCAGAGTTCTGTCTCTGTGAAGGGTAGAAAAG
r               CTCCCGCCGCCGCCGTCATCATCTTTAGTCAGAGATCGGTGTCTATGGACGGTAGGAAAA
                ** * *            ** ** ***      * ** ** **  ** *    ** * *
                                                                  1 1
h               TAACAAGGAAGGCCCCT-TAAAAGAAAGGCCTTTTAA-AGTGGAGGT-CTCAGTTCACAA
c               TAACAAGGAAGGCCCCT-TAAAAGAAAGGCCTTTTAA-AGTGGAGGT-CTCAGTTCACAA
m               GAACAAGGGAGACTTCTTCAGAAGAAAGTCCCTTTGAGAGCAGTGGTTCTCAGCCTGTAG
r               CAACAAGGGAGAGTCCT-CAGAAGAAAGTCCCTCTGAGAGCAGTGATTCTCAACCTGTGG
                 ******* **    **  * ******* ** * * * **  * * * ****

h               ACTTTAAAAGTTTACGAAGTAGTATTTTAGA-ATT--------------TTAA--TCACC
c               ACTTTAAAAGTTTACGAAGTAGTATTTTAGA-ATT--------------TTAA--TCACC
m               GTCGAGACCCCTTTGGTGGCTGAATATCAGATA----------------TTAGTGTTACA
r               GTCGTGACCCCTCCGGTGGCTGCGTGTCAGATAGCCAGCTTATCAGAAATTAGTGT-GCC
                      *    *   *  *  *  * * *** *                ***   *  *
                                                  0.379
h               AGATATATTGGGGGAATTTATTACTTCCTATTCATATGGAAGAGGACAGCAGAGGTAAAT
c               AGATATATTGGGGGAATTTATTACTTCCTATTCATATGGAAGAGGACAGCAGAGGTAAAT
m               ACACATGGCAGTAGCAAG-ATTGCC----ATT---ATGAAGTAG--CAACACAAATAATT
r               ATACAT--CGGTAGCAAG-ATCACA----GTT---ATGAAGCAG--CAGTAAAAATAATT
                * * **    *  * *   **  *      **   *** *  **  **  * *  *** *

h               ATTTATGTAACTTGTATTTTAAAAGCTTACTCTTGATTAAAATGTTTGATCTATCGACTT
c               ATTTATGTAACTTGTATTTTAAAAGCTTAGTCTTGATTAAAATGTTTGATCTATCGACTT
m               --TTATGA----------CTGGGGGGTCA-CCACAACACAGGGGACTGTTTTAAAGGGA-
r               --TTATGA----------TGGGGGGTTCA-CCACAACCCAGGGAACTGTTATAA-GGGT-
                  *****               * * *  *   *   *      ** * **  *

h               CTGTATCCTTAGAAGTTTGA-----ATTGCTGGACATATTTGACTGGGGTTTTTA--CAA
c               CTGAATCCTTAGAAGTTTGA-----ATTGCTGGACATATTTGACTGGGGTTTTTA--CAA
m               CATAGCATTTGGAAGGTTGAGAGACACTGCTG-------TGGAGGGAGATTCTTATCCGT
r               CAGAGCATTTGGAAGGTTGAGAGCCACTGCTA-------TAGAAGGAGATTCTAATCCAC
                *       ** **** ****     * ****        * **  * * ** * *  *
```

TCF4 exon 5c

Aligned regions

h chr18:51298148-51298550
c chr18:51871746-51872156
m chr18:69605796-69606177
r chr18:66119076-66119463

exon 5c sequence identity between h and m: 57%
note: exon 5c is located in SINE (MIRb) element containing region.

```
                                                                  MIRb ←
h               TAATATCCTCCCCCCGCCCCCCCCGCTTTTTTTTTTTTTTTTTTTTGGAGTTCCTACTGTG
c               TAGTATCCTCCCCCCGCCCCCCTGCTTTTTTTTTTTTTTT-------GAGTTCCTACTGTG
m               TAATAAAATTG----GGAGTTATGTCTATACAGTTAT----------AA-TCCTACTATG
r               TAATAAAATTG----GGAGTTATACCTATAGATTTTT----------AGCTCCTACTACG
                ** **   *      *          * *    ** *          *  *******  *
                                          2   1
h               TTCCAGGTTCCATGGTAACATTTTCCTTGTTATCTGTAATCATCACAACAGTACTGGAGG
c               TTCCAGGTTCCATGGTAACATTTTCCTTGTTATCTGTAATCATCACAACAGTACTGGAGG
m               TTCCAGGTTTTATGGCAACACTTTCC-------CTATACCTTTCATGACAAT-CCAGAGA
r               TTCAAGGTTTCATGGTAACACTTTCC-------CTATACTTTTCACAACAGT-CCAGAGA
                *** *****  **** **** *****       ** **    ***  *** * *  ***
                                                                   0.367
h               GTAAATACGGTCATTCCCATTTTAC-AGATGAGGATACATTCAATGT-GAGATGTTAAAG
c               GTAAATACGGTCATTCCCATTTTAC-AGATGAGGATACATTCAATGT-GAGATGTTAAAG
m               GCAGAAGCAGTGACTCCCATTTTGTTGGATGAGAATCCATACAGTAT-GAGATATCAAAA
r               GGACAAGCAGTCCTTACCATTTTGTTGGATGAGACGCCATACAGTACAGAGATATCAAAA
                * * *  * **   * *******    ******    *** ** *   ***** * ***
                0.389
h               ATGTTCTTCAGAATTTCCCAGGTAGCAGGCAGCTCA--------------TGGTATTCA
c               ACGTTCTTCAGAATTGCCCAGGTAGCAGGCAGCTCACCACACCACCATGGGTGGTATTCA
m               ATGTCCCCCAG--TTGTCCGCAAAGCAGGCAGCTCCGCACACTCTCGTGGGTGGTGCTCG
r               AGGTCCTCAAA--TTGTCCGCAAAGCAGGCAGCTGAGTACACTCTCGTGGGTGGTGTGGG
                * ** *   *   **  **    ***********                 ****

h               AAAGGTAATGGTTAAGAGAAAGAATGGATAATGAATATCTTGTGTCTATTTTCTTGAGAC
c               AAAGGTAATGGTTAAGAGAAAGAATGGATAATGAATATCTTGTGTCTATTTTCTTGAGAC
m               GAAGGTAATGGT--GGAAAGGCGACA----CTAAATGCATTCTGTTTGTTTTCTTGAGAC
r               AAAGGTAATGGT--GGAACGGCGATAGATCCTAAATGCATTCTGTTTATTTTCTTGAGAC
                 ***********   **      *       * ***   ** *** * ************

h               AGACGAGACTGAGCAGCCTGAGAACATTTGGTGCTCAGAGGGAGGAAATGCTAAGAGGGG
c               AGACGAGACTGAGCAGCCTGAGAACATTTGGTGCTCAGAGGGAGGAAATGCTAAGAGGGG
m               AGCC-----TGAGCAGCCTGAGAACATTTGGTGCTCAGAGGGAGGAAGCGCTGAGAGGGG
r               AGAC-----TGAGCAGCCTGAGAACATTTGGTGCTCAGAGGGAGGAAGTGCTGAGAGGAG
                ** *     ************************************  *** ***** *

h               AAAACAAGTCTGAAAGGGAAAGAAGGATAGTTCAAAGACCAGTTTCCCCTGCTTCAGAAT
c               AAAACAAGTCTGAAAGGGAAAGAAGGATAGTTCAAAGACCAGTTTCCCCTGCTTCACAAT
m               AAAACAAGTCTGAGTGG-AAAGGAGGGTAGTTCAAAGACCAGTTTCCCCTGCTTCACCGT
r               AAAACAAGTCTGAATGG-GAAGGAGGAGAGTTCAAAGACCAGTTTCCCCTGCGTCGTCGT
                *************  **  *** ***  ************************ **    *
```

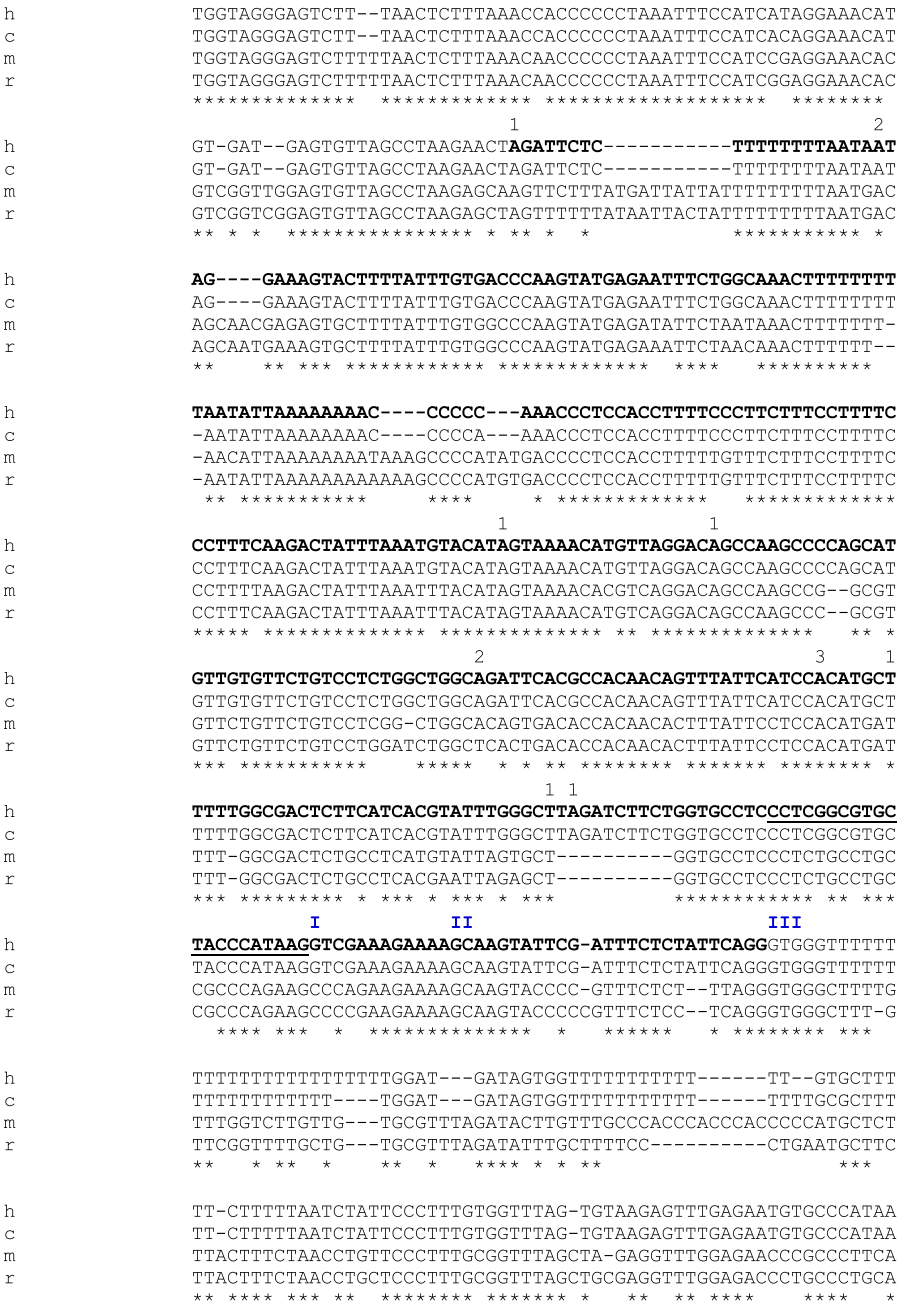
TCF4 exons 7a (I, II, III)

Aligned regions

h chr18:51240265-51240881
c chr18:51813659-51814272
m chr18:69660167-69660803
r chr18:66174462-66175075

exon 7a sequence identity between h and m: 76%

```
h               AACTTTTCAGGGAAGAAAAAAA--CACACACACACACTAAAAGCACCAGATGCAAATTAA
c               AACTTTTCAGGGAAGAAAAAAA--CACACACACACACTAAAAGCACCAGATGCAAATTAA
m               AACTTTTCAGGGAAGAAAAAAAAACACACACACACACTAAAAGCACCAGATGCAAATTAA
r               AACTTTTCAGGGAAGAAAAAAA--CACACACACACACTAAAAGCACCAGATGCAAATTAA
                **********************  ************************************
```


h TGGTAGGGAGTCTT--TAACTCTTTAAACCACCCCCCTAAATTTCCATCATAGGAAACAT
c TGGTAGGGAGTCTT--TAACTCTTTAAACCACCCCCCTAAATTTCCATCACAGGAAACAT
m TGGTAGGGAGTCTTTTTAACTCTTTAAACAACCCCCCTAAATTTCCATCCGAGGAAACAC
r TGGTAGGGAGTCTTTTTAACTCTTTAAACAACCCCCCTAAATTTCCATCGGAGGAAACAC
1 2
h GT-GAT--GAGTGTTAGCCTAAGAACTAGATTCTC-----------TTTTTTTTAATAAT
c GT-GAT--GAGTGTTAGCCTAAGAACTAGATTCTC-----------TTTTTTTTAATAAT
m GTCGGTTGGAGTGTTAGCCTAAGAGCAAGTTCTTTATGATTATTATTTTTTTTTAATGAC
r GTCGGTCGGAGTGTTAGCCTAAGAGCTAGTTTTTTATAATTACTATTTTTTTTTAATGAC
h AG----GAAAGTACTTTTATTTGTGACCCAAGTATGAGAATTTCTGGCAAACTTTTTTTT
c AG----GAAAGTACTTTTATTTGTGACCCAAGTATGAGAATTTCTGGCAAACTTTTTTTT
m AGCAACGAGAGTGCTTTTATTTGTGGCCCAAGTATGAGATATTCTAATAAACTTTTTTT-
r AGCAATGAAAGTGCTTTTATTTGTGGCCCAAGTATGAGAAATTCTAACAAACTTTTTT--
h TAATATTAAAAAAAAC----CCCCC---AAACCCTCCACCTTTTCCCTTCTTTCCTTTTC
c -AATATTAAAAAAAAC----CCCCA---AAACCCTCCACCTTTTCCCTTCTTTCCTTTTC
m -AACATTAAAAAAAATAAAGCCCCATATGACCCCTCCACCTTTTTGTTTCTTTCCTTTTC
r -AATATTAAAAAAAAAAAAGCCCCATGTGACCCCTCCACCTTTTTGTTTCTTTCCTTTTC
1 1
h CCTTTCAAGACTATTTAAATGTACATAGTAAAACATGTTAGGACAGCCAAGCCCCAGCAT
c CCTTTCAAGACTATTTAAATGTACATAGTAAAACATGTTAGGACAGCCAAGCCCCAGCAT
m CCTTTTAAGACTATTTAAATTTACATAGTAAAACACGTCAGGACAGCCAAGCCG--GCGT
r CCTTTCAAGACTATTTAAATTTACATAGTAAAACATGTCAGGACAGCCAAGCCC--GCGT
2 3 1
h GTTGTGTTCTGTCCTCTGGCTGGCAGATTCACGCCACAACAGTTTATTCATCCACATGCT
c GTTGTGTTCTGTCCTCTGGCTGGCAGATTCACGCCACAACAGTTTATTCATCCACATGCT
m GTTCTGTTCTGTCCTCGG-CTGGCACAGTGACACCACAACACTTTATTCCTCCACATGAT
r GTTCTGTTCTGTCCTGGATCTGGCTCACTGACACCACAACACTTTATTCCTCCACATGAT
1 1
h TTTTGGCGACTCTTCATCACGTATTTGGGCTTAGATCTTCTGGTGCCTCCCTCGGCGTGC
c TTTTGGCGACTCTTCATCACGTATTTGGGCTTAGATCTTCTGGTGCCTCCCTCGGCGTGC
m TTT-GGCGACTCTGCCTCATGTATTAGTGCT----------GGTGCCTCCCTCTGCCTGC
r TTT-GGCGACTCTGCCTCACGAATTAGAGCT----------GGTGCCTCCCTCTGCCTGC
I II III
h TACCCATAAGGTCGAAAGAAAAGCAAGTATTCG-ATTTCTCTATTCAGGGTGGGTTTTTT
c TACCCATAAGGTCGAAAGAAAAGCAAGTATTCG-ATTTCTCTATTCAGGGTGGGTTTTTT
m CGCCCAGAAGCCCAGAAGAAAAGCAAGTACCCC-GTTTCTCT--TTAGGGTGGGCTTTTG
r CGCCCAGAAGCCCCGAAGAAAAGCAAGTACCCCCGTTTCTCC--TCAGGGTGGGCTTT-G
h TTTTTTTTTTTTTTTTTGGAT---GATAGTGGTTTTTTTTTT------TT--GTGCTTT
c TTTTTTTTTTTT----TGGAT---GATAGTGGTTTTTTTTTT------TTTTGCGCTTT
m TTTGGTCTTGTTG---TGCGTTTAGATACTTGTTTGCCCACCCACCCACCCCCATGCTCT
r TTCGGTTTTGCTG---TGCGTTTAGATATTTGCTTTTCC----------CTGAATGCTTC
h TT-CTTTTTAATCTATTCCCTTTGTGGTTTAG-TGTAAGAGTTTGAGAATGTGCCCATAA
c TT-CTTTTTAATCTATTCCCTTTGTGGTTTAG-TGTAAGAGTTTGAGAATGTGCCCATAA
m TTACTTTCTAACCTGTTCCCTTTGCGGTTTAGCTA-GAGGTTTGGAGAACCCGCCCTTCA
r TTACTTTCTAACCTGCTCCCTTTGCGGTTTAGCTGCGAGGTTTGGAGACCCTGCCCTGCA

TCF4 exons 7b (I,II), 7

Aligned regions

h chr18:51221530-51222336
c chr18:51794937-51795741
m chr18:69679014-69679822
r chr18:66192823-66193608

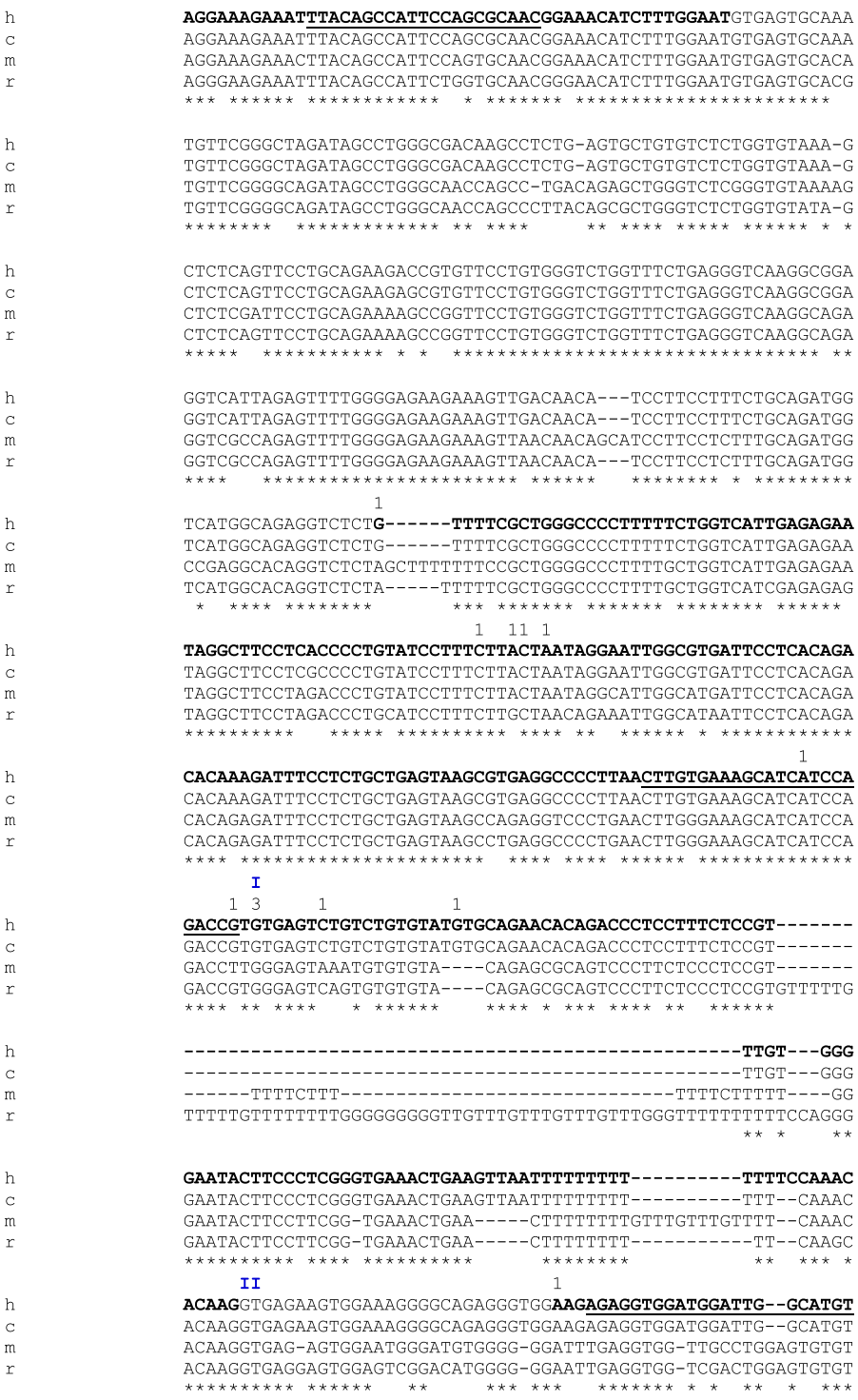
exon 7b sequence identity between h and m: 86%

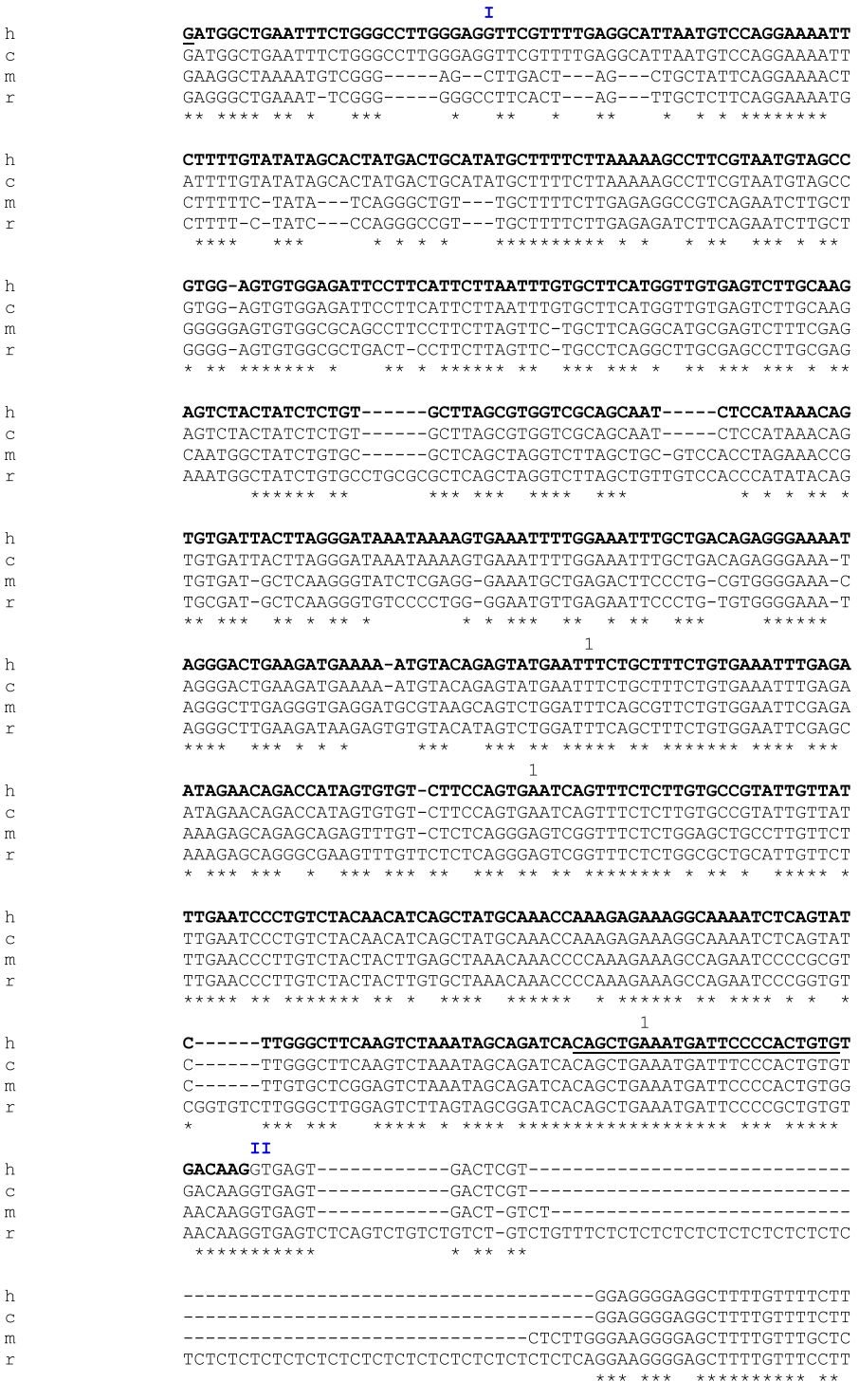
```
h               GTAGGCAATTTTTCAAAAGCCCTGTTTCCCAAATGTAAATTCCATCATCA-TCTTGATCA
c               GTAGGCAATTTTTCAAAAGCCCTGTTTCCCAAATGTAAATTCCATCATCA-TCTTGATCA
m               ATAGGTAATTTCCCCAA-ACCCCCTTTCCCAAGGGGG----GCCCCATCACTGCTGCTGG
r               CTAGGTGACTTCTCCAA-GCCCCCTTTCCCAAGTGTG----CCCT---------------
                 ****  * **  * **  ***  ********  *       *

h               TGATGGGGTTCACTCCTGGG-AAAGGGAGTTTTTAAAAAAGAG-AGAGAGAG--------
c               TGATGGGGTTCACTCCTGGG-AAAGGGAGTTTTTAAAAAAGAG-AGAGAGAG--------
m               TAATGGGGATCACTCCTGGGGAAAGGGAGTTTTTGAAAGGGAGGAGAGAGCAGTGGGGGG
r               ------GGATCAGTCCTGGGGAAAGGGAGTTTTTAAAAGGGAGGAGAGAGAGGTGGGGGG
                      ** *** ******* ************* ***  *** ******
                    2
h               -AGAGACTCAGAATTTTGAATAGTTTCTTTGCTAAATTCACCATCTGGAGAGCCTGTGAT
c               -AG--ACTCAGAATTTTGAATAGTTTCTTTGCTAAATTCACCATCTGGAGAGCCTGTGAT
m               AAGGGTCTTAGAATTTTGAACAGTTTCTTTGCGAGATTCACCATCTGGAGAGCCCATGAT
r               AGGGGGCTTAGAATTTTGAACAGTTTCTTTGCTAAATTCACCATCTGGAGAACCCATGAT
                  *   ** *********** *********** * ***************** **  ****

h               TGATTAGTTTTGGCCTCACTATTAAGTGTCAATATAATCAATGGGAATTACTAGGCTAGG
c               TGATTAGTTTTGGCCTCACTATTAAGTGTCAATATAATCAATGGGAATTGCTAGGCTAGG
m               TGATTAGCTTTGGCCTCCCTGTTAAGTGTCAATATAATCAATGGGAAGTACCAGGCTAGG
r               TGATTAGCCTTGGCCTCCCTGTTAAGTGTCAATATAATCAATGGGAAGTACCAGGCTAGG
                *******  ******** ** ************************* * * ********

h               GTCATTACATAGATAATTTGCCTGCTGTGTAAATAAAAATGAAATTGTTTAAATTTTGTG
c               GTCATTACATAGATAATTTGCCTGCTGTGTAAATAAAAATGAAATTGTTTAAATTTTGTG
m               GTCATTACATAGATCATTTGCCTGCTGTGTAAATAAAGATGAAATGGTCCAAATTTTGCG
r               GTCATTACATAGATCATTTGCCTGCTGTGTAAATAAAAGTGAAATGGTTCAAATTTGGTG
                ************** ********************  ****** **  ****** * *

h               CAAGAGAAAG----TGAAAATAA--GGAGTTACGATTTGTTTGTTA-GAGAAAGTGGAGG
c               CAAGAGAAAG----TGAAAATAA--GGAGTTACGATTTGTTTGTTA-GAGAAAGTGGAGG
m               AGAGAGAGAGAACGTGAAAAACAA-GGAGTGACAATTTGTTTGTTAAGAGAAAGTGGAGG
r               AGAGAGGAAGAACGTGAAAAAAAAAGGAATGACAATTTGTTTGTTA-GAGAAAGTGGAGG
                  ****  **    ******  *  *** * ** ************ **************
                  1                  11      2  1                 0.514
h               CCATTGGAATGACAGTTTTTGGAAGTGTGGAGCAGTTTGGCTAAGAATAGGAATGAAGGA
c               CCATTGGAATGACAGTTTTTGGAAGTGTGGAGCAGTTTGGCTAAGAATAGGAATGAAGGA
m               CCTCTGGAATGACAGTTTTTGGAAGTGTGGAGCAGTTTGGCTAAGAATAGGAATGAAGGA
r               CCACTGGAATGACAGTTTTTGGAAGTGTGGAGCAGTTTGGCTAAGAATAGGAATGAAAGA
                **  ************************************************** **
                1 11  1    (11)  1                  1       1 2 3
h               TATTTTTTT-CCAGTTTATCATAGCCAGAGTGAGGAAGTGTTATTCCCTCTCTTGCTTGC
c               TATTTTTTT-CCAGTTTATCATAGCCAGAGTGAGGAAGTGTTATTCCCTCTCTTGCTTGC
m               TATTTTTTTTCCAGTTTCTTGTCGCCAGAGTGAGGAAGCGCCATTCCCTGTCTTGCTTGC
r               TATTTTTTT-CCAGTTCCTTGTCGCCAGAGTGAGGAAGCGCCATTCCCTGTCTTGCTTGC
                ********* ******  *  * *************** *  ******* **********
                 11       2               I 1
h               ATACATTGCCAGTAGTGCCTACTTTACGTATGTAAACATCGGGGAAGAGCAGTAGATGTC
c               ATACATTGCCAGTAGTGCCTACTTTACGTATGTAAACATCGGGGAAGAGCAGTAGATGTC
m               ATACATTGCCAGTTGTGGCTCCCCTACGTATGTAAACATCGGAGCTGGGGAGTGGATCTC
r               ATACATTGCCAGTTGTGTCTCCTCTACGTACGTAAACATCCGAGATGGGGAGTGGATCTC
                ************* *** ** *  ****** ********* * *  * * *** *** **

h               TGTTACCTG--GAGGAGATTTTTTTTTAAGGAAAAACACAATTTACACTTTTTTGTCGTT
c               TGTTACCTG--GAGGAGATTTTTTTTTAAGGAAAAACACAATTTACACTTTTTTGTCGTT
m               TGTTACCTGCAGGGGAAGTTTGCTTTAGAGGAAAAGCACGATTCATCCTGTTCTGTTGTT
r               AGTTACCTGCAGGGGAGGTTTGCTTTAAAGGAAAAGCACGATTTATCCTGTTTTATTGTT
                 ********  * ***  ***  ***  ******* *** *** *  ** ** * * ***
```

```
                        II
h               TCTCTTGGCAGGTAAAACAGAAAGGGGCTCATACTCATCTTATGGGAGAGAATCAAACTT
c               TCTCTTGGCAGGTAAAACAGAAAGGGGCTCATACTCATCTTATGGGAGAGAATCAAACTT
m               TCTCTTGGCAGGTAAAACAGAAAGGGGCTCATACTCATCTTATGGGAGAGAA---AACGT
r               TCTCTCGGCAGGTAAAACAGAAAGGGGTTCATACTCATCTTATGGGAGAGAA---AACGT
                ***** ******************** ************************   *** *

h               ACAGGGTTGCCACCAGGTAAGTGAGAATCTCTGT-TGCGGAAGGGTTGAGAGGATTAACC
c               ACAGGGTTGCCACCAGGTAAGTGAGAATCTCTGT-TGCGGAAGGGTTGAGAGGATTAACC
m               TCAGGGTTGCCACCAGGTAAGTCCCACATT-GGGCTGGGGACGAGGTGAGAGGGCTAACC
r               ACAGGGTTGCCACCAGGTAAGTGCCACTCTCGGTCTGGGGAAGAGGTGCCAGAACTAAGG
                 ********************   *   *  *  ** *** * * **  **   ***

h               TGGTAAATCA--GTTTGTTTTCTAAATTAGATGATGCATTTTAAACTGTGACACCAAAGA
c               TGGTAAATCA--GTTTGTTTTCTAAATTAGATGATGCATTTTAAACTGTGACACCAAAGA
m               ATGGAAATGATTGCCTGCCTT----------------GCTTTCAGTGGTGACACTGAAGG
r               AGGGAAATGGT-GCCTGTCTT----------------GCTCTCAGTTGTGATACTAAAGG
                  * ****    *  **  **                 * * *   **** **  ***

h               TGCACCGGCAAG--TCAGA------TGATGGCTCTCTCTTTAAGTAATGAAATAACACAG
c               TGCACCGGCAAG--TCAGA------TGATGGCTCTCTCTTTAAGTAATGAAATAACACAG
m               ---CCAGGGAGGGTTGAGAGTCTCCTGATGGCTGTCT--TTAAGGAATGAGGCAAACCGG
r               ---ACAGGGAGGGCTGAGAGCCTC-TGACGGCTGTCT--TTAAGGAACGAAGCAAATCGG
                    * ** * *  * ***      *** **** ***  ***** ** **   **  * *
```

TCF4 exons 8a, 8b (I,II) and 8c (I,II)

```
        Aligned regions
h       chr18:51219075-51220606
c       chr18:51792487-51794015
m       chr18:69680798-69682354
r       chr18:66194620-66196287
exon 8a sequence identity between h and m: 83%
exon 8b sequence identity between h and m: 79%
exon 8c sequence identity between h and m: 69%

h               AA--ATGTAAGCAAAGTTTTGTTTTTGTAATTAAATCGAGCCTATGAGGAAAGTTTCCTT
c               AA--ATGTAAGCAAAGTTTTGTTTTTGTAATTAAATCGAGCCTATGAGGAAAGTTTCCTT
m               ACCCATGTAAGCCCACATGTGCTTCAGTAA-----CTGGGCAC-TGAGGGAAGTCTCTGT
r               AAC-ACGTAAGCCCACATGTGCCTCAGTAATGTAACTGGGCAC-TGAAGGAAGTTTTCCT
                *   * ******  *  * **  *  ****       * **   *** * **** *   *

h               TAC-GTACCAGACAT-----AGGAAGGTACGACTTCTGGTGGGTTGGT-------GTGTT
c               TAC-GTACCAGACAT-----AGGAAGGTACGACTTCTGGTGGGTTGGT-------GTGTT
m               ACCAGGAGCGCTCATTTCTCATGAAGGTA-GAATTCTGGAGGATTGGTTCTGGACGGGTT
r               TCCAGGAGCACCCGTTGATCATGAAGGTA-GAATTCTGGAGGATTGGTTCTGGAAGGGTT
                  * * * *   * *     * ******* ** ****** ** *****       * ***
                                                          1
h               TGTTTTT---TCCCTG------TATATATTTGCATTTTTAAAGTTTACTCCCTTGGCCCA
c               TGTTTTT---TCCCTG------TATATATTTGCATTTTTAAAGTTTACTCCCTTGGCCCA
m               TTCTTTCCCCTCCCCGCTGCTGTATATATTTGCATTTTAAAAGTTAACTCCTCTGGCCCA
r               TTCTTTCCCCTCCCCGCCG---TATATATTTGCATTTTAAAAGTTGACTCCTCTGGTCCA
                *  ***    **** *      ***************** ****** *****  *** ***
                   1           3         2
h               GCCACCAATTCAGTCATATTC-GGCATCTGAAAGATTTGTTGCATTGAGGTGAACAGTGC
c               GCCACCAATTCAGTCATATTC-GGCATCTGAAAGATTTGTTGCATTGAGGTGAACAGTGC
m               GTTGCCAATTCAGTCATACTCCAGCATCTAAAAGATTTGTTGCGATGAGCAGCGCAGCTC
r               GTTGCCAATTCAGTCATCTCCCAGCATCTAAAAGATTCGTGGCGATGAGCTGCACAGCTC
                *   *************   *  ****** ******* ** **  ****  *  ***  *
                      1
h               TTGGTTAAGAGCTCCTGCCGGAGTG-CAGTTCATGGCAGTTTGTTTGGAAAGAATTCTGC
c               TTGGTTAAGAGCTCCTGCCGGAGTG-CAGTTCATGGCAGTTTGTTTGGAAAGAATTCTGC
m               TGGGTTGAGAACTCTCGCAGGAGGGGCAGCCCATGGCAGTCTGTGTGGAAAGAATTCTGC
r               TTGGTTGAGAAGTCTTCCAGGAGGG-CAGCCCATGGCAGTTTGTGTGGAAAGAATTCTGC
                * **** ***  **   * **** * ***  ********* *** ***************
```


h AGGAAAGAAATTTACAGCCATTCCAGCGCAACGGAAACATCTTTGGAATGTGAGTGCAAA
c AGGAAAGAAATTTACAGCCATTCCAGCGCAACGGAAACATCTTTGGAATGTGAGTGCAAA
m AGGAAAGAAACTTACAGCCATTCCAGTGCAACGGAAACATCTTTGGAATGTGAGTGCACA
r AGGGAAGAAATTTACAGCCATTCTGGTGCAACGGGAACATCTTTGGAATGTGAGTGCACG
h TGTTCGGGCTAGATAGCCTGGGCGACAAGCCTCTG-AGTGCTGTGTCTCTGGTGTAAA-G
c TGTTCGGGCTAGATAGCCTGGGCGACAAGCCTCTG-AGTGCTGTGTCTCTGGTGTAAA-G
m TGTTCGGGGCAGATAGCCTGGGCAACCAGCC-TGACAGAGCTGGGTCTCGGGTGTAAAAG
r TGTTCGGGGCAGATAGCCTGGGCAACCAGCCCTTACAGCGCTGGGTCTCTGGTGTATA-G
h CTCTCAGTTCCTGCAGAAGACCGTGTTCCTGTGGGTCTGGTTTCTGAGGGTCAAGGCGGA
c CTCTCAGTTCCTGCAGAAGAGCGTGTTCCTGTGGGTCTGGTTTCTGAGGGTCAAGGCGGA
m CTCTCGATTCCTGCAGAAAAGCCGGTTCCTGTGGGTCTGGTTTCTGAGGGTCAAGGCAGA
r CTCTCAGTTCCTGCAGAAAAGCCGGTTCCTGTGGGTCTGGTTTCTGAGGGTCAAGGCAGA
h GGTCATTAGAGTTTTGGGGAGAAGAAAGTTGACAACA---TCCTTCCTTTCTGCAGATGG
c GGTCATTAGAGTTTTGGGGAGAAGAAAGTTGACAACA---TCCTTCCTTTCTGCAGATGG
m GGTCGCCAGAGTTTTGGGGAGAAGAAAGTTAACAACAGCATCCTTCCTCTTTGCAGATGG
r GGTCGCCAGAGTTTTGGGGAGAAGAAAGTTAACAACA---TCCTTCCTCTTTGCAGATGG
h TCATGGCAGAGGTCTCTG------TTTTCGCTGGGCCCCTTTTTCTGGTCATTGAGAGAA
c TCATGGCAGAGGTCTCTG------TTTTCGCTGGGCCCCTTTTTCTGGTCATTGAGAGAA
m CCGAGGCACAGGTCTCTAGCTTTTTTTCCGCTGGGGCCCTTTTGCTGGTCATTGAGAGAA
r TCATGGCACAGGTCTCTA-----TTTTTCGCTGGGCCCCTTTTGCTGGTCATCGAGAGAG
h TAGGCTTCCTCACCCCTGTATCCTTTCTTACTAATAGGAATTGGCGTGATTCCTCACAGA
c TAGGCTTCCTCGCCCCTGTATCCTTTCTTACTAATAGGAATTGGCGTGATTCCTCACAGA
m TAGGCTTCCTAGACCCTGTATCCTTTCTTACTAATAGGCATTGGCATGATTCCTCACAGA
r TAGGCTTCCTAGACCCTGCATCCTTTCTTGCTAACAGAAATTGGCATAATTCCTCACAGA
h CACAAAGATTTCCTCTGCTGAGTAAGCGTGAGGCCCCTTAACTTGTGAAAGCATCATCCA
c CACAAAGATTTCCTCTGCTGAGTAAGCGTGAGGCCCCTTAACTTGTGAAAGCATCATCCA
m CACAGAGATTTCCTCTGCTGAGTAAGCCAGAGGTCCCTGAACTTGGGAAAGCATCATCCA
r CACAGAGATTTCCTCTGCTGAGTAAGCCTGAGGCCCCTGAACTTGGGAAAGCATCATCCA
h GACCGTGTGAGTCTGTCTGTGTATGTGCAGAACACAGACCCTCCTTTCTCCGT-------
c GACCGTGTGAGTCTGTCTGTGTATGTGCAGAACACAGACCCTCCTTTCTCCGT-------
m GACCTTGGGAGTAAATGTGTGTA----CAGAGCGCAGTCCCTTCTCCCTCCGT-------
r GACCGTGGGAGTCAGTGTGTGTA----CAGAGCGCAGTCCCTTCTCCCTCCGTGTTTTTG
h ----------TTGT---GGG
c ----------TTGT---GGG
m ------TTTTCTTT----------TTTTCTTTTT----GG
r TTTTTGTTTTTTTTGGGGGGGGGTTGTTTGTTTGTTTGTTTGGGTTTTTTTTTTCCAGGG
h GAATACTTCCCTCGGGTGAAACTGAAGTTAATTTTTTTTT----------TTTTCCAAAC
c GAATACTTCCCTCGGGTGAAACTGAAGTTAATTTTTTTTT----------TTT--CAAAC
m GAATACTTCCTTCGG-TGAAACTGAA-----CTTTTTTTTGTTTGTTTGTTTT--CAAAC
r GAATACTTCCTTCGG-TGAAACTGAA-----CTTTTTTTT-----------TT--CAAGC
h ACAAGGTGAGAAGTGGAAAGGGGCAGAGGGTGGAAGAGAGGTGGATGGATTG--GCATGT
c ACAAGGTGAGAAGTGGAAAGGGGCAGAGGGTGGAAGAGAGGTGGATGGATTG--GCATGT
m ACAAGGTGAG-AGTGGAATGGGATGTGGGG-GGATTTGAGGTGG-TTGCCTGGAGTGTGT
r ACAAGGTGAGGAGTGGAGTCGGACATGGGG-GGAATTGAGGTGG-TCGACTGGAGTGTGT


I
h GATGGCTGAATTTCTGGGCCTTGGGAGGTTCGTTTTGAGGCATTAATGTCCAGGAAAATT
c GATGGCTGAATTTCTGGGCCTTGGGAGGTTCGTTTTGAGGCATTAATGTCCAGGAAAATT
m GAAGGCTAAAATGTCGGG-----AG--CTTGACT---AG---CTGCTATTCAGGAAAACT
r GAGGGCTGAAAT-TCGGG-----GGGCCTTCACT---AG---TTGCTCTTCAGGAAAATG
h CTTTTGTATATAGCACTATGACTGCATATGCTTTTCTTAAAAAGCCTTCGTAATGTAGCC
c ATTTTGTATATAGCACTATGACTGCATATGCTTTTCTTAAAAAGCCTTCGTAATGTAGCC
m CTTTTTC-TATA---TCAGGGCTGT---TGCTTTTCTTGAGAGGCCGTCAGAATCTTGCT
r CTTTT-C-TATC---CCAGGGCCGT---TGCTTTTCTTGAGAGATCTTCAGAATCTTGCT
h GTGG-AGTGTGGAGATTCCTTCATTCTTAATTTGTGCTTCATGGTTGTGAGTCTTGCAAG
c GTGG-AGTGTGGAGATTCCTTCATTCTTAATTTGTGCTTCATGGTTGTGAGTCTTGCAAG
m GGGGGAGTGTGGCGCAGCCTTCCTTCTTAGTTC-TGCTTCAGGCATGCGAGTCTTTCGAG
r GGGG-AGTGTGGCGCTGACT-CCTTCTTAGTTC-TGCCTCAGGCTTGCGAGCCTTGCGAG
h AGTCTACTATCTCTGT------GCTTAGCGTGGTCGCAGCAAT-----CTCCATAAACAG
c AGTCTACTATCTCTGT------GCTTAGCGTGGTCGCAGCAAT-----CTCCATAAACAG
m CAATGGCTATCTGTGC------GCTCAGCTAGGTCTTAGCTGC-GTCCACCTAGAAACCG
r AAATGGCTATCTGTGCCTGCGCGCTCAGCTAGGTCTTAGCTGTTGTCCACCCATATACAG
h TGTGATTACTTAGGGATAAATAAAAGTGAAATTTTGGAAATTTGCTGACAGAGGGAAAAT
c TGTGATTACTTAGGGATAAATAAAAGTGAAATTTTGGAAATTTGCTGACAGAGGGAAA-T
m TGTGAT-GCTCAAGGGTATCTCGAGG-GAAATGCTGAGACTTCCCTG-CGTGGGGAAA-C
r TGCGAT-GCTCAAGGGTGTCCCCTGG-GGAATGTTGAGAATTCCCTG-TGTGGGGAAA-T
1
h AGGGACTGAAGATGAAAA-ATGTACAGAGTATGAATTTCTGCTTTCTGTGAAATTTGAGA
c AGGGACTGAAGATGAAAA-ATGTACAGAGTATGAATTTCTGCTTTCTGTGAAATTTGAGA
m AGGGCTTGAGGGTGAGGATGCGTAAGCAGTCTGGATTTCAGCGTTCTGTGGAATTCGAGA
r AGGGCTTGAAGATAAGAGTGTGTACATAGTCTGGATTTCAGCTTTCTGTGGAATTCGAGC
1
h ATAGAACAGACCATAGTGTGT-CTTCCAGTGAATCAGTTTCTCTTGTGCCGTATTGTTAT
c ATAGAACAGACCATAGTGTGT-CTTCCAGTGAATCAGTTTCTCTTGTGCCGTATTGTTAT
m AAAGAGCAGAGCAGAGTTTGT-CTCTCAGGGAGTCGGTTTCTCTGGAGCTGCCTTGTTCT
r AAAGAGCAGGGCGAAGTTTGTTCTCTCAGGGAGTCGGTTTCTCTGGCGCTGCATTGTTCT
1
h TTGAATCCCTGTCTACAACATCAGCTATGCAAACCAAAGAGAAAGGCAAAATCTCAGTAT
c TTGAATCCCTGTCTACAACATCAGCTATGCAAACCAAAGAGAAAGGCAAAATCTCAGTAT
m TTGAACCCTTGTCTACTACTTGAGCTAAACAAACCCCAAAGAAAGCCAGAATCCCCGCGT
r TTGAACCCTTGTCTACTACTTGTGCTAAACAAACCCCAAAGAAAGCCAGAATCCCGGTGT
1
h C------TTGGGCTTCAAGTCTAAATAGCAGATCACAGCTGAAATGATTCCCCACTGTGT
c C------TTGGGCTTCAAGTCTAAATAGCAGATCACAGCTGAAATGATTTCCCACTGTGT
m C------TTGTGCTCGGAGTCTAAATAGCAGATCACAGCTGAAATGATTCCCCACTGTGG
r CGGTGTCTTGGGCTTGGAGTCTTAGTAGCGGATCACAGCTGAAATGATTCCCCGCTGTGT
II
h GACAAGGTGAGT------------GACTCGT------------------------------
c GACAAGGTGAGT------------GACTCGT------------------------------
m AACAAGGTGAGT------------GACT-GTCT-----------------------------
r AACAAGGTGAGTCTCAGTCTGTCTGTCT-GTCTGTTTCTCTCTCTCTCTCTCTCTCTC
h ---------------------------------------GGAGGGGAGGCTTTTGTTTTCTT
c ---------------------------------------GGAGGGGAGGCTTTTGTTTTCTT
m --------------------------------CTCTTGGGAAGGGGAGCTTTTGTTTGCTC
r TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAGGAAGGGGAGCTTTTGTTTCCTT

```
h               CTGTTTACTACATGATTGAGATACGGAATAGGAAACCTTCAGATGGTGTGCCCTAAGGAG
c               CTGTTTACTACATGATTGAGATACGGAATAGGAAACCTTCAGATGGTGTGCCCTAAGGAG
m               C-GTTTGCCA-ACTAGTGAGATGCTGGACAGAAAACCTTTAGATGGAGTAACCCCAGGAG
r               C-ATTCGCCA-GCTAGTGAGTCCCTGGACAGAAAACCTTTACATGGGGTAGTCCCAGGAG
                *  **  * *    * ****   * * * ** ******* * **** **   *  *****

h               CTCCAAAAGCTGCTTTATCGTATTATAACAGTGTA-TCCTATGGACTTAAAGGGGTGAGC
c               CTCCAAAAGCTGCTTTATCGTATTATAACAGTGTA-TCCTATGGACTTAAAGGGGTGAGC
m               CTGCAAAAGCTGCCTTTTCATATGATAACAATGTA-GCCTGTGAACCTAAAGGGATGAGC
r               CTGCAAA-GCTGCCTTTTCATATGATAACAATGTAAGCCTGTGAACCCAAAGGGATGGGC
                ** **** ***** ** ** *** ****** ****  *** ** **  ****** ** **
```

TCF4 exon 8d

Aligned regions

```
h       chr18:51169899-51170366
c       chr18:51742009-51742476
m       chr18:69722547-69723022
r       chr18:66238485-66238964
exon 8d sequence identity between h and m: 95%

h               GGGGCAGAGTCATAATTGGCCTATTTAGAGTATTTTGCATGTGAATAGTGTGTTAGTGAA
c               GGGGCAGAGTCATAATTGGCCTATTTAGAGTATTTTGCATGTGAATAGTGTGTTAGTGAA
m               GGGGCAGAGTCATAATTGGCCTATTTAGAGTATTTTGCATGTGAATAGTGTGTTAGTGAA
r               GGGGCAGAGTCATAATTGGGCTATTTAGAGTATTTTGCATGTGAATAGTGTGTTAGTGAA
                ****************** ****************************************

h               GCATCCCTAAGTCTGTTGTGAGTGACATGGTGATTAGAATTTAGATTAGGGAAAACACAT
c               GCATCCCTAAGTCTGTTGTGAGTGACATGGTGATTAGAATTTAGATTAGGGAAAACACAT
m               GCATCCCTAAGTCTGTTGTGAGTGACATGGTGATTAGAATTTAGATTAGGGAAAACACGT
r               GCATCCCTAAGTCTGTTGTGAGTGACACGGTGATTAGAATTTAGATTAGGGAAAACACAT
                ************************** ****************************** *
                                              1   1
h               TCTGTACTTTATCAAGGACAGTAATTAGTTCAGTCAGTAAAAGTTGATTACAAGTAACGA
c               TCTGTACTTTATCAAGGACAGTAATTAGTTCAGTCAGTAAAAGTTGATTACAAGTAACGA
m               TCTGTACTTTATCAAGGACAGTAATTAGTTCAGTCAGTAAAAGTTGATTACAAGTAACGA
r               TCTGTACTTTATCAAGGACAGTAATTAGTTCAGTCAGTAAAAGTTGATTACAAGTAACGA
                ************************************************************

h               TAAAGTTAGATTTTT-AGTACTTAATTTTAAAATTTCTTTATTAAATAGTAAATTTGTTC
c               TAAAGTTAGATTTTT-AGTACTTAATTTTAAAATTTCTTTATTAAATAGTAAATTTGTTC
m               TAAAGTTAGATTTTT-AGTACTTAATTTTAAAATTTCTTTATTAAATAGCAAATTTGTCC
r               TAAAGTTAGATTTTTTAGTACTTAATTTTAAAATTTCTTTATTAAATAGCAAATTTGTTC
                *************** ********************************** ******** *

h               TTAATATAAATGGTACTATCTACTCTTGTTTTATTGCATATCAGAATAATTAAAAGAACA
c               TTAATATAAATGGTACTATCTACTCTTGTTTTATTGCATATCAGAATAATTAAAAGAACA
m               TTAATATAAATGGTACTATCTACACTTGTTTTATTGCCTATCAGAATAATTTAAAGAACA
r               TTAATATAAACGGTGCTCTCTACACTCGTTTTATTGCCTATCAGAATAATTTAAAGGACA
                ********** *** ** ***** ** ********** ************* **** ***

h               ATGTGTTCAATGCAGAGTATAATCAAGCGATTTATATATTATGGGTAAGTTTCTATGTAG
c               ATGTGTTCAATGCAGAGTATAATCAGGCGATTTATATATTATGGGTAAGTTTCTATGTAG
m               ATGTGTCCAATGCAGAGTATAATCAGGCGGTTTGTGTATTGTGGGTAAGTTCCTATGTAG
r               ATGTGTCCAAAGCACAGCGTAACCAGGCGGTTTATGTATTGTGGGTAAGTTCCTAGGTAG
                ****** *** *** **  *** ** *** *** * **** ********** *** ****

h               TTTTTAGAAAATGTAATATTATATAATCTTTCTGATGTCATCATTTTT-AGAAGTTTAAT
c               TTTTTAGAAAATGTAATATTATATAATCTTTCTGATGTCATCATTTTT-AGAAGTTTAAT
m               CTTTTAGGAAATGCAATATTATGTAATCTGTCTTATGTCATCATTTTTCAAAAGCTTAGC
r               CTTTTAGGAAACATAATATTATGTAATCCGTCTTATGTCATCATTTTTCGAAAGTTTAGC
                 ****** ***   ******** *****  *** ***************   *** ***

h               AATT-GTCACACTGT---GATAGGT------TATTGTGGAAGTTTTCTAGTGTTCAGGTA
c               AATT-GTCACACTGT---GATAGGT------TATTGTGGAAGTTTTCTAGTGTTCAGGTA
m               AATTTTTCACACCGT---GACAACTTGATGATGTTGTGGGAATGTTCTAGTGTTCAGGTA
r               AATTTGTCACACCGTCGTGACAGCTCAATGATGTTGTGGGAACGTTCTAGTGTTCAGGTA
                ****  ****** **   ** *  *      * ****** *   ****************
```

TCF4 exon 10a

Aligned regions

```
h       chr18:51139749-51140227
c       chr18:51711821-51712299
m       chr18:69752920-69753388
r       chr18:66269376-66269845
exon 10a sequence identity between h and m: 99%

h               TCTCGCGTATGCATTAGGATTTCGAGCTCCATCAGGGCCGGGCTGCCTTCCAGCTCCGCT
c               TCTCGCGTATGCATTAGGATTTCGAGCTCCATCAGGGCCGGGCTGCCTTCCAGCTCCGCT
m               TCTCGAGTGCTC----GGCTGCGGCGCTGCAC--AGGCCGGGCCGCCATCCAGC----CT
r               TCTCGAGTGCTC----GGCTGCGGCGCTGCAC--AGGCCGGGCCGCCATCCAGC----CT
                ***** **   *    ** *   * *** **    ******** *** ******    **

h               GCGCAAAGTTGAGCGTCTGACAGCAGCGCCGCGGCCTCCCCC-GCCCGCCAGGAATGGTC
c               GCGCAAAGTTGAGCGTCTGACAGCAGCGCCGCGGCCTCCCCC-GCCCGCCAGGAATGGTC
m               GCGCGAACTTGAGCGTCTGACAGCAGCGCCGGGGCCTCCCCC-GCCCGCCAGGAATGGTC
r               GCGCGAACTTGAGCGTCTGACAGCAGCGCCGCGGCCTCCCCCCGCCCGCCAGGAATGGTC
                **** ** ********************** ********** *****************
                      1             4                1412   13      4     1
h               TCTTCC[bold]TGCTTTGCATATTCACCACGCTTGGCCCGGCCATATGGGAAAATGCAACTGAAG[/bold]
c               TCTTCCTGCTTTGCATATTCACCACGCTTGGCCCGGCCATATGGGAAAATGCAACTGAAG
m               TCTTCCTGCTTTGCATATTCACCACGCTTGGCCCGGCCATATGGGAAAATGCAACTGAAG
r               TCTTCCTGCTTTGCATATTCACCACGCTTGGCCCGGCCATATGGGAAAATGCAACTGAAG
                ************************************************************
                     3 12 21  2                                  1  1
h               [bold]GAATTGTTGTGAAAAAAGGGACAGCGAGTTTGAAATAAAAGTTGTTAGAGTGGTACTGAG[/bold]
c               GAATTGTTGTGAAAAAAGGGACAGCGAGTTTGAAATAAAAGTTGTTAGAGTGGTACTGAG
m               GAATTGTTGTGAAAAAAGGGACAGCGAGTTTGAAATAAAAGTTGTTAGAGTGGTACTGAG
r               GAATTGTTGTGAAAAAAGGGACAGCGAGTTTGAAATAAAAGTTGTTAGAGTGGTACTGAG
                ************************************************************
                   2 1              0.726     1                 0.491     1
h               [bold]GAGAAAAAAGAATCCGAGACCATGTACTGCGCATACACAATCCCGGGCATGGGCGGCAAC[/bold]
c               GAGAAAAAAGAATCCGAGACCATGTACTGCGCATACACAATCCCGGGCATGGGCGGCAAC
m               GAGAAAAAAGAATCCGAGGCCATGTACTGCGCATACACCATCCCGGGCATGGGCGGCAAC
r               GAGAAAAAAGAATCCGAGGCCATGTACTGCGCATACACCATCCCGGGCATGGGCGGCAAC
                ****************** ****************** **********************
                     4 0.440
h               [bold]TCTTTGATGTACTACTATAATGGGAAAGCG[/bold]GTAAGCGAATGCGGGCTGCGAGGGGGATTT
c               TCTTTGATGTACTACTATAATGGGAAAGCGGTAAGCGAATGCGGGCTGCGAGGGGGATTT
m               TCTTTGATGTACTACTATAATGGGAAAGCGGTAAGCGAACTCGGGCTGCGAGGGGCATTG
r               TCTTTGATGTACTACTATAATGGGAAAGCGGTAAGCGAACTCGGGCTGCGAGGGGCATTG
                ************************************  ************** ***

h               TTGACAAGTTTTATTGTCGCAGGGATGAGCATCCTAGGTGGTGATCATTGTTTGCAAAGT
c               TTGACAAGTTTTATTGTCGCAGGGATGAGCATCCTAGGTGGTGATCATTGTTTGCAAAGT
m               GGGACAAGTTTTATGGTCACAGGGATGAGCCTTCTAGGCGAGGAGCATTGTTTGCTAAGT
r               GGGACAAGTTTTATGGTCACAGGGATGAGCCTTCTAGGCGGGGAGCATTGTTTGCGAAGT
                  ************ *** *********** * ***** *  ** ********** ****

h               TGTGCAGGACTGCCGCACCTTCCTGGGTGCCATGAAATATTTATCATTTGCCAATCCTAA
c               TGTGCAGGACTGCCGCACCTTCCTGGGTGCCATGAAATATTTATCATTTGCCAATCCTAA
m               TGTGCAGGACTGCCGCACCTTTCTCGGTACCATGAAATATTTATCATCTGCGAGTCCTAG
r               TGTGCAGGACTGCCGCACCTTTCTCGGTGCCATGAAATATTTATCATCTGCGAGTCCTAG
                ********************** ** *** ****************** *** * *****
```

TCF4 exon 10b

Aligned regions

```
h       chr18:51139327-51139667
c       chr18:51711398-51711739
m       chr18:69753455-69753790
r       chr18:66269909-66270246
exon 10b sequence identity between h and m: 88%
```

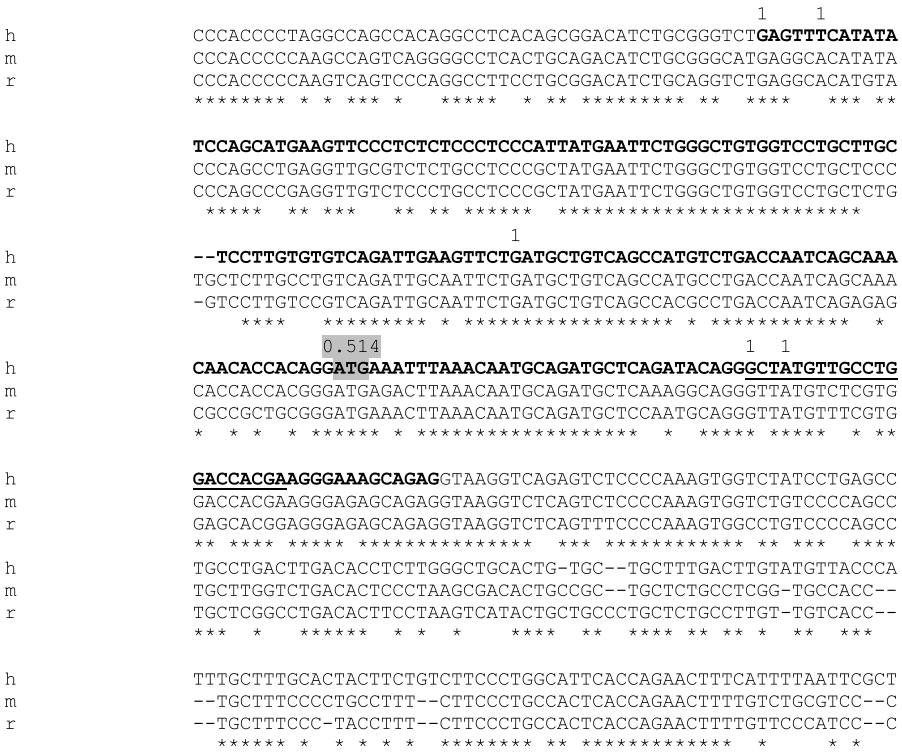
```
h               GAGCTTTAGAGA-GTGGATCCTAGAGCGAAA-GTTTGCAGTCAGGTAAAGTTGAGAGGTG
c               GAGCTTTAGAGA-GTGGATCCTAGAGCGAAA-GTTTGCAGTCAGGTAAAGTTGAGAGGTG
m               AAGCCCCTGAAA-GTAGATCCCAGAGAGAAA-GTTTGCAGTCAGGTGAACCTAGGGGGTT
r               AAGCTCCTGAAAAGTAGATCCCAGAGAGAAAAGTTTGCAGTCAGGTGAACTTAGGAGGTT
                 ***    ** * ** ***** **** **** ************** **  *  * ***

h               GTATGATTT-------TTTTTTTTT------AAAGAGTTTTGG--AAAATTCTTATTCCT
c               GTATGATTT-------TTTTTTTTTT-----AAAGAGTTTTGG--AAAATTCTTATTCCT
m               GTACGATTTAAAATAATACTAATTTTTAAAAAAGGACTTTTGGCAAAA-TTATTTTTGCT
r               GTACGATTTAAAATAATACTAATTTT---AAAAGGATTTTTGGCAAAAATCCTTTTTGCT
                *** *****       *  *  ***      ** ** ******  *** *  ** ** **
                                                          1
h               TC-CAAATGACTGTTAGAAAGGAAAAGCAGATATGAGAAGC AGAATTGAGAAAGCCCAAG
c               TC-CAAATGACTGTTAGAAAGGAAAAGCAGATATGAGAAGCAGAATTGAGAAAGCCCAAG
m               CT-CAAATGATTG----GAAGGAGGAGCAGAT-TATGAAG-AGAACTGAGAAAACCCAAG
r               TTTCAAATGATTG----GAAGGAGGAGCAGAT-TATGAAG-AGAACCGAGAAAGCCCAAG
                   ******* **     *****  ******* *  **** ****  ****** ******

h               TTAGGCTGAGAGGAAATTTGGGGTAAGTTTCTCCCTGGATTGAGGTTTTGTGATATCTAA
c               TTAGGCTGAGAGGAAATTTGGGGTAAGTTTCTCCCTGGATTGAGGTTTTGTGATATCTGA
m               TTAGCCTGAGAGAAAATTGGGGGTAAGTTTCACTCCGGATTGAGGTTTTGTGCTGTCTGA
r               TTAGCCTGAGAGAAAATTGGGGGTAAGTTTCTCTCCGGATTGAGGGTCTGTACTATCTAA
                **** ******* ***** ************ * * ********* * ***  * *** *

h               AGTCTTACTAAGTTGGAAGCTTGAGTGCAGTGGATAGAGTGGAGAGAACCTGAAAGGTTA
c               AGTCTTACTAAGTTGGAAGCTTGAGTGCAGTGGATAGAGTGGAGAGAACCTGAAAGGTTA
m               AGTCTTACTAAGTTGGAAACC--------ATGGG--GAGTGGACAATGGATAAGAGATTA
r               AGTCTTACTAAGTTGGAAACC--------ATGGG--GAGTGGACAAGAGATAAGAGATTA
                ****************** *          ***   ******* *     * * ** ***

h               AAATAGAGCAAACACATAAAA-CTGAGAGATCAAATTAAGGAGTAGATTATATAAGCTTA
c               AAATAGAGCAAACACATAAAA-CTGAGAGATCAAATTAAGGAGTAGATTATATAAGCTTA
m               AAATAGACCAAAAACTTAAAA-CGGACATATCTGAT---GGAGTAGATTATATAAGCTTG
r               AAATAGACCAAAAACTTAAAAACGGACATATCTGAT---GGAGTAGATTATATAAGCTTG
                ******* **** ** ***** * ** * ***  **   ********************
```

TCF4 exon 10c

Aligned regions

```
h       chr18:51120517-51121010
m       chr18:69771959-69772448
r       chr18:66,287976-66288465
exon 10c sequence identity between h and m: 85%
Note: the respective chimp sequence is not available in the latest March 2006
assembly.
```

```
h               CTAAACGGGCTCCAGATGCAGCTTTATCCAGGG-GACGCCTCATGCCTGGTCATTTGGGT
m               CTAAGGATGGAGTGGATGTGGCTCACACCAGGATGATGCCTGGAGCCTGGTCATTTGGGT
r               CTAGCCATGTTCTAGACGGGGCTCACACCAGGATGATGCCCAGAGCCTGGTCATTTGGGT
                ***     *     ** *  ***    *****  ** ***    ****************

h               GGCTGGGCTTGTCCCACCTCCCCTTCTTCATGCCCTCCTCCCGGTCTCATCCTAGGCCCT
m               GTCTGGGCCTGTCCCACCTCCCCTCCC-CATGTACTCCTCCCACCCTCCTCCTGGGCCCT
r               GGCTGGGCCTGTCCCACCTCCCCTCCC-CATGTACTCCTCCCAGCCTTCTCCTAGGCCCT
                * ****** *************** *  ****  ********   **  **** ******
```


1 1
h CCCACCCCTAGGCCAGCCACAGGCCTCACAGCGGACATCTGCGGGTCTGAGTTTCATATA
m CCCACCCCCAAGCCAGTCAGGGGCCTCACTGCAGACATCTGCGGGCATGAGGCACATATA
r CCCACCCCCAAGTCAGTCCCAGGCCTTCCTGCGGACATCTGCAGGTCTGAGGCACATGTA
h TCCAGCATGAAGTTCCCTCTCTCCCTCCCATTATGAATTCTGGGCTGTGGTCCTGCTTGC
m CCCAGCCTGAGGTTGCGTCTCTGCCTCCCGCTATGAATTCTGGGCTGTGGTCCTGCTCCC
r CCCAGCCCGAGGTTGTCTCCCTGCCTCCCGCTATGAATTCTGGGCTGTGGTCCTGCTCTG
1
h --TCCTTGTGTGTCAGATTGAAGTTCTGATGCTGTCAGCCATGTCTGACCAATCAGCAAA
m TGCTCTTGCCTGTCAGATTGCAATTCTGATGCTGTCAGCCATGCCTGACCAATCAGCAAA
r -GTCCTTGTCCGTCAGATTGCAATTCTGATGCTGTCAGCCACGCCTGACCAATCAGAGAG
0.514
1 1
h CAACACCACAGGATGAAATTTAAACAATGCAGATGCTCAGATACAGGGCTATGTTGCCTG
m CACCACCACGGGATGAGACTTAAACAATGCAGATGCTCAAAGGCAGGGTTATGTCTCGTG
r CGCCGCTGCGGGATGAAACTTAAACAATGCAGATGCTCCAATGCAGGGTTATGTTTCGTG
h GACCACGAAGGGAAAGCAGAGGTAAGGTCAGAGTCTCCCCAAAGTGGTCTATCCTGAGCC
m GACCACGAAGGGAGAGCAGAGGTAAGGTCTCAGTCTCCCCAAAGTGGTCTGTCCCCAGCC
r GAGCACGGAGGGAGAGCAGAGGTAAGGTCTCAGTTTCCCCAAAGTGGCCTGTCCCCAGCC
h TGCCTGACTTGACACCTCTTGGGCTGCACTG-TGC--TGCTTTGACTTGTATGTTACCCA
m TGCTTGGTCTGACACTCCCTAAGCGACACTGCCGC--TGCTCTGCCTCGG-TGCCACC--
r TGCTCGGCCTGACACTTCCTAAGTCATACTGCTGCCCTGCTCTGCCTTGT-TGTCACC--
h TTTGCTTTGCACTACTTCTGTCTTCCCTGGCATTCACCAGAACTTTCATTTTAATTCGCT
m --TGCTTTCCCCTGCCTTT--CTTCCCTGCCACTCACCAGAACTTTTGTCTGCGTCC--C
r --TGCTTTCCC-TACCTTT--CTTCCCTGCCACTCACCAGAACTTTTGTTCCCATCC--C

Supporting Table S1. Accession numbers of representative mRNA or EST sequences for alternative *TCF4* transcripts and complete coding sequences of TCF4 isoforms cloned in full-length in this study.

Transcripts with alternative 5′ exons		
Transcript	**Human**	**Mouse**
1a	CD557562	
1aΔ3	FR748202	
1b(1)	BC031056	
1b(1)Δ3	DB055535	
1b(1,2)	FR748203	
1b(1,2)Δ3	DB443719	
1b(2)	DB040143	
1b(2)Δ3	FR748204	
3a	FR748205	
3aΔ3	DC398422	
3b	AK315074	AK133885
3bΔ3	AK299169	AK051958
3c	DB106801	AK081012
3cΔ3	FR748206	
3d	M74719	X91753
3dΔ3	BP206418	
4a-I	BP293388	
4a-II	DC337265	
4a-III	DC390184	
4b	BI544467	
4c	DC358747	CJ115804
5a-I	AU279553	
5a-II	FR748207	
5b	AK096862	
5c	DC398249	
7a-I	DC358599	
7a-II	AK300612	BY247629
7a-III	DC375084	
7b-I	AK095041	
7b-II	DC350124	CD350230
8a	AK316165	BY286412
8b-I	FR748208	
8b-II	AK300636	BY252182
8c-I	DC326149	
8c-II	FR748209 CA393351	BB663894
8d	BP230382	
10a	AK300038	U16321
10b	BP214032	BU058820
10c	DA664480	BY333068

Transcripts with alternative internal splicing		
Transcript	**Human**	**Mouse**
8-9 present	DB106801	AK133885
Δ8-9	AK315074	
8 (acceptor I) 130 bp	AK096862	AK133885
8 (acceptor II) 127 bp	AK095041	BY259217
15 (acceptor I) 77 bp	M74718	AK133885
15 (acceptor II) 74 bp	CN404346	BE652641
18^{-} (donor I) 151 bp	AK316165	U16321
18^{+} (donor II) 163 bp	AK315074	AK133885

Full-length coding sequences	
Isoform	**Human**
TCF4-B^{+}	FR748210
TCF4-B^{-}	FR748211
TCF4-B^{+}Δ	FR748212
TCF4-B^{-}Δ	FR748213
TCF4-C^{-}	FR748214
TCF4-C^{-}Δ	FR748215
TCF4-E^{-}	FR748216
TCF4-F^{-}	FR748217
TCF4-G^{-}	FR748219
TCF4-D^{-}	FR748218
TCF4-A^{+}	FR748220
TCF4-A^{-}	FR748221
TCF4-H^{-}	FR748222
TCF4-I^{-}	FR748223

The nucleotide sequences deposited by this study are under Accession Numbers FR748202-FR748223.

Supporting Table S2. List of oligonucleotides used in this study.

Specificity	Forward primer (5´- 3´)	Reverse primer (5´- 3´)	Products (bp)	Tm (°C)	Notes
1a-8	GAGGCATGAGAGCGAGCAAG	GTTTGGTGGGCGAAAGGGTTCC	536, 444	60	37
1-5	AGAACATTATCTCTTGGACGGGCA	GGACCCTGAGCTACTTCTGTCTTC	260, 168 (381, 289)	59	35
1b,2-5	AGGAGGCCCTTGGAGATGGAG	GGACCCTGAGCTACTTCTGTCTTC	330, 238	61	35
3a-8	AGACAGGAATCTTCACCATGTGTG	GTTTGGTGGGCGAAAGGGTTCC	530, 438	59	37
3b-5	CCCGGATGTGAATGGATTACAATG	GGACCCTGAGCTACTTCTGTCTTC	689, 597	59	40
3c-5	CTGTTCAGGCTTCAGATTGTAACTG	GGACCCTGAGCTACTTCTGTCTTC	239, 147	57	37
3d-5	GGTTAGACATGAACGCCGCCTC	GGACCCTGAGCTACTTCTGTCTTC	258, 166	59	37
4aI-5	CCCCTTTCTGCTTGTTGCACTTTC	GGACCCTGAGCTACTTCTGTCTTC	169	59	40
4aIII-11	GCTCTCCTGTTAGAGACGAAATGAG	TGGATGCAGGCTACAGTAGCTG	739	59	40
4b-5	GAGATGGAAATGGACATTGCCACG	GGACCCTGAGCTACTTCTGTCTTC	138	57	40
4c-11	AGCTCTGAAGGCACAAGTGTC	TGGATGCAGGCTACAGTAGCTG	762	57	37
5a-11	GGGAGCTTTCCCTGACATTCTTTC	TGGATGCAGGCTACAGTAGCTG	665	59	40
5b-11	CATGGAGCACAGGAGAGTAACAAG	TGGATGCAGGCTACAGTAGCTG	800	59	40
5c-11	CTGGAGGGTAAATACGGTCATTCC	TGGATGCAGGCTACAGTAGCTG	743	55	45
7a-8	CCTCGGCGTGCTACCCATAAG	GTTTGGTGGGCGAAAGGGTTCC	185, 159, 147	57	37
7b-8	TCTTGCTTGCATACATTGCCAG	GTTTGGTGGGCGAAAGGGTTCC	160, 262	59	40
8a-11	TTACAGCCATTCCAGCGCAAC	TGGATGCAGGCTACAGTAGCTG	444	59	40
8b-11	CTTGTGAAAGCATCATCCAGACCG	TGGATGCAGGCTACAGTAGCTG	542	57	40
8cI-11	AGAGGTGGATGGATTGGCATGTG	TGGATGCAGGCTACAGTAGCTG	457, 956	59	40
8cII-11	CAGCTGAAATGATTCCCCACTGTG	TGGATGCAGGCTACAGTAGCTG	439	57	40
8d-11	GTCAGTAAAAGTTGATTACAAGTAACG	TGGATGCAGGCTACAGTAGCTG	599	57	40
10a-11	ACCATGTACTGCGCATACACAATC	TGGATGCAGGCTACAGTAGCTG	300	57	37
10b-11	GAGAAAGCCCAAGTTAGGCTGAG	TGGATGCAGGCTACAGTAGCTG	263	59	40
10c-11	GCTATGTTGCCTGGACCACGA	TGGATGCAGGCTACAGTAGCTG	262	57	40
5-11	AGAAGACAGAAGTAGCTCAGGGTC	TGGATGCAGGCTACAGTAGCTG	628, 448	57	35
10-20	TATGCTCCATCAGCAAGCACTG	CATCTGTCCCATGTGATTCGATGCG	1461(BglII:1090+371), 1449	57	35
10-11	TATGCTCCATCAGCAAGCACTG	TGGATGCAGGCTACAGTAGCTG	225	61	q
17-18	GCCATTCTCTTCTGCCAAACCA	TCATCACCCTCGTCATCGGATT	172	60	q
19-20	TCCTCAGTCTGGAGCAGCAA	TCGATGCGTCTCCCATTCCA	144	56	q

Specificity	Forward primer (5′- 3′)	Reverse primer (5′- 3′)	Products (bp)	Tm (°C)	Notes
HPRT1	CAGTCCCAGCGTCGTGATTA	AGCAAGTCTTTCAGTCCTGTC	141	58	q
GAPDH*	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG	87	60	q
HMBS*	GGCAATGCGGCTGCAA	GGGTACCCACGCGAATCAC	64	58	q
SDHA*	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCATG	86	58	q/32
UBC	GATTTGGGTCGCAGTTCTTG	CTCGATGGTGTCACTGGGCTCA	122	58	q
3-11(RPA)	caccATGCATCACCAACAGCGAATGG	TGGATGCAGGCTACAGTAGCTG	777+4 (597+4)		c
10-16(ISH)	CTCCATCAGCAAGCACTG	GTTCCATACCCTGAGCCCAGAC	760		c
8-9(NLS)	ctaagatctaCCCCGAAGGAGGCCTCTTCA	cacggtaccGGAGGAACTTTTCGAACTTTC	68+19		c
TCF4 NLS m1	TCTAGCAATAATCCCgcAgcGgcGCCTCTTCA CAGTAGT	ACTACTGTGAAGAGGCgcCgcTgcGGGATT ATTGCTAGA			m
TCF4 NLS m2	CATGGAGGTACAGACAgcGgcAGTTgcAgcAG TTCCTCCAGGTTTG	CAAACCTGGAGGAACTgcTgcAACTgcCgc TGTCTGTACCTCCATG			m
3×μE5	CTAGAGATCTGAACACCTGCAAGAACACCTGC AAGAACACCTGCAAG	GATCCTTGCAGGTGTTCTTGCAGGTGTTCT TGCAGGTGTTCAGATCTCTAGAGCT			a
Specificity	**Forward primer (5′- 3′)**	**Reverse primer (5′- 3′)**	**Isoform**		**Notes**
3	caccATGCATCACCAACAGCGAATGG		TCF4-B		c
4	tcgaattcaccATGTTTTCACCTCCTGTGAGC		TCF4-C		c
3c	caccATGCAACGGGCAAAAACTGAAC		TCF4-E		c
5b,5	caccATGGAAGAGGACAGCAGAGATGTA		TCF4-F		c
8	tcgaattcaccATGGATATGGGCAACCCAG		TCF4-D		c
7b	caccATGAAGGATATTTTTTTCCAGTTTATC		TCF4-G		c
10a	cACCATGTACTGCGCATACACAATC		TCF4-A		c
10c	caccATGAAATTTAAACAATGCAGATGCTC		TCF4-H		c
10,11	cacCATGCAAGATGGCCATCACAG		TCF4-I		c
20		CATCTGTCCCATGTGATTCGATGCG	all		c
20stop		ctgagctcTcACATCTGTCCCATGTGATTC	all		c
21		ccactcgaGCGTCTGCGATTCATAACTAC	all		c
Specificity	**Forward siRNA (5′- 3′)**	**Reverse siRNA (5′- 3′)**	**Species specificity**		**Notes**
12 (I)	ACCAUUACAGCACCUCUUCtt	GAAGAGGUGCUGUAAUGGUTT	human and mouse		si
12 (II)	CGGAACAGACAGUAUAAUGtt	CAUUAUACUGUCUGUUCCGTT	mouse		si
20	GAAGGGAGGAAGAGAAGGUtt	ACCUUCUCUUCCUCCCUUCTT	human and mouse		si

The oligonucleotides were used for detection of TCF4 transcripts and housekeeping gene transcripts by RT-PCR, amplification of TCF4 coding sequences and fragments for cRNA probe synthesis for cloning, site-directed mutagenesis of TCF4 coding sequence, and annealing of μE5 E-boxes. TCF4 primer pairs are designated with the exon range spanned by the amplified product. For clarity TCF4 PCR product lengths are given only for transcripts containing longer variants of exons 8 and 15. In the 'Notes' column, the number of cycles is given for the primer pairs used for end-point PCR analysis; primers used for quantitative PCR, cloning or annealing are indicated with 'q' , 'c' or 'a', respectively; siRNAs are indicated with 'si'. Primers marked with asterisk (*) have been described before [55]. Nucleotides in lowercase indicate adapter sequences (CACC Kozak consensus and restiction sites) or mutated positions.

PAPER II

Sepp M, Pruunsild P, Timmusk T
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. 2012;21(13):2873-2888.

Human Molecular Genetics, 2012, Vol. 21, No. 13 **2873–2888**
doi:10.1093/hmg/dds112
Advance Access published on March 28, 2012

Pitt–Hopkins syndrome-associated mutations in *TCF4* lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant-negative effects

Mari Sepp, Priit Pruunsild and Tõnis Timmusk*

Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

Received January 31, 2012; Revised and Accepted March 20, 2012

Transcription factor TCF4 (alias ITF2, SEF2 or E2-2) is a broadly expressed basic helix–loop–helix (bHLH) protein that functions as a homo- or heterodimer. Missense, nonsense, frame-shift and splice-site mutations as well as translocations and large deletions encompassing *TCF4* gene cause Pitt–Hopkins syndrome (PTHS), a rare developmental disorder characterized by severe motor and mental retardation, typical facial features and breathing anomalies. Irrespective of the mutation, *TCF4* haploinsufficiency has been proposed as an underlying mechanism for PTHS. We have recently demonstrated that human *TCF4* gene is transcribed using numerous 5′ exons. Here, we re-evaluated the impact of all the published PTHS-associated mutations, taking into account the diversity of TCF4 isoforms, and assessed how the reading frame elongating and missense mutations affect TCF4 functions. Our analysis revealed that not all deletions and truncating mutations in *TCF4* result in complete loss-of-function and the impact of reading frame elongating and missense mutations ranges from subtle deficiencies to dominant-negative effects. We show that (i) missense mutations in TCF4 bHLH domain and the reading frame elongating mutation damage DNA-binding and transactivation ability in a manner dependent on dimer context (homodimer versus heterodimer with ASCL1 or NEUROD2); (ii) the elongating mutation and the missense mutation at the dimer interface of the HLH domain destabilize the protein; and (iii) missense mutations outside of the bHLH domain cause no major functional deficiencies. We conclude that different PTHS-associated mutations impair the functions of TCF4 by diverse mechanisms and to a varying extent, possibly contributing to the phenotypic variability of PTHS patients.

INTRODUCTION

Pitt–Hopkins syndrome (PTHS; OMIM 610954) was first described in 1978, but presumably remained underdiagnosed with only a few cases reported in almost 30 years (1,2). In 2007, the diagnosis was simplified by the identification of *TCF4* (transcription factor 4) haploinsufficiency as the underlying cause of PTHS (3–5) and to date over 100 patients have been described with autosomal dominant mostly *de novo* frameshift, splice-site, nonsense or missense mutations in *TCF4*, and large deletions encompassing all or part of *TCF4* gene (6–23). *TCF4* is located on chromosome 18q21.2 and encodes a basic helix–loop–helix (bHLH) transcription factor TCF4 also known as ITF2 (immunoglobulin transcription factor 2), SEF2 (murine leukaemia virus SL3-3 enhancer factor 2) and E2-2 (24,25). In this study, the term PTHS is used only for genetically confirmed syndrome and is distinguished from the Pitt-Hopkins-Like syndromes 1 (OMIM 610042) and 2 (OMIM 600565) that have similar clinical phenotypes, but different genetic basis involving recessive defects in *CNTNAP2* (contactin associated protein-like 2) and *NRXN1* (neurexin 1) genes, respectively (26).

The characteristic symptoms of PTHS patients include specific dysmorphic features, severe motor and mental retardation (delayed walking, absent language), stereotypic movements, breathing abnormalities (hyperventilation episodes, apnoea), ophthalmological disorders (strabismus, myopia), hypotonia and seizures (8,15,18). In addition to PTHS, *TCF4* has been linked to other neurodevelopmental disorders. Two recent

*To whom correspondence should be addressed. Tel: +372 6204445; Fax: +372 6204401; Email: tonis.timmusk@ttu.ee

studies have analysed the contribution of *TCF4* hemizygosity to the phenotype of individuals with large 18q deletions encompassing many genes (13,27). In this context, hemizygosity of *TCF4* has been found to be a factor for the increased risk of autistic-like behaviours, and to be responsible for the absent cognitive and motor development beyond the developmental milestones normally acquired by 1 year of age, *corpus callosum* abnormalities and higher probability of premature death due to aspiration-related complications (13,27). Different single-nucleotide polymorphisms (SNPs) in *TCF4* have been associated with Fuchs corneal dystrophy (28–31) and schizophrenia (32–34). One of the schizophrenia-associated SNPs in *TCF4* (rs9960767) has been demonstrated to have opposite effects on two endophenotypes of schizophrenia—the risk allele is associated with impaired sensorimotor gating and less pronounced deficits in verbal declarative memory (35,36). However, the association mechanisms remain unclear, since *TCF4* expression in the adult brain was found not to be influenced by the polymorphism (37). Yet, impairment of sensorimotor gating and cognitive functions has been linked to moderate *Tcf4* overexpression in transgenic mice (38). Neurological phenotypes of *Tcf4* null mice have not been characterized in-depth. It is known that homozygous *Tcf4* knockout mice die shortly after birth, whereas heterozygous knockout mice are viable, but both have grossly normal brains (39,40). Development of the brain and eyes has been shown to be delayed by reduction of *tcf4* expression in zebrafish (41).

TCF4, together with TCF3/E2A and TCF12/HEB, belongs to the family of E-proteins that are homologous to *Drosophila* protein Da (Daughterless) and exhibit a broad expression pattern (42). E-proteins form homodimers and heterodimers with tissue-specific bHLH factors and the dimers bind DNA at Ephrussi box (E-box) sequences (CANNTG, 43). TCF4 is highly expressed in the nervous system (9,44,45) and is known to heterodimerize with several bHLH transcription factors that play important roles in the development of the nervous system, e.g. ATOH1 (atonal homolog 1, alias MATH1), ASCL1 (achaete–scute complex homolog 1, alias MASH1), NEUROD1 (neurogenic differentiation 1, alias BETA2) and NEUROD2 (alias NDRF) (38,40,46,47). A specialized function described for the TCF4-ATOH1 heterodimer in the differentiation of pontine nucleus neurons in mice illustrates the potential complexity of the functions carried out by different TCF4 heterodimers in the nervous system (40). Furthermore, we and others have shown that the usage of alternative promoters for transcribing the human *TCF4* gene enables production of many amino-terminally distinct TCF4 isoforms (TCF4-A–TCF4-R) that differ in their subcellular localization and transactivational capacity (25,45). All TCF4 isoforms contain the C-terminal bHLH domain that mediates dimerization and DNA binding, and the transcriptional activation domain 2 (AD2), whereas the longer isoforms have an additional transcriptional activation domain (AD1) and a nuclear localization signal (NLS) in their N-terminal part (45). Alternative splicing of *TCF4* gene allows the production of + or − and full-length or Δ protein isoforms that include or lack a four amino acid insertion and the NLS-containing region, respectively (25,45).

Regardless of mutation type, haploinsufficiency has been proposed as an underlying mechanism of PTHS, but extensive functional analyses to validate this concept are lacking. Thus, the present study was undertaken to determine how different PTHS-associated mutations in the *TCF4* gene affect the transcription factor function. To this purpose, we re-evaluated the impact of all the published PTHS-associated mutations so as to take into account the presence of alternative 5′ exons in the *TCF4* gene. Moreover, we analysed the effects of a series of reading frame elongating and missense mutations on TCF4 stability, subcellular localization and ability to heterodimerize, to bind DNA and to transactivate E-box-controlled reporter gene transcription. Our data provide evidence that PTHS-causing mutations create hypomorphic, non-functional and dominant-negative *TCF4* alleles.

RESULTS

Mapping of PTHS-causing mutations in *TCF4*

We have previously demonstrated that the human *TCF4* gene is transcribed using 21 5′ initial exons (designated with a number and a letter, e.g. exon 3a) that are interspersed with internal exons 1–9 and followed by constitutive internal exons 10–20 and the 3′ exon 21 (45). As a result, a variety of *TCF4* transcripts containing different number of internal exons and potentially encoding for TCF4 protein isoforms with 18 different N-termini are generated. To specify the impact of the PTHS-associated mutations, we first mapped all the mutations described to date to the composite structure of *TCF4* gene. Figure 1A shows 12 PTHS-associated translocations and large intragenic deletions relative to the entire *TCF4* gene. Out of these, the translocation with a breakpoint between exons 10c and 13 (15) and the deletion spanning exons 1a to 10c-11 (4) lead to loss of all 5′ exons, so that no natural start site for transcription initiation of the remaining *TCF4* constitutive exons is left. The translocation with a breakpoint between exons 4 and 5a from a patient with milder clinical phenotype than classical PTHS (48) encompasses 9 upper 5′ exons, but leaves the rest of the gene intact and consequently, fully functional transcripts could be transcribed from the remaining 12 initial exons. Similarly, the deletions extending from exon 5b to 8c (18) or from 7a–7b to exon 10b (11) are anticipated to allow the synthesis of functional *TCF4* transcripts from downstream 5′ exons 8d, 10a, 10b and 10c or from 10c only, respectively. It is uncertain whether and how many utilizable transcription initiation sites are present in the two patients with deletions extending from exons 7a–7 to exons 9–10c (18). Of note, in the four latter cases it is highly unlikely that functional mRNAs could be produced using the remaining 5′ exons upstream of the deletion, since all these mRNAs would contain out-of-phase deletions of internal exons 5–7 or 7–9. The rest of the PTHS-associated intragenic deletions encompass at least one of the constitutive exons 10–21 (11,14,18) and thus, affect all of the alternative *TCF4* transcripts. These deletions are out-of-phase and/or lead to the loss of the AD2 or bHLH domain encoded by exons 14–16 or 19, correspondingly. Therefore, no functional *TCF4* transcripts can be produced from the alleles with deletions encompassing exons 10–21.

Next, we examined the PTHS-associated splice-site, nonsense and small indel mutations. The splice-site mutations

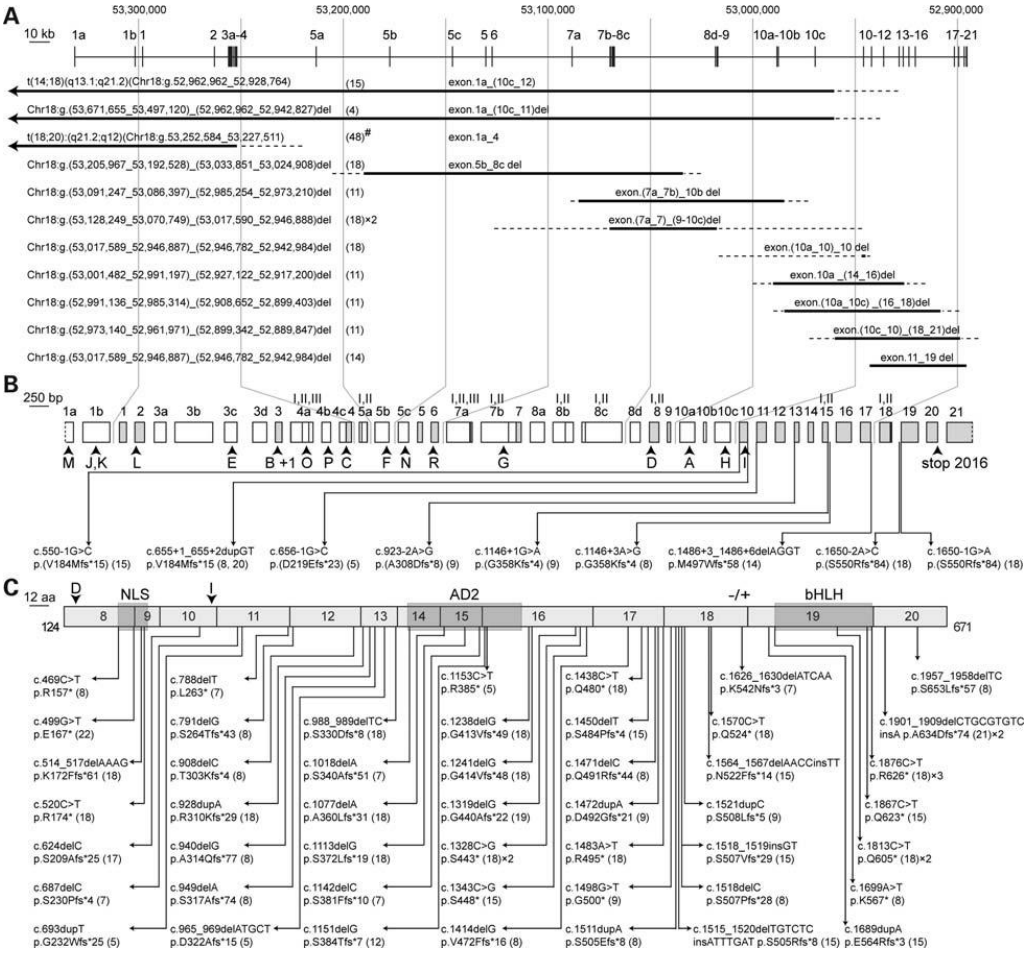


Figure 1. Schematic depiction of Pitt–Hopkins disease-causing mutations. (**A**) The published deletions and translocations that encompass part of *TCF4* gene. The minimum deleted or translocated regions are depicted as solid lines and the maximum regions as dotted lines in relation to the *TCF4* genomic organization with introns drawn in scale. Arrows designate deletions or translocations that extend beyond the depicted chromosomal region. The coordinates are given according to human genome assembly hg19. The translocation denoted with '#' is from a patient with milder phenotype than classical PTHS. (**B**) *TCF4* splice-site mutations depicted in relation to the gene structure with exons drawn in scale. White boxes represent 5′ exons and grey boxes represent internal and 3′ exons. Exon names are shown above the gene structure in (A) and (B). Roman numerals indicate alternative splice donor or acceptor sites. Start codons of different TCF4 isoforms and the stop codon are shown with arrowheads. (**C**) Nonsense and small indel mutations in the *TCF4*-coding region depicted in relation to the schematic of C-terminal part of TCF4 protein encoded by exons 8–20. The NLS, AD2 and bHLH domains are indicated with dark shading. The site of a four amino acid insertion in $TCF4^{+}$ isoforms compared with the $TCF4^{-}$ isoforms is indicated with −/+. The first methionines of TCF4 isoforms D and I are shown with arrowheads. Compared with the full-length isoforms, the respective TCF4-Δ isoforms and also TCF4 isoforms A, H and I lack the amino acids coded by exons 8–9. The localization of mutations is indicated with arrows and the coordinates of mutations are given according to the TCF4 isoform B^{+} encoding cDNA (NM_001083962.1) in (B) and (C). For recurrent mutations, the number of patients is indicated after the citation in (A)–(C).

are mapped to the schematic of all *TCF4* exons in Figure 1B, and nonsense and small indel mutations are depicted in Figure 1C in relation to the C-terminal part of TCF4 protein encoded by exons 8–20. All nine reported splice-site mutations affect the constitutive exons and are predicted to shift the reading frame as a result of out-of-phase exon skipping or novel splice-site selection. Forty-four of the 48 nonsense and small indel mutations described in 53 patients map to the constitutive exons 10–19 and lead to the generation of a premature stop codon. As a result, most of the transcripts containing these mutations would encode a protein without the bHLH domain. Only two nonsense mutations c.1867C>T p.Q623* (15) and c.1876C>T p.R626* (recurrent in 3 patients, 18) located at the end of exon 19 allow the translation of a protein with the majority of the bHLH present. However, all these transcripts are predicted to be subjected to the

nonsense-mediated decay pathway. There are two PTHS-associated indels that do not cause a premature stop: mutations c.1901_1909delinsA p.A634Dfs*74 and c.1957_1958del p.S653Lfs*57 (8,21), which map to exon 20 and elongate TCF4 reading frame into the 3′ UTR exon 21. The effect of the elongations on the function of the protein is not known. Additionally, in contrast to the mutations in constitutive exons 10–20 that affect all the alternative *TCF4* transcripts, there are four premature stop-causing mutations reported in upper exons 8–9—c.469C>T, c.499G>T, c.514_517del and c.520C>T (8,18,22). These mutations should have no effect on the Δ-isoforms encoding transcripts, which have exons 8–9 spliced out, and the shorter *TCF4* transcripts initiated at 5′ exons 10a, 10b and 10c. Taken together, the analysis above reveals that although many of the PTHS-associated deletions and premature stop-inducing mutations probably lead to complete loss-of-function of the affected allele, there are some that are expected to permit the production of the Δ-isoforms of TCF4 and/or the shorter isoform(s) TCF4-D, TCF4-A, TCF4-I and/or TCF4-H.

Subsequently, we focused on the PTHS-associated missense mutations. Twenty-two PTHS patients with missense mutations situated at 9 amino acid positions in the constitutive part of TCF4 and giving rise to 12 different substitutions have been described to date (Fig. 2A). Only two of the missense mutations are situated outside of the bHLH domain: mutation p.G358V (8) is located in the AD2 domain at a position conserved in mammals, and mutation p.D535G (9) is situated between the AD2 and bHLH domain at a position conserved in vertebrates. The majority of the described missense mutations are located in the bHLH domain at positions conserved also in *Drosophila melanogaster* protein Da and *Caenorhabditis elegans* protein HLH2 (Fig. 2A). To examine the location of the missense mutation sites in the bHLH domain 3D structure, we used the published co-crystal structures of DNA-bound E47 bHLH homodimer and E47-NEUROD1 bHLH heterodimer (49,50) as templates to build the structure models of DNA-bound TCF4 bHLH homo- and heterodimer. We replaced NEUROD1 with NEUROD2 in the heterodimer structure since NEUROD2 was included in the functional studies described in what follows. In the modelled region, TCF4 and E47 share 85% sequence identity, and NEUROD2 is 97% identical to NEUROD1 (Fig. 2B). As expected from the high similarity of the proteins, the obtained TCF4 models closely resembled the E47 template structures (root mean square deviation <0.5 Å). It has been shown for the E47 homodimer that the CAC and CAG half-sites of the CACCTG E-box are bound non-equivalently by the 'specific' and 'non-specific' subunits, respectively, and only the E-protein subunit in the 'specific' conformation makes contacts with the central base pairs of the E-box (49). Similar DNA-binding geometry of the subunits is preserved in the TCF4 homodimer structure model (Fig. 2C). The heterodimer model includes DNA containing the CATCTG E-box with TCF4 in 'non-specific' conformation oriented to the CAT half-site (Fig. 2D), in the same way as described for the E47 heterodimer (50).

We mapped the PTHS-associated missense mutation sites in the bHLH domain to the TCF4 homo- and heterodimer structure models and examined the specific functions of the affected residues. First, we looked at the arginines that have been found to be mutated in PTHS and are located in the basic DNA-binding region and helix 1 of the bHLH domain (Fig. 2E and F). According to the models, R576 makes contacts with an E-box-flanking base in the 'specific' subunit or with the DNA backbone in the 'non-specific' subunit. The predicted role of R578 is to form hydrogen bonds with the DNA backbone in both conformations and additionally make the only contact with an E-box central base in the 'specific' subunit. In the models, R580 contacts the DNA backbone and forms salt bridges with E577 stabilizing the direct interaction(s) of the latter with the first DNA base(s) of the E-box (Fig. 2E and F). R582 and R569 do not contact DNA according to the models, but the former forms salt bridge(s) with the side chains of D583 and E586 in the first helix (Fig. 2E). Second, we examined the PTHS-affected alanines that are positioned at the dimer interface in the TCF4 structure models (Fig. 2G and H). A587 is situated in the first helix and packs against TCF4 L611 or the corresponding leucine in NEUROD2 in the second helix of the opposite subunit in the homo- or heterodimer model, respectively. A614 is located in the second helix and, according to the models, is buried in the interface of the dimer's four helices. Packing interactions are formed between A614 and L591 from both TCF4 subunits or the corresponding residues in NEUROD2 in case of the homo- or heterodimer model, respectively (Fig. 2G and H). On the basis of these data, we hypothesized that the PTHS-associated missense mutations in TCF4 could have an effect on the protein's DNA-binding, dimerization and/or transactivation ability.

PTHS-associated mutations in the second helix of the bHLH domain and C-terminal part of TCF4 induce protein destabilization and aggregation

In order to elucidate the underlying molecular mechanisms of the PTHS-associated missense mutations, we used site-directed mutagenesis to introduce seven different missense mutations to the TCF4-B$^-$ expression construct. One substitution was made for each TCF4 position that had been shown to be mutated in PTHS at the time of the design of this study—these were G358V, D535G, R576Q, R578H, R580W, R582P and A614V. In addition, to examine the impact of a frameshift-inducing deletion c.1957_1958delTC p.S653Lfs*57, we cloned the TCF4-B$^+$-encoding sequence together with the 3′ UTR and introduced the reading frame elongating S653Lfs*57 mutation by PCR. To differentiate between loss-of-function and gain-of-function effects of the frameshift, we also generated the S653* mutant. First, we assessed the expression levels of the PTHS-associated mutant proteins in HEK293 cells. We co-transfected wt or mutant TCF4-B-encoding constructs and pEGFP vector into HEK293 cells and detected the overexpressed proteins by western blotting (Fig. 3A). For quantification, TCF4 levels were determined densitometrically and normalized for transfection efficiency and loading to EGFP levels (Fig. 3B). We found that the steady-state levels of TCF4-B$^-$ A614V mutant were decreased to approximately one-third of the wt TCF4-B$^-$ levels, and the expression of TCF4-B$^+$ S653Lfs*57 mutant was indiscernible, whereas the other mutations did not have a significant effect

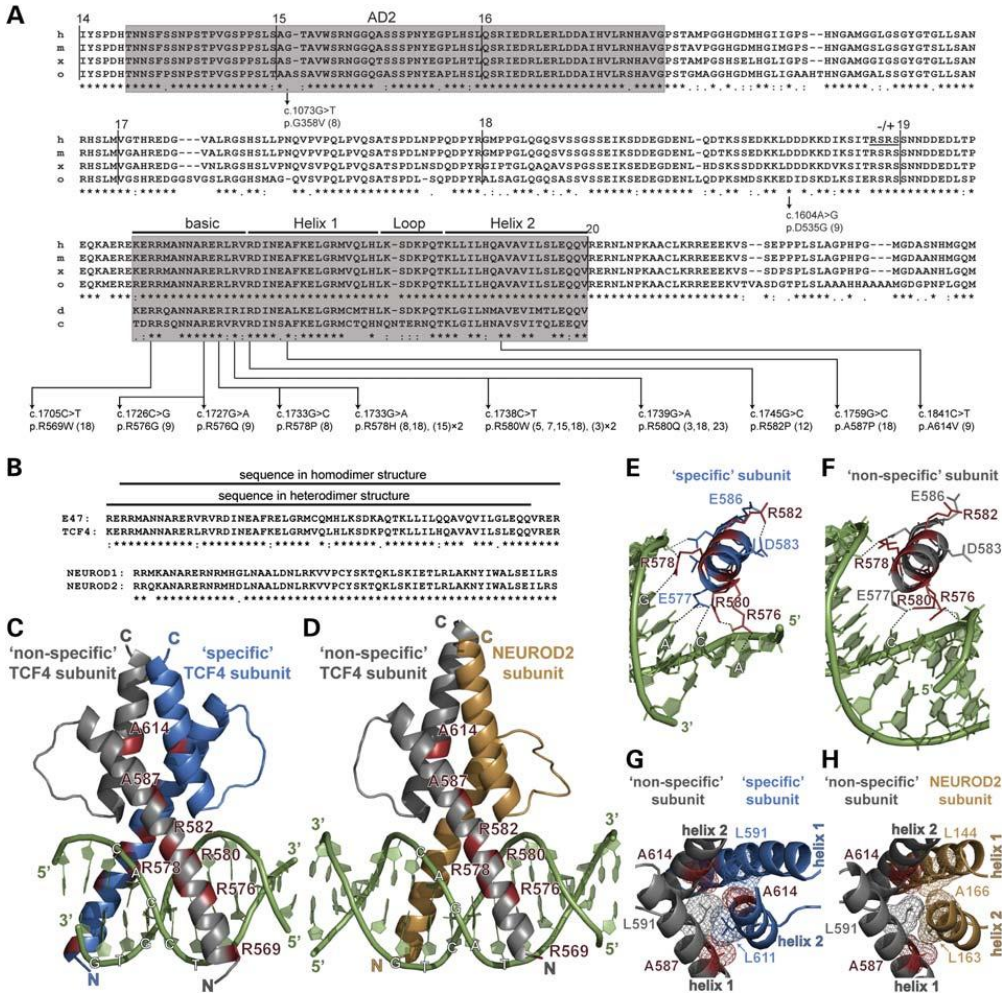


Figure 2. Mapping of Pitt–Hopkins disease-causing missense mutations. (**A**) The published *TCF4* missense mutations depicted in relation to the alignment of exons 14–20 encoded TCF4 protein sequences from *Homo sapiens* (h), *Mus musculus* (m), *Xenopus laevis* (x) and *Oryzias latipes* (o). The alignment of the TCF4 bHLH domain with the respective sequences from Da of *Drosophila melanogaster* (d) and HLH2 of *Caenorhabitis elegans* (c) is additionally shown. AD2 and bHLH domains are indicated with dark shading and the coordinates of mutations are given according to the TCF4 isoform B^+ encoding cDNA (NM_001083962.1). For recurrent mutations, all the studies and, if necessary, the number of patients per study are indicated. (**B**) Alignment of protein sequences of the bHLH domains of E47 and NEUROD1 to TCF4 and NEUROD2, respectively. (**C**) Ribbon drawing of the TCF4 bHLH homodimer bound to DNA, modelled using the crystal structure of the E47 bHLH homodimer bound to CACCTG E-box as a template. (**D**) Ribbon drawing of TCF4-NEUROD2 bHLH heterodimer bound to DNA, modelled using the crystal structure of E47-NEUROD1 bHLH heterodimer bound to CATCTG E-box as a template. (**E** and **F**) Close-up view of the PTHS-affected arginines in DNA-bound 'specific' subunit of the homodimer (E) or 'non-specific' subunit of the heterodimer (F). Shown are ribbons and sticks for TCF4 amino acids 576–613. Dashed lines indicate the presumed hydrogen bonds with DNA and salt bridges between the side chains. The bases contacted by the depicted TCF4 residues are labelled. (**G** and **H**) Structural environment of the PTHS-affected alanines at the dimerization interface of the TCF4 homodimer (G) or TCF4-NEUROD2 heterodimer (H). The relevant amino acids are shown as mesh-contoured sticks. For clarity, A587 is presented only on 'non-specific' TCF4 subunit and the C-terminal part of helix 2 of 'specific' TCF4 or NEUROD2 subunit is cut away. The amino acids that have been found to be mutated in PTHS are shown in red in (C)–(H).

on TCF4-B expression. Examination of mRNA expression levels by reverse transcription and quantitative PCR demonstrated that the levels of transcripts coding for A614V and S653Lfs*57 TCF4-B mutants were not decreased compared with wt transcripts in transfected HEK293 cells (Supplementary Material, Fig. S1). To determine whether proteasomal degradation is responsible for the reduced levels of A614V and S653Lfs*57 TCF4 mutant proteins, we treated transfected

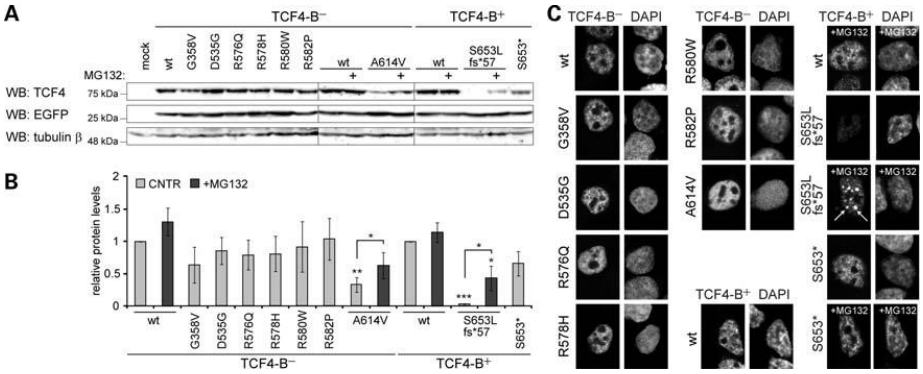


Figure 3. Expression and localization of TCF4-B mutants in cultured cells. (**A**) Western blot analysis of extracts from HEK293 cells co-transfected with pEGFP vector and wt or mutant TCF4-B-encoding constructs. Cells were treated with MG132 for 7 h where indicated. Signals of TCF4, EGFP (transfection control) and tubulin β (loading control) were detected by immunoblotting. Molecular mass markers are shown at the left. (**B**) Quantification of TCF4-B expression levels in (A) from four independent experiments. TCF4-B levels were normalized to EGFP levels and the values are given in relation to the levels of wt TCF4 in untreated control (CNTR) cells. Error bars indicate standard deviations. Statistical significance shown with asterisks is relative to the levels of wt TCF4-B. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *t*-test. (**C**) Localization of wt and mutant TCF4-B proteins in HEK293 cells visualized by immunocytochemical staining with TCF4 antibodies and confocal microscopy. Nuclei were stained with DAPI. Cells were treated with MG132 for 7 h where indicated. Examples of protein aggregates are indicated with arrows.

HEK293 cells with the proteasome inhibitor MG132 for 7 h and performed western blot analyses as described above. As shown in Figure 3A and B, the levels of wt TCF4-B^{-} and TCF4-B^{+} were only slightly elevated in MG132-treated cells compared with untreated cells, whereas the levels of TCF4-B^{-} A614V were increased almost two times and the expression of TCF4-B^{+} S653Lfs*57 became evident.

Next, we studied the subcellular distribution of wt and mutant TCF4-B proteins overexpressed in HEK293 cells. Indirect immunofluorescence analysis demonstrated that similar to wt TCF4-B, all mutant TCF4-B proteins localized to the cell nucleus (Fig. 3C). This result was expected since none of the PTHS-associated mutations affected the NLS of TCF4-B. However, S653Lfs*57 mutant protein was clearly detectable only in MG132-treated cells where it accumulated into nuclear dots, suggesting that S653Lfs*57 mutation renders TCF4-B prone to aggregation (Fig. 3C). All other mutant TCF4-B proteins, including the S653* mutant, and wt TCF4-B were distributed diffusely in the nuclei and the dots were not present in wt or S653* mutant TCF4-B^{+}-expressing cells even after MG132 treatment (Fig. 3C). Similar results were obtained when wt and mutant TCF4-B proteins were overexpressed in rat primary cortical and hippocampal neurons (Supplementary Material, Fig. S2A). Together these results indicate that A614V and S653Lfs*57 mutations destabilize TCF4 in cells by inducing protein misfolding and/or proteasomal degradation. The destabilization of S653Lfs*57 mutant is caused by the acquired 56 amino acids and not by the loss of 19 native C-terminal amino acids of TCF4.

TCF4 mutant proteins form heterodimeric complexes with ASCL1 and NEUROD2 in cells

Since TCF4 functions as obligate dimer, we enquired whether the PTHS-associated TCF4 mutant proteins are able to dimerize in cells. To answer this, we performed nuclear redirection assays, taking advantage of the knowledge that TCF4 proteins lacking the NLS can be transported to the cell nucleus through heterodimerization with NLS-containing bHLH partner proteins (45). We included ASCL1 and NEUROD2 in this analysis since they represent two separate phylogenetically defined families of the bHLH transcription factors and have been demonstrated to interact with TCF4 (38,46,51). Additionally, these factors might be of relevance to PTHS because of parallels between the phenotypes of the disorder and *Ascl1* or *Neurod2* null mice (52,53). We analysed the PTHS-associated mutations located in the bHLH domain or the C-terminal part of TCF4 in the context of EGFP–bHLH fusion protein, and the mutations located N-terminal to the bHLH domain in the context of TCF4-A^{-} isoform. When overexpressed alone in HEK293 cells, all wt and mutant EGFP-bHLH and TCF4-A^{-} proteins were present in the cell nucleus and cytoplasm (Fig. 4A), whereas co-expression with ASCL1 or E2-tagged NEUROD2 led to the accumulation of TCF4 proteins into the nucleus (Fig. 4B and C). Similar NEUROD2-E2-induced nuclear redirection of all overexpressed wt and mutant EGFP-bHLH proteins was seen in rat primary neurons (Supplementary Material, Fig. S2B). From these data, we concluded that the PTHS-associated TCF4 mutant proteins have maintained the ability to heterodimerize with ASCL1 and NEUROD2.

PTHS-associated mutations C-terminal to the AD2 domain lead to variable impairment of TCF4 DNA-binding ability

Subsequently, we examined the effect of the PTHS-associated mutations on the ability of TCF4-B to bind DNA *in vitro*. We used *in vitro*-translated proteins and determined their binding to the μE5 E-box (CACCTG) containing oligonucleotides by electrophoretic mobility shift assay (EMSA). All wt and

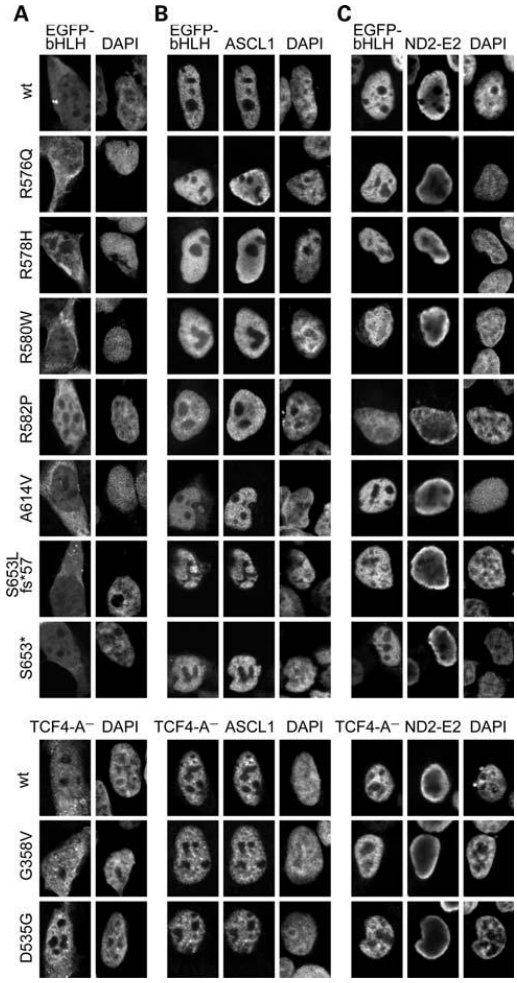


Figure 4. Heterodimerization of TCF4 mutants with NEUROD2 and ASCL1 in cultured cells. Nuclear redirection assay with wt and mutant EGFP-bHLH or TCF4-A proteins overexpressed in HEK293 cells (**A**) alone, (**B**) together with ASCL1 or (**C**) together with NEUROD2-E2 (ND2-E2). Immunocytochemical staining was carried out with TCF4, ASCL1 and/or E2 antibodies. EGFP-fused proteins were visualized by direct fluorescence, and nuclei were counterstained with DAPI.

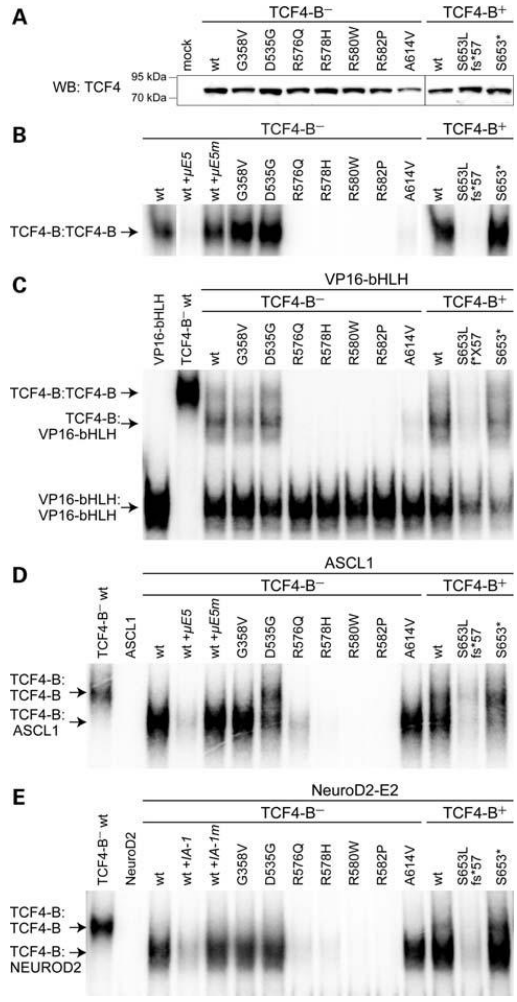


Figure 5. Ability of TCF4 mutants to bind DNA *in vitro*. (**A**) Western blot analysis of *in vitro*-translated wt and mutant TCF4-B proteins. Molecular mass markers are shown at the left. (**B**–**D**) Electrophoretic mobility shift analysis of *in vitro*-translated wt and mutant TCF4-B proteins' ability to bind the μE5 E-box (CACCTG) DNA 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). (**E**) Electrophoretic mobility shift analysis of *in vitro*-translated wt and mutant TCF4-B proteins' ability to bind the IA-1 E-box (CATCTG) DNA as heterodimers with NEUROD2 (TCF4-B:NEUROD2). Unlabelled wt (*μE5* or *IA-1*) or mutated (*μE5m* or *IA-1m*) E-box oligonucleotides were added to the binding reaction for competition where indicated in italics.

mutant TCF4-B proteins were efficiently translated as revealed by western blot analysis (Fig. 5A). The binding of wt TCF4-B$^-$ to the μE5 E-box was specific since competition with unlabelled wt and not mutant μE5 oligonucleotides diminished binding to the labelled DNA (Fig. 5B). Compared with wt TCF4-B, six of the nine mutant proteins displayed impaired binding to the μE5 E-box. Particularly, in case of R576Q, R578H, R580W and R582P mutations, the binding was completely abrogated, and in case of A614V and S653Lfs*57, the binding was severely reduced (Fig. 5B). We additionally looked at the homodimerization ability of TCF4 proteins with mutations in the HLH domain by crosslinking *in vitro*-translated proteins with glutaraldehyde. Homodimers of wt or R582P TCF4 were readily detected, whereas the levels of A614V homodimers were diminished (Supplementary Material, Fig. S3), indicating that the reduction in the

DNA-binding ability of A614V mutant is at least partly caused by its reduced capacity to homodimerize.

To elaborate further on the DNA-binding characteristics of the PTHS-associated mutant TCF4 proteins, we studied the DNA-binding ability of the intra-TCF4 heterodimers that contain one wt and one mutant TCF4 subunit. To allow discrimination of intra-TCF4 heterodimer from homodimers, we utilized a heterologic protein that contains VP16 transactivation domain fused to wt TCF4 bHLH domain and migrates faster than native TCF4 isoforms in EMSA. When wt TCF4-B isoform and VP16-bHLH were co-translated, the μE5 E-box binding of the heterodimer was detected as a complex with intermediate mobility compared with the TCF4-B homodimer and VP16-bHLH homodimer (Fig. 5C). Similar to the binding ability of mutant TCF4-B homodimers, the binding of intra-TCF4 heterodimers containing wt VP16-bHLH and mutant TCF4-B was severely impaired in case of R576Q, R578H, R580W, R582P, A614V and S653Lfs*57 mutations, indicating that in these instances mutation in one subunit is sufficient to abolish DNA binding.

Next, we studied the DNA-binding ability of ASCL1 heterodimers with wt or mutant TCF4-B proteins. In our assays, ASCL1 alone did not bind the μE5 E-box, whereas when ASCL1 was co-translated with wt TCF4-B$^-$ or TCF4-B$^+$, the binding of heterodimer was detected as a complex migrating faster than the TCF4-B homodimer (Fig. 5D). The binding of the ASCL1 and TCF4-B heterodimer to the μE5 E-box was specific since 10-fold excess of unlabelled wt but not mutant μE5 E-box oligonucleotides out-competed the labelled probe. DNA-binding of ASCL1 and TCF4-B heterodimer was abolished by R578H, R580W and R582P mutations and severely reduced by R576Q and S653Lfs*57 mutations (Fig. 5D). Compared with wt TCF4 that was detected predominantly in heterodimeric complex with ASCL1, the levels of DNA-bound heterodimers were decreased in favour of the TCF4-B homodimers by D535G, S653* and S653Lfs*57 mutations (Fig. 5D). Notably, TCF4 A614V mutant, and to a minor extent also R576Q mutant, was able to bind the μE5 E-box in the heterodimeric complex with ASCL1, in contrast to impaired or absent DNA-binding of the mutants in the homodimeric or intra-TCF4 heterodimeric complexes.

Finally, we examined the effect of PTHS-associated mutations on the DNA-binding ability of NEUROD2 and TCF4-B heterodimer. In these assays, we used oligonucleotides containing the insulinoma-associated antigen-1 promoter (IA-1) E-box CATCTG that is known to be bound by NEUROD1 and E47 heterodimers (54). As shown in Figure 5E, we did not detect the binding of the NEUROD2 homodimer, but were able to discriminate the binding of NEUROD2 and wt TCF4-B heterodimer that migrated faster than the TCF4-B homodimer. The specificity of the heterodimer binding to the IA-1 E-box was demonstrated by out-competing the probe with 20-fold excess of unlabelled wt and not mutant IA-1 E-box oligonucleotides. The DNA-binding ability of NEUROD2 and TCF4-B heterodimers was below the detection limit in case of R580W and R582P mutants, and very low in case of R576Q, R578H and S653Lfs*57 mutants (Fig. 5E). All other studied TCF4-B mutants were competent in DNA binding in the heterodimeric complex with NEUROD2.

In sum, these findings demonstrate that the studied PTHS-associated mutations affect the DNA-binding ability of TCF4 differentially: first, mutations of arginines in the bHLH domain and the frameshift mutation in the C-terminal part of TCF4 impede DNA binding; second, the dimerization interface mutation impairs the binding of intra-TCF4 dimers but allows the binding of TCF4 heterodimers with ASCL1 and NEUROD2; third, the nonsense and frameshift mutations in the C-terminal part of TCF4 and the mutation between the bHLH and AD2 domains change the ASCL1 and TCF4 heterodimer versus TCF4 homodimer binding preference; and fourth, the mutation in the AD2 domain has no effect on DNA binding.

Interaction of mutant bHLH domain containing TCF4 proteins with cellular chromatin is decreased

To gain insight into the capacity of the PTHS-associated TCF4 proteins to interact with chromatin in cells, we separated the extracts from transfected HEK293 cells into 0.5% Triton X-100 soluble fraction that contains cytosol and nucleosol, and insoluble pellet fraction that includes chromatin (Fig. 6A). In case of the severely destabilized S653Lfs*57 mutant, we treated transfected HEK293 cells with proteasome inhibitor MG132 for 7 h before the fractionation. The material from equal number of cells for both fractions was subjected to western blot analysis with TCF4-specific antibodies (Fig. 6B). To monitor sample-specific fractionation efficiency, we used endogenous tubulin β and CREB1 (cAMP responsive element binding protein 1) as markers for cytoplasm and chromatin, respectively. As an additional control, we determined the fractionation profile of an overexpressed nucleoplasmic protein, EGFP fused to TCF4 NLS (45). The proportions of the studied proteins in soluble and pellet fractions were determined by densitometric quantification (Fig. 6C). As expected, most of tubulin β and EGFP-NLS was extracted into the soluble fraction, whereas CREB1 remained in the pellet fraction, confirming the liability of the method. Compared with wt, G358V and D535G TCF4-B$^-$, 90–100% of which co-distributed with chromatin, the fractionation profile differed significantly in case of R576Q, R578H, R580W, R582P and A614V mutants, 35–75% of which were extracted into the soluble fraction. Similar to wt TCF4-B$^+$, virtually all of the S653* mutant protein distributed to the chromatin-containing fraction. MG132 treatment led to a similar ∼10% increase in the amount of 0.5% Triton X-100 soluble TCF4-B$^+$ in HEK293 cells expressing either wt or S653Lfs*57 mutant TCF4-B$^+$. Overall, these results, demonstrating that PTHS-associated mutations in the bHLH domain lead to reduced co-fractionation of TCF4-B with cellular chromatin, are in good accordance with their impairing effects on the ability to bind DNA *in vitro*. The S653Lfs*57 mutant remained in the insoluble fraction despite defective DNA binding probably because of aggregation.

PTHS-associated TCF4 mutants have defects of varying extent in the capacity to activate transcription

As a final step, we investigated the impact of the PTHS-associated mutations on the ability of TCF4 to activate transcription. For this, we performed reporter assays using HEK293 cells transfected with firefly luciferase *luc2P* vector

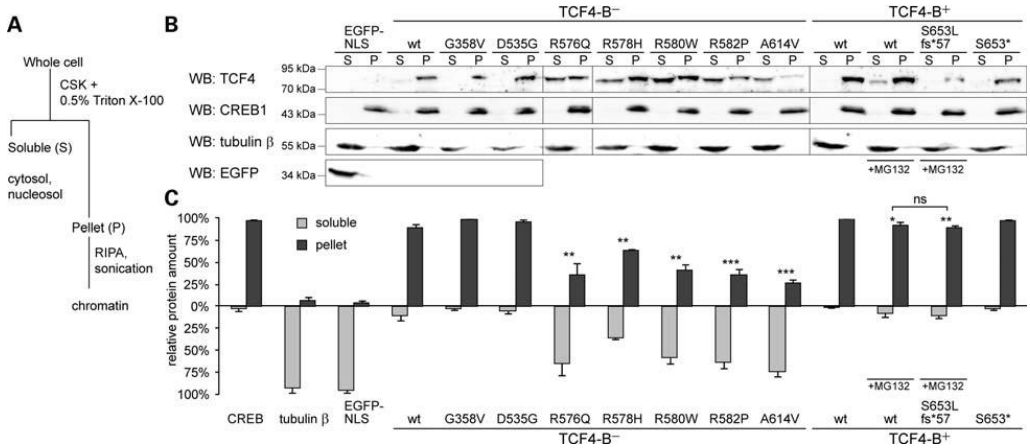


Figure 6. Co-fractionation of TCF4 mutants with cellular chromatin. (**A**) Schematic of cell extraction protocol. (**B**) Western blot analysis of nucleosol- and cytosol-containing soluble (S) and chromatin-containing pellet (P) fractions from HEK293 cells overexpressing EGFP-NLS, wt or mutant TCF4-B proteins. Endogenous tubulin β and CREB1 served as fractionation controls for soluble and pellet fractions, respectively. Molecular mass markers are shown at the left. Cells were treated with MG132 for 7 h where indicated. (**C**) Quantification of tubulin β, CREB1, EGFP-NLS and TCF4 levels in (B) from three independent experiments. Values are given as proportions of protein in soluble and pellet fractions for each protein. Error bars indicate standard deviations. Statistical significance shown with asterisks is relative to the proportions of the respective wt TCF4-B or between the bars connected with lines. CSK, cytoskeleton buffer; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; ns, not significant; *t*-test.

carrying 12 μE5 E-boxes in front of the minimal promoter and wt or mutant TCF4-B encoding constructs. *Renilla* luciferase *hRLucP* construct with the minimal promoter was co-transfected for normalization. As shown in Figure 7A, the normalized reporter activity measured from cells expressing wt, G358V or D535G mutant TCF4-B^{-}, and wt or S653* mutant TCF4-B^{+}, was more than 160 times higher than in empty vector transfected cells. Compared with the respective wt TCF4-B, the transactivation fold was severely reduced by A614V mutation and essentially lost by R576Q, R578H, R580W, R582P and S653Lfs*57 mutations.

Next, we studied the ability of wt and PTHS-associated mutant TCF4-B proteins to activate μE5 E-box-dependent transcription in a heterodimeric complex with ASCL1. Reporter assays were performed with HEK293 cells overexpressing either ASCL1 or TCF4-B, or both proteins. When TCF4-B proteins were expressed individually, the results obtained were similar to the results described above, but since less TCF4-B construct was transfected, the transactivation folds were smaller (Fig. 7B). Compared with empty vector-transfected cells, a small but significant increase in normalized luciferase activity was observed in ASCL1-expressing cells ($6.6 + 3.0$, -2.0; $P < 0.01$, $n = 4$), whereas co-expression of wt TCF4-B and ASCL1 led to a robust, ∼1000-fold activation of reporter transcription (Fig. 7B). Compared with the effect of wt TCF4-B in combination with ASCL1, the induced reporter levels were similar in case of G358V, D535G and A614V mutations, decreased in case of S653* and R576Q mutations, severely reduced in case of S653Lfs*57 mutation and not induced at all relative to the impact of ASCL1 alone in case of R578H, R580W and R582P mutations.

We used the data in Figure 7B to calculate the cooperation index for each wt and PTHS-associated mutant TCF4-B protein. The index expresses how many times the transactivation fold is increased in cells co-expressing TCF4-B and ASCL1 compared with the sum of transactivation folds from cells expressing both proteins separately. The index value of 1.0 is indicative of simple summation of the independent effects of both proteins, whereas a value above 1.0 implies synergism and below 1.0 antagonism. The cooperation index was almost 14 for wt TCF4-B^{-} and almost 6 for wt TCF4-B^{+}, indicating synergism with ASCL1 (Fig. 7C). Relative to the respective wt TCF4-B isoform, the index was increased 1.4-fold by G358V, 2.7-fold by R576Q and 6.7-fold by A614V mutation, whereas S653Lfs*57 mutation reduced the index 1.8-fold. In case of R578H, R580W and R582P mutants, the index did not differ from or was below 1.0, meaning that these mutants did not act synergistically with ASCL1; conversely, a small but statistically significant antagonistic effect was revealed for R582P mutant (Fig. 7C).

Since A614V and S653Lfs*57 mutant proteins are destabilized in HEK293 cells, we asked whether proteasome inhibition has an effect on transactivation capacity of these mutants. To answer this, we treated transfected HEK293 cells with MG132 for 7 h before measuring the reporter activities. As both luciferases were stabilized by MG132 treatment, the normalized values were not significantly increased in empty vector-transfected MG132-treated cells compared with untreated cells. Importantly, proteasome inhibition did not rescue the impaired transactivation capacity of A614V mutant expressed individually or S653Lfs*57 mutant expressed alone or together with ASCL1 (Fig. 7D and E). These data suggest that deficiencies other than reduced

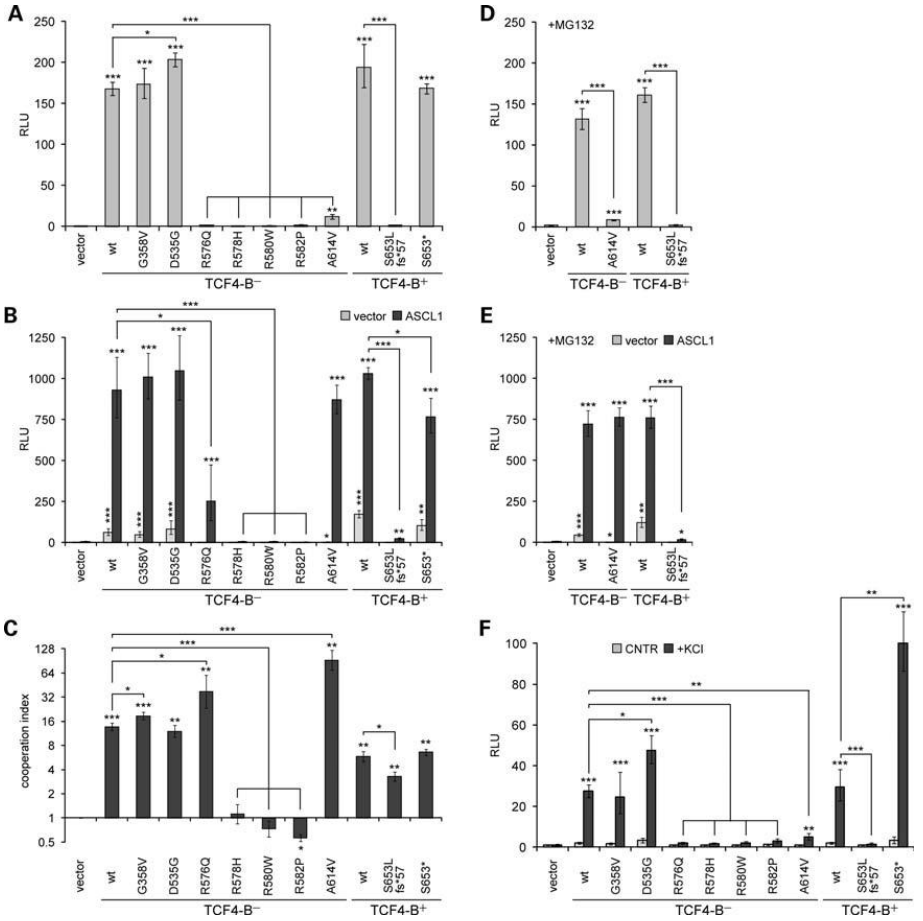


Figure 7. Activation of E-box-controlled transcription by TCF4 mutants in cultured cells. (**A**, **B**, **D** and **E**) Luciferase reporter assays with wt and mutant TCF4-B proteins overexpressed in HEK293 cells alone (A and D) or together with ASCL1 (B and E). Cells were co-transfected with firefly luciferase construct carrying 12 μE5 E-boxes in front of the minimal promoter and Renilla luciferase construct with minimal promoter for normalization. Cells were left untreated (A and B) or treated with MG132 for 7 h (D and E). (**C**) Index of cooperation between wt or mutant TCF4-B proteins and ASCL1 calculated using the data in (B). (**F**) Luciferase reporter assays with wt and mutant TCF4-B proteins overexpressed in rat primary cortical and hippocampal neurons. Cells were co-transfected with firefly luciferase construct carrying 12 μE5 E-boxes in front of TK promoter and Renilla luciferase construct with the EF1α promoter for normalization. Neurons were left untreated (CNTR) or treated with 25 mM KCl for 8 h. Luciferase activities were measured and data are presented as fold induced levels above the signals obtained from empty vector-transfected cells in (A), B and (D)–(F). Shown are the mean results from three independent experiments performed in duplicates. Error bars indicate standard deviations; statistical significance shown with asterisks is relative to the empty vector-transfected cells or cells overexpressing only ASCL1 [dark bars in (B) and (C)] or between the bars connected with lines in (A)–(F). $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *t*-test. RLU, relative luciferase unit.

expression levels are sufficient to damage the ability of A614V and S653Lfs*57 mutants to activate transcription.

Given that neurons are the disease-relevant cells in PTHS, we carried out similar reporter assays as described above, using cultures of rat primary cortical and hippocampal neurons transfected with firefly luciferase *luc2P* vector carrying 12 μE5 E-boxes in front of the thymidine kinase (TK) promoter and *Renilla* luciferase *hRLucP* construct with the elongation factor 1α (EF1α) promoter for normalization. We measured luciferase activities in neurons grown in basal conditions and in neurons treated with 25 mM KCl for 8 h to provoke membrane depolarization and Ca^{2+} signalling. More than 10-fold upregulation of reporter levels was induced by KCl treatment in wt TCF4-B expressing neurons (Fig. 7F). This suggests that the activation of μE5 E-box-controlled transcription by TCF4 overexpression in neurons is dependent on neuronal activity. Work is in progress to investigate the mechanisms underlying this phenomenon; however, increased expression of CMV promoter-controlled TCF4 following depolarization cannot be ruled out. Nevertheless, comparison

of wt and PTHS-associated mutant TCF4-B proteins by the ability to activate transcription in depolarized primary neurons revealed that the reporter levels were severely decreased by R576Q, R578H, R580W, R582P, A614V and S653Lfs*57 mutations, similar to the results obtained in HEK293 cells.

Of note, we additionally analysed G358V and D535G mutations, which did not impair transactivational capacity of TCF4-B, in the context of shorter TCF4-A^{-} isoform, and G358V mutation in the context of a protein where TCF4 AD2 is fused to heterologic GAL4 DNA-binding domain. We did not observe any significant impairment in the ability of the G358V and D535G mutants to activate transcription from μE5 E-boxes or GAL4-binding sites containing promoter, respectively, in HEK293 cells or rat primary neurons (Supplementary Material, Fig. S4).

Collectively, the results from the reporter assays are consistent with the *in vitro* DNA binding data. First, the R578H, R580W, R582P and S653Lfs*57 mutations impair both DNA-binding and transactivation ability of TCF4. Second, the deficiencies of R576Q and A614V mutants in DNA binding as well as transactivation capacity are partially or fully rescued when in heterodimeric complex with ASCL1. This recovery is highlighted by the increased cooperation index of R576Q and A614V mutants. Third, the disfavouring of ASCL1 and the TCF4 heterodimer versus the TCF4 homodimer in case of S653Lfs*57 mutation in DNA-binding assays correlates with the reduced cooperation index of S653Lfs*57 mutant in reporter assays.

DISCUSSION

In this study, we sought to gain a better understanding of the molecular mechanisms underlying PTHS. First, we mapped all published PTHS mutations to the recently described *TCF4* gene structure and found that some of the PTHS-associated deletions and truncating mutations do not damage all *TCF4* alternative transcripts and thus, lead to only partial loss-of-function of the affected allele. Second, we carried out functional analyses of PTHS-associated reading frame elongating and missense mutations and demonstrated that the resultant TCF4 proteins are impaired to variable extent and via different mechanisms, including protein destabilization, alteration of dimerization preferences and loss of DNA-binding and transactivation ability. Additionally, our analyses provide insights into how residues outside of the basic region can play a role in the DNA-binding ability of TCF4, and residues outside of the bHLH domain may affect dimerization properties of TCF4.

It has been found that the DNA-binding geometry of different bHLH dimers is extremely similar, making it possible to compare the contacts made by conserved residues with the E-box containing DNA (55). In our structure models of DNA-bound TCF4 homo- and heterodimer, the conserved arginines that are located in the basic DNA-binding region and are found to be mutated in PTHS make direct contacts with DNA. The only exception is R569, which does not bind DNA in our models, but since the respective residue has been found to make contacts with E-box flanking DNA in related bHLH co-structures (55), we suggest that this arginine may also contact DNA. Based on seven bHLH co-structures, the estimated average changes in DNA interaction energy induced by *in silico* mutations corresponding to R576Q, R578H and R580W in TCF4 are small (55). Nonetheless, according to our results, loss of even a single DNA contact made by any of these charged arginines is detrimental for DNA binding. Namely, we demonstrated that R576, R578 and R580 are not needed for dimerization, but are important for DNA-binding of TCF4 both *in vitro* and *in vivo*. This is consistent with previous studies demonstrating that mutating the corresponding arginines in bHLH-PAS protein BMAL1 or E-proteins E12 or E47 has no effect on dimerization but impairs DNA binding (56–58).

It has been suggested that two intact basic regions are required for the interaction of bHLH dimers with DNA (58). This is in accordance with our data demonstrating that in case of TCF4 homodimers, R576Q, R578H and R580W mutations lead to impairment of DNA binding regardless of whether the mutation is present in one or both subunits. However, we also showed that in case of heterodimers, not all mutations in the TCF4 basic region arginines are *trans*-dominant. Specifically, although R578H and R580W mutants impaired DNA-binding of TCF4 heterodimers with ASCL1 or NEUROD2, the R576Q mutation in TCF4 did not eliminate the DNA-binding ability of TCF4-ASCL1 heterodimer. Correspondingly, R578H and R580W mutants were not able to cooperate with ASCL1 in activating μE5 E-box-dependent transcription in HEK293 cells, whereas the transactivation deficiency of R576Q mutant was partially rescued in heterodimeric complex with ASCL1. However, since R576 can contact an E-box-flanking base, it is possible that the binding specificity of the mutant TCF4 containing heterodimers could be altered *in vivo*. Our data are consistent with reporter assays performed with R580 mutations encompassing TCF4-ASCL1 heterodimers in previous studies (5,9), but differ from the results obtained with R576G substitution that did not allow TCF4 to cooperate with ASCL1 in activating transcription from the *Delta1* promoter in SKNBE(2)C cells (9). The latter could reflect substitution-, cell-line- or promoter-specific effects. Notably, across all bHLH families, the position corresponding to R576 in TCF4 is less conserved than other arginines in the basic region, and substitutions corresponding to PTHS-associated mutations R576G or R576Q in TCF4 are present in some natural bHLH proteins (55). In sum, as the PTHS-associated mutations in the basic region arginines impair the DNA-binding and transactivation ability of TCF4 to a varying degree, these mutations should not be regarded as functional equivalents.

We determined that, similar to the arginines in the basic region, R582 located in the beginning of helix 1 of TCF4 HLH domain was dispensable for dimerization but essential for DNA binding. Accordingly, the R582P mutant was not able to activate transcription either as a homo- or heterodimer in rat primary neurons or HEK293 cells. Furthermore, TCF4 with R582P mutation antagonized ASCL1-induced transcription, indicating that this mutant acts in a dominant-negative manner. Since proline destabilizes α-helices (59), we propose that R582P mutation interferes with the conformational change of the basic region from a random coil to a

structured α-helix that has been suggested to be induced in bHLH factors upon DNA-binding in solution (60). How mutations outside the basic region can influence DNA-binding of TCF4 is additionally exemplified by the A614V mutation located in the middle of helix 2 of the HLH domain. We observed that A614V mutation leads to overall reduced stability of TCF4 and has differential effects on TCF4 homo- and heterodimers. Specifically, the mutation severely reduced homodimerization and DNA-binding of homodimers in a *trans*-dominant manner, whereas it had no effect on heterodimerization, DNA-binding of TCF4-ASCL1 and TCF4-NEUROD2 heterodimers or activation of μE5 E-box-dependent transcription by TCF4-ASCL1 dimers. Our results are in agreement with a previous study showing that mutating the alanine, corresponding to A614 in TCF4, to aspartate in E47 does not abrogate the formation of homodimers, but leads to weakened DNA-binding ability of homodimers (58). The alanine at the position corresponding to A614 in TCF4 is found in proteins from all bHLH families, except bHLH-PAS and Id factors (61), indicating that the packing interactions formed by this residue at the dimer interface could be influential in majority of bHLH homo- and heterodimers. However, it is plausible that the extended helix 1 of E-proteins places more constraints on this position in homodimers, whereas the substitution is tolerated in heterodimers because of shorter helix 1 of the partner protein. All this suggests that, in E-proteins, the alanine corresponding to A614 in TCF4 is not essential for heterodimer formation, but possibly contributes to the dynamic properties of HLH homodimers that in turn influences the efficiency of conformational transition of the basic region required for DNA-binding.

We studied two PTHS-associated missense mutations, G358V and D535G, located upstream of the bHLH domain and one reading frame elongating mutation, S653Lfs*57, downstream of the bHLH domain. No major deficiencies were revealed for G358V or D535G mutant TCF4 proteins that were able to dimerize, bind DNA and activate E-box-dependent transcription in HEK293 cells and rat primary neurons. On the other hand, the S653Lfs*57 elongation triggered destabilization and aggregation of TCF4, and impaired the ability of homo- and heterodimers to bind DNA and to activate transcription in a *trans*-dominant manner. Gain of 56 amino acids and not loss of 19 native C-terminal amino acids was responsible for these deficiencies since they were not observed for TCF4 with nonsense mutation S653*, which was analysed as a control. Notably, since the same C-terminal amino acids are introduced by another PTHS-associated elongating mutation A634Dfs*74, which was described during the preparation of this manuscript (21), we suggest that the pathogenic mechanisms of the elongating mutations might be similar. We additionally showed that compared with wt TCF4, the S653Lfs*57, S653* and D535G mutants display shifted affinity of homo- versus heterodimerization when translated together with ASCL1. Namely, in co-translation and EMSA experiments, we detected wt TCF4 predominantly in DNA-bound heterodimeric complexes with ASCL1 or NEUROD2, whereas in case of S653 mutants and D535G mutant the prevalence of heterodimers was lost or even surpassed by TCF4 homodimers when translated together with ASCL1, but not with NEUROD2. Of note, since the DNA probe was kept in excess relative to the proteins in shift assays, the ratio of bound dimers is expected to reflect the ratio of formed dimers. The change in homodimer–heterodimer equilibrium was less evident in reporter assays, however, which showed that the cooperation between TCF4 and ASCL1 was reduced only by S653Lfs*57 mutation. It is possible that slight differences in dimerization preferences are less influential in experiments with overexpressed proteins, but are of importance in an endogenous context. Several previous studies have demonstrated that regions outside the HLH domain affect dimerization of E-proteins. For instance, amino acids C-terminal to the bHLH domain are required for the dimerization of E-proteins *in vivo* (62), acidic sequences preceding the bHLH domain impact the dimerization specificity of E12 (63), insertion upstream of the bHLH domain interferes with dimerization and DNA-binding of REB (64) and phosphorylation of a residue outside of the bHLH domain regulates dimerization specificity of E47 (65,66). Together with our results, this implies that even small changes outside of the bHLH domain could have profound effects on dimerization preferences of E-proteins.

Our finding that D535G mutant has only a minor deficiency and G358V mutant functions similarly to wt TCF4 raises questions of whether and how these mutations are responsible for the PTHS phenotype in the patients. Further studies are needed to uncover the pathogenic mechanisms of these mutations and to verify that the patients do not carry any additional mutations that may contribute to the disorder. Then again, the results presented here also indicate that even partial loss-of-function of the mutated *TCF4* allele is sufficient to cause PTHS. This is particularly evident from the PTHS-associated deletions and premature stop-inducing mutations that leave some of the TCF4 isoforms intact. There are several non-exclusive explanations for the underlying mechanisms. First, it is possible that all *TCF4* alternative promoters are needed to produce sufficient quantity of the protein to maintain normal function. Second, the deficit of AD1 and/or NLS containing TCF4 isoforms might not be compensated by the remaining isoforms that lack these functional regions. Third, the cell-type- and age-dependent regulation of *TCF4* alternative promoters could also play a role in the functional divergence of different isoforms. Given that a patient with a translocation that leaves 12 of the 21 5′ exons in place has a milder phenotype (48), we tried to analyse whether the PTHS patients with mutations that result in partial loss of TCF4 isoforms have less severe symptoms. Unfortunately, such correlations are difficult to make since systematic graded characterization of the patients is not available. However, in a recent study where a PTHS clinical diagnosis score was introduced, the patient with hypomorphic deletion had the lowest relative score (18). Therefore, it is possible that a continuum of phenotypes from moderate mental retardation to full-scale PTHS might be caused by different mutations in the *TCF4* gene. This would mean that, in addition to PTHS, *TCF4* should be considered a candidate gene in cases of milder mental retardation that could be caused by mutations in individual TCF4 5′ exons or in regulatory elements of the alternative promoters. Additionally, given that different phenotypes have been described in two siblings carrying the same mutation in *TCF4* gene and born to a healthy mosaic

mother, genetic background could also have an impact on disease severity (21). Furthermore, there is considerable variability of symptoms among the PTHS patients and many of the features considered characteristic of the disorder are not uniformly present in all patients. For instance, <60% of patients have hyperventilation episodes and <40% have seizures (18). The only phenotype–genotype correlation suggested so far is that individuals with missense mutations are more likely to develop seizures than other PTHS patients (11), but this was disproved in a recent study (18). Our results demonstrate that impacts of PTHS-causing mutations vary in severity from partial loss-of-function to dominant-negative effects and functionally unequal mutations are found among deletions and truncating mutations as well as missense mutations in *TCF4*. This provides a starting point for making PTHS phenotype–genotype correlations that would not be based on mutation type solely, but would take into account the functional impacts of the mutations.

MATERIALS AND METHODS

Comparative modelling of protein structures

The coordinates of 2.8 Å crystal structure of E47 bHLH dimer complexed with DNA were provided by T. Ellenberger (49), and the coordinates of 2.5 Å crystal structure of NEUROD1 and E47 bHLH heterodimer bound to DNA were downloaded from Protein Data Bank (PDB code: 2QL2) (50). The structures were employed as templates for the generation of the TCF4 bHLH homodimer and NEUROD2-TCF4 bHLH heterodimer models using Modeller 9v9 (67). For modelling the NEUROD2-TCF4 bHLH heterodimer, both biological units of the 2QL2 asymmetric unit were employed as templates. The DNA geometry was restrained rigid and 100 models were generated for each complex. The models were assessed using DOPE-HR and QMEAN (67,68) scores, and the ones with the best scores and preserved DNA-binding geometry were selected for presentation. Figures of structures were generated using the PyMOL Molecular Graphics System, Version 1.4.1 (Schrödinger).

Constructs

pCDNA3.1 (Invitrogen) constructs encoding human TCF4 isoforms A^- and B^-, the pQM (Icosagen) constructs encoding C-terminally E2-tagged mouse NEUROD2 and GAL4-AD2 fusion proteins, the pEGFP (Clontech)-based constructs encoding human TCF4 NLS or bHLH containing EGFP fusion proteins and pGL4.29 (Promega)-derived constructs pGL4.29[luc2P/12μE5/Hygro] and pGL4[hRlucP/min/Hygro] have been described previously (45). The pG5luc vector that carries GAL4-binding sites is from Promega. For pGL4[luc2P/12μE5/TK/Hygro], the minimal promoter in pGL4.29[luc2P/12μE5/Hygro] was replaced with the TK promoter from pRL-TK (Promega). For pGL4.83[hRlucP/EF1α/Puro], the EF1α promoter (from −218 to +995 relative to the transcription start site) was PCR-amplified from human genomic DNA and inserted into pGL4.83[hRlucP/Hygro] (Promega). Coding sequences of human ASCL1 and TCF4-B^+ S653* and the coding and 3′ UTR sequence of TCF4-B^+ were amplified by PCR from human brain cDNA and inserted into pCDNA3.1. For pCDNA-Neurod2-E2, Neurod2-E2 sequence from pQM-Neurod2-E2 was inserted into pCDNA3.1 vector behind the T7 promoter. For pACT-bHLH, the sequence coding TCF4-B^- amino acids 541–667 was inserted into pACT (Promega) behind the GAL4 NLS and VP16 transactivation domain-encoding sequence. PTHS-associated mutations to TCF4 coding sequence were introduced by PCR with complementary primers against the target sequence containing the respective mutation and Phusion High-Fidelity DNA Polymerase (Finnzymes). The missense mutations were introduced into pCDNA-TCF4-B^-; for S653Lfs*57 mutation, pCDNA-TCF4-B^+ 3′UTR was used as a template. The mutations containing sequences were subcloned into pEGFP-bHLH, pCDNA-TCF4-A^- or pQM-GAL4-AD2-E2. All constructs were verified by sequencing. The primers used are listed in Supplementary Material, Table S1.

In vitro translation and electrophoretic mobility shift assay

In vitro-translated proteins were produced using TnT T7 Quick Coupled Transcription/Translation System (Promega) according to manufacturer's instructions, using unlabelled methionine. In case of co-translations, the ratios of constructs were as follows—pCDNA-TCF4-B:pACT-bHLH 2:1, pCDNA-TCF4-B:pCDNA-ASCL1 or pCDNA-Neurod2-E2 1:2. μE5 and I-A1 E-box oligonucleotides (Supplementary Material, Table S1) were ^{32}P-labelled with T4 polynucleotide kinase (Fermentas) as suggested by the manufacturer. For binding reactions, 1 μl of translated protein mixture was incubated with 0.1 pmol of labelled annealed oligonucleotides in 15 μl of reaction buffer (20 mM HEPES-KOH, pH 7.9, 50 mM KCl, 5% glycerol, 1 mM EDTA, 1 mM dithiothreitol, 13.3 ng/μl poly(dI-dC)) for 20 min at room temperature. Where indicated, 1–2 pmol of unlabelled annealed oligos were included in the reaction. The DNA–protein complexes were resolved in 5% non-denaturing polyacrylamide gel and visualized by autoradiography.

Cell culture and transfection

Human embryonic kidney HEK293 cells were grown in Eagle's minimum essential medium (PAA) supplemented with 10% fetal bovine serum (PAA), 100 U/ml penicillin (PAA) and 0.1 mg/ml streptomycin (PAA) at 37°C in 5% CO_2. For rat hippocampal and cortical mixed neuronal cultures, E22.5 embryos were obtained from time-mated pregnant Sprague–Dawley rats. All animal procedures were performed in compliance with the local ethics committee. The cells were attained as described previously (69) and plated onto poly-L-lysine (Sigma)-coated plates in Neurobasal A medium (Gibco) supplemented with 1 mM L-glutamine (PAA), B27 (Gibco), 100 U/ml penicillin (PAA) and 0.1 mg/ml streptomycin (PAA). Mitotic inhibitor 10 μg/ml 5-fluoro-2′-deoxyuridine (Sigma) was added to the medium at 2 DIV.

For transfection of HEK293 cells, 0.375 μg of DNA and 0.75 μl of LipoD293 reagent (SignaGen) were used per well of a 48-well plate or scaled up accordingly. Neuronal cultures

were transfected at 6–7 DIV, using 0.5 μg of DNA and 1 μl of Lipofectamine 2000 reagent (Invitrogen) per well of a 48-well plate. Forty hours post-transfection, 25 mM KCl was added to the culture medium for 8 h where indicated. In case of co-transfections, the ratios of transfected constructs were as follows—pCDNA-TCF4-B:pGL4.83[hRlucP/EF1α/Puro] 30:1, pEGFP-bHLH:pCDNA-ASCL1 1:2, pCDNA-TCF4-A$^-$:pCDNA-ASCL1 2:1, pCDNA-TCF4-A$^-$ or pEGFP-bHLH:pQM-Neurod2-E2 2:1, pGL4.29[luc2P/12μE5/Hygro] or pG5luc:pGL4[hRlucP/min/Hygro]:pCDNA-TCF4-B or pCDNA-TCF4-A$^-$ or pQM-GAL4-AD2-E2 1:1:1, pGL4.29[luc2P/12 μE5/Hygro]:-pGL4[hRlucP/min/Hygro]:pCDNA-TCF4-B:pCDNA-ASCL1 3:3:2:1 and pGL4[luc2P/12μE5/TK/Hygro] or pG5luc:pGL4.83[hRlucP/EF1α/Puro]:pCDNA-TCF4-B or pCDNA-TCF4-A$^-$ or pQM-GAL4-AD2-E2 25:1:25.

Reverse transcription and quantitative PCR

RNA was purified with RNeasy Plus Mini Kit (Qiagen) and treated with TURBO DNase (Ambion). cDNAs were synthesized using Superscript III reverse transcriptase (Invitrogen) and oligo(dT) primers. Quantitative PCR was performed with LightCycler 480 DNA SYBR Green I Master reagent (Roche), LightCycler 2.0 engine (Roche) and polycarbonate qPCR capillaries (Bioron) in a volume of 10 μl containing 1/100 of reverse transcription reaction as a template. The primers used are given in Supplementary Material, Table S1.

Cell extracts

Whole-cell lysates were prepared by extraction in RIPA buffer [50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS, 1 mM dithiothreitol, 1 mM phenylmethanesulfonylfluoride, protease inhibitors cocktail Complete (Roche)]. The extracts were sonicated briefly and centrifuged to remove insoluble debris. Protein concentrations were determined using BCA assay (Pierce). Equal amounts of proteins were separated in 8–10% SDS–PAGE and transferred to PVDF membrane (BioRad). Subcellular fractionation was performed essentially as described earlier (70). Briefly, cells were extracted twice with CSK (cytoskeleton) buffer [10 mM PIPES, pH 7.0, 100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 1 mM EGTA, 1 mM dithiothreitol, 1 mM phenylmethanesulfonylfluoride, protease inhibitors cocktail Complete (Roche)] containing 0.5% Triton X-100. After 5 min incubation on ice, soluble and insoluble fractions were separated by centrifugation at 5200*g* for 3 min. The first and the second soluble fractions were combined. The pellet was suspended in RIPA buffer and sonicated briefly. Protein from equivalent number of cells was loaded to SDS–PAGE.

Western blotting

For western blotting, the antibodies were diluted in 2% skim milk and 0.1% Tween 20 in PBS as follows: rabbit polyclone anti-TCF4/ITF2 (CeMines) 1:1000, mouse monoclone anti-tubulin β (Developmental Studies Hybridoma Bank) 20 ng/ml, rabbit polyclone anti-CREB1 (Upstate) 1:2000, rabbit polyclone anti-EGFP (71) 1:10^6, HRP-conjugated goat anti-mouse/rabbit IgG (Pierce) 1:5000. Chemiluminescent signal was detected using SuperSignal West Femto Chemiluminescent Substrate (Pierce) and ImageQuant LAS 4000 CCD camera system (GE Healthcare). Signals from non-saturated exposures were quantified using the ImageQuantTL software (GE Healthcare). Background signals were subtracted by the rolling ball algorithm and volumes of each band were calculated.

Immunocytochemistry

Transfected cells grown on poly-L-lysine (Sigma)-coated cover slips were fixed in 4% paraformaldehyde, quenched with 50 mM NH_4Cl in PBS, permeabilized in 0.5% Triton X-100 in PBS and blocked with 2% bovine serum albumin (BSA) in PBS. The primary and secondary antibodies were diluted in 0.2% BSA and 0.1% Tween 20 in PBS as follows: rabbit polyclone anti-TCF4/ITF2 (CeMines) 1:200, goat polyclone anti-TCF4 (Abcam) 1:100, rabbit polyclone anti-ASCL1/Ash1 (CeMines) 1:200, mouse monoclone anti-E2 (Icosagen, 5E11) 1:500, Alexa 488- or Alexa 568-conjugated $F(ab')_2$ fragment of goat anti-mouse/rabbit IgG or rabbit anti-goat IgG (Molecular Probes) 1:2000. Cover slips were washed three times with 0.1% Tween 20 in PBS after both reactions. ProLong Gold antifade reagent with DAPI (Molecular Probes) was used for mounting, and the samples were analysed by confocal microscopy (LSM Duo, Zeiss).

Reporter assay

Luciferase assays and data analysis were performed as described previously (45). Briefly, reporter signals were measured 24–48 h post-transfection using Passive Lysis Buffer (Promega) and Dual-Glo Luciferase assay (Promega). The co-operation index was calculated as described earlier (72). For data analysis, background subtraction, normalization, log-transformation, autoscaling and *t*-tests were performed. For graphical representation, the data were back-transformed to the original scale.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Epp Väli and Maila Rähn for excellent technical assistance, Kristi Luberg and Kaur Jaanson for help in protein modelling, Indrek Koppel and Kaja Kannike for critical discussions and Andres Merits for the EGFP antibody.

Conflict of Interest statement. The authors declare no conflict of interest.

FUNDING

This work was supported by Estonian Ministry of Education and Research (0140143) and Estonian Science Foundation (7257, 8844).

REFERENCES

1. Pitt, D. and Hopkins, I. (1978) A syndrome of mental retardation, wide mouth and intermittent overbreathing. *Aust. Paediatr. J.*, **14**, 182–184.
2. Peippo, M.M., Simola, K.O., Valanne, L.K., Larsen, A.T., Kahkonen, M., Auranen, M.P. and Ignatius, J. (2006) Pitt-Hopkins syndrome in two patients and further definition of the phenotype. *Clin. Dysmorphol.*, **15**, 47–54.
3. Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N.P., Lyonnet, S., Munnich, A. *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.
4. Brockschmidt, A., Todt, U., Ryu, S., Hoischen, A., Landwehr, C., Birnbaum, S., Frenck, W., Radlwimmer, B., Lichter, P., Engels, H. *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.
5. Zweier, C., Peippo, M.M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., Reardon, W., Saraiva, J., Cabral, A., Gohring, I. *et al.* (2007) Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am. J. Hum. Genet.*, **80**, 994–1001.
6. Andrieux, J., Lepretre, F., Cuisset, J.M., Goldenberg, A., Delobel, B., Manouvrier-Hanu, S. and Holder-Espinasse, M. (2008) Deletion 18q21.2q21.32 involving TCF4 in a boy diagnosed by CGH-array. *Eur. J. Med. Genet.*, **51**, 172–177.
7. Giurgea, I., Missirian, C., Cacciagli, P., Whalen, S., Fredriksen, T., Gaillon, T., Rankin, J., Mathieu-Dramard, M., Morin, G., Martin-Coignard, D. *et al.* (2008) TCF4 deletions in Pitt-Hopkins syndrome. *Hum. Mutat.*, **29**, E242–E251.
8. Zweier, C., Sticht, H., Bijlsma, E.K., Clayton-Smith, J., Boonen, S.E., Fryer, A., Greally, M.T., Hoffmann, L., den Hollander, N.S., Jongmans, M. *et al.* (2008) Further delineation of Pitt-Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. *J. Med. Genet.*, **45**, 738–744.
9. de Pontual, L., Mathieu, Y., Golzio, C., Rio, M., Malan, V., Boddaert, N., Soufflet, C., Picard, C., Durandy, A., Dobbie, A. *et al.* (2009) Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. *Hum. Mutat.*, **30**, 669–676.
10. Kato, Z., Morimoto, W., Kimura, T., Matsushima, A. and Kondo, N. (2010) Interstitial deletion of 18q: comparative genomic hybridization array analysis of 46, XX,del(18)(q21.2.q21.33). *Birth Defects Res. A Clin. Mol. Teratol.*, **88**, 132–135.
11. Rosenfeld, J.A., Leppig, K., Ballif, B.C., Thiese, H., Erdie-Lalena, C., Bawle, E., Sastry, S., Spence, J.E., Bandholz, A., Surti, U. *et al.* (2009) Genotype-phenotype analysis of TCF4 mutations causing Pitt-Hopkins syndrome shows increased seizure activity with missense mutations. *Genet. Med.*, **11**, 797–805.
12. Takano, K., Lyons, M., Moyes, C., Jones, J. and Schwartz, C. (2010) Two percent of patients suspected of having Angelman syndrome have TCF4 mutations. *Clin. Genet.*, **78**, 282–288.
13. Hasi, M., Soileau, B., Sebold, C., Hill, A., Hale, D.E., O'Donnell, L. and Cody, J.D. (2011) The role of the TCF4 gene in the phenotype of individuals with 18q segmental deletions. *Hum. Genet.*, **130**, 777–787.
14. Lehalle, D., Williams, C., Siu, V.M. and Clayton-Smith, J. (2011) Fetal pads as a clue to the diagnosis of Pitt-Hopkins syndrome. *Am. J. Med. Genet. A*, **155A**, 1685–1689.
15. Marangi, G., Ricciardi, S., Orteschi, D., Lattante, S., Murdolo, M., Dallapiccola, B., Biscione, C., Lecce, R., Chiurazzi, P., Romano, C. *et al.* (2011) The Pitt-Hopkins syndrome: report of 16 new patients and clinical diagnostic criteria. *Am. J. Med. Genet. A*, **155A**, 1536–1545.
16. Taddeucci, G., Bonuccelli, A., Mantellassi, I., Orsini, A. and Tarantino, E. (2010) Pitt-Hopkins syndrome: report of a case with a TCF4 gene mutation. *Ital. J. Pediatr.*, **36**, 12.
17. Takano, K., Tan, W.H., Irons, M.B., Jones, J.R. and Schwartz, C.E. (2011) Pitt-Hopkins syndrome should be in the differential diagnosis for males presenting with an ATR-X phenotype. *Clin. Genet.*, **80**, 600–601.
18. Whalen, S., Heron, D., Gaillon, T., Moldovan, O., Rossi, M., Devillard, F., Giuliano, F., Soares, G., Mathieu-Dramard, M., Afenjar, A. *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.
19. Armani, R., Archer, H., Clarke, A., Vasudevan, P., Zweier, C., Ho, G., Williamson, S., Cloosterman, D., Yang, N. and Christodoulou, J. (2012) Transcription Factor 4 and Myocyte Enhancer Factor 2C mutations are not common causes of Rett syndrome. *Am. J. Med. Genet. A*, **158A**, 713–719.
20. Maini, I., Cantalupo, G., Turco, E.C., De Paolis, F., Magnani, C., Parrino, L., Terzano, M.G. and Pisani, F. (2012) Clinical and polygraphic improvement of breathing abnormalities after valproate in a case of Pitt-Hopkins syndrome. *J. Child Neurol*, doi: 10.1177/0883073811435917.
21. Steinbusch, C.V., van Roozendaal, K.E., Tserpelis, D., Smeets, E.E., Kranenburg-de Koning, T.J., de Waal, K.H., Zweier, C., Rauch, A., Hennekam, R.C., Blok, M.J. *et al.* (2012) Somatic mosaicism in a mother of two children with Pitt-Hopkins syndrome. *Clin. Genet*, doi: 10.1111/j.1399-0004.2012.01857.x.
22. Verhulst, S.L., De Dooy, J., Ramet, J., Bockaert, N., Van Coster, R., Ceulemans, B. and De Backer, W. (2012) Acetazolamide for severe apnea in Pitt-Hopkins syndrome. *Am. J. Med. Genet. A*, **158A**, 932–934.
23. Ghosh, P.S., Freidman, N.R. and Ghosh, D. (2012) Pitt-Hopkins syndrome in a boy with Charcot Marie Tooth disease type 1A: a rare co-occurrence of 2 genetic disorders. *J. Child Neurol*, doi: 10.1177/0883073812437242.
24. 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.
25. Corneliussen, B., Thornell, A., Hallberg, B. and Grundstrom, T. (1991) Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. *J. Virol.*, **65**, 6084–6093.
26. Zweier, C., de Jong, E.K., Zweier, M., Orrico, A., Ousager, L.B., Collins, A.L., Bijlsma, E.K., Oortveld, M.A., Ekici, A.B., Reis, A. *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.
27. O'Donnell, L., Soileau, B., Heard, P., Carter, E., Sebold, C., Gelfond, J., Hale, D.E. and Cody, J.D. (2010) Genetic determinants of autism in individuals with deletions of 18q. *Hum. Genet.*, **128**, 155–164.
28. Baratz, K.H., Tosakulwong, N., Ryu, E., Brown, W.L., Branham, K., Chen, W., Tran, K.D., Schmid-Kubista, K.E., Heckenlively, J.R., Swaroop, A. *et al.* (2010) E2-2 protein and Fuchs's corneal dystrophy. *N. Engl. J. Med.*, **363**, 1016–1024.
29. Li, Y.J., Minear, M.A., Rimmler, J., Zhao, B., Balajonda, E., Hauser, M.A., Allingham, R.R., Eghrari, A.O., Riazuddin, S.A., Katsanis, N. *et al.* (2011) Replication of TCF4 through association and linkage studies in late-onset Fuchs endothelial corneal dystrophy. *PLoS One*, **6**, e18044.
30. Riazuddin, S.A., McGlumphy, E.J., Yeo, W.S., Wang, J., Katsanis, N. and Gottsch, J.D. (2011) Replication of the TCF4 intronic variant in late-onset Fuchs corneal dystrophy and evidence of independence from the FCD2 locus. *Invest. Ophthalmol. Vis. Sci.*, **52**, 2825–2829.
31. Thalamuthu, A., Khor, C.C., Venkataraman, D., Koh, L.W., Tan, D.T., Aung, T., Mehta, J.S. and Vithana, E.N. (2011) Association of TCF4 gene polymorphisms with Fuchs' corneal dystrophy in the Chinese. *Invest. Ophthalmol. Vis. Sci.*, **52**, 5573–5578.
32. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O.P., Mors, O., Mortensen, P.B. *et al.* (2009) Common variants conferring risk of schizophrenia. *Nature*, **460**, 744–747.
33. Li, T., Li, Z., Chen, P., Zhao, Q., Wang, T., Huang, K., Li, J., Li, Y., Liu, J., Zeng, Z. *et al.* (2010) Common variants in major histocompatibility complex region and TCF4 gene are significantly associated with schizophrenia in Han Chinese. *Biol. Psychiatry*, **68**, 671–673.
34. Steinberg, S., de Jong, S., Andreassen, O.A., Werge, T., Borglum, A.D., Mors, O., Mortensen, P.B., Gustafsson, O., Costas, J., Pietilainen, O.P. *et al.* (2011) Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum. Mol. Genet.*, **20**, 4076–4081.
35. Quednow, B.B., Ettinger, U., Mossner, R., Rujescu, D., Giegling, I., Collier, D.A., Schmechtig, A., Kuhn, K.U., Moller, H.J., Maier, W. *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.
36. Lennertz, L., Rujescu, D., Wagner, M., Frommann, I., Schulze-Rauschenbach, S., Schuhmacher, A., Landsberg, M.W., Franke, P., Moller, H.J., Wolwer, W. *et al.* (2011) Novel schizophrenia risk gene TCF4 influences verbal learning and memory functioning in schizophrenia patients. *Neuropsychobiology*, **63**, 131–136.

37. Williams, H.J., Moskvina, V., Smith, R.L., Dwyer, S., Russo, G., Owen, M.J. and O'Donovan, M.C. (2011) Association between TCF4 and schizophrenia does not exert its effect by common nonsynonymous variation or by influencing *cis*-acting regulation of mRNA expression in adult human brain. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **156B**, 781–784.
38. Brzozka, 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.
39. 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.
40. 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. USA*, **104**, 15382–15387.
41. Brockschmidt, A., Filippi, A., Charbel Issa, P., Nelles, M., Urbach, H., Eter, N., Driever, W. and Weber, R.G. (2011) Neurologic and ocular phenotype in Pitt-Hopkins syndrome and a zebrafish model. *Hum. Genet.*, **130**, 645–655.
42. Massari, M.E. and Murre, C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.*, **20**, 429–440.
43. 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.
44. 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. *Brain Res. Mol. Brain Res.*, **25**, 176–180.
45. 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.
46. Persson, P., Jogi, A., Grynfeld, A., Pahlman, 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.
47. 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.
48. Kalscheuer, V.M., Feenstra, I., Van Ravenswaaij-Arts, C.M., Smeets, D.F., Menzel, C., Ullmann, R., Musante, L. and Ropers, H.H. (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.
49. Ellenberger, T., Fass, D., Arnaud, M. and Harrison, S.C. (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.*, **8**, 970–980.
50. Longo, A., Guanga, G.P. and Rose, R.B. (2008) Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. *Biochemistry*, **47**, 218–229.
51. Atchley, W.R. and Fitch, W.M. (1997) A natural classification of the basic helix-loop-helix class of transcription factors. *Proc. Natl Acad. Sci. USA*, **94**, 5172–5176.
52. Hirsch, M.R., Tiveron, M.C., Guillemot, F., Brunet, J.F. and Goridis, C. (1998) Control of noradrenergic differentiation and Phox2a expression by MASH1 in the central and peripheral nervous system. *Development*, **125**, 599–608.
53. Olson, J.M., Asakura, A., Snider, L., Hawkes, R., Strand, A., Stoeck, J., Hallahan, A., Pritchard, J. and Tapscott, S.J. (2001) NeuroD2 is necessary for development and survival of central nervous system neurons. *Dev. Biol.*, **234**, 174–187.
54. Breslin, M.B., Zhu, M. and Lan, M.S. (2003) NeuroD1/E47 regulates the E-box element of a novel zinc finger transcription factor, IA-1, in developing nervous system. *J. Biol. Chem.*, **278**, 38991–38997.
55. De Masi, F., Grove, C.A., Vedenko, A., Alibes, A., Gisselbrecht, S.S., Serrano, L., Bulyk, M.L. and Walhout, A.J. (2011) Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Res.*, **39**, 4553–4563.
56. Hosoda, H., Motohashi, J., Kato, H., Masushige, S. and Kida, S. (2004) A BMAL1 mutant with arginine 91 substituted with alanine acts as a dominant negative inhibitor. *Gene*, **338**, 235–241.
57. Beltran, A.C., Dawson, P.E. and Gottesfeld, J.M. (2005) Role of DNA sequence in the binding specificity of synthetic basic-helix-loop-helix domains. *Chembiochem*, **6**, 104–113.
58. Voronova, A. and Baltimore, D. (1990) Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proc. Natl Acad. Sci. USA*, **87**, 4722–4726.
59. Pace, C.N. and Scholtz, J.M. (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.*, **75**, 422–427.
60. Fairman, R., Beran-Steed, R.K. and Handel, T.M. (1997) Heteronuclear (1H, 13C, 15N) NMR assignments and secondary structure of the basic region-helix-loop-helix domain of E47. *Protein Sci.*, **6**, 175–184.
61. Chavali, G.B., Vijayalakshmi, C. and Salunke, D.M. (2001) Analysis of sequence signature defining functional specificity and structural stability in helix-loop-helix proteins. *Proteins*, **42**, 471–480.
62. 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.
63. 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.
64. Klein, E.S., Simmons, D.M., Swanson, L.W. and Rosenfeld, M.G. (1993) Tissue-specific RNA splicing generates an ankyrin-like domain that affects the dimerization and DNA-binding properties of a bHLH protein. *Genes Dev.*, **7**, 55–71.
65. Sloan, S.R., Shen, C.P., McCarrick-Walmsley, R. and Kadesch, T. (1996) Phosphorylation of E47 as a potential determinant of B-cell-specific activity. *Mol. Cell. Biol.*, **16**, 6900–6908.
66. Lluis, F., Ballestar, E., Suelves, M., Esteller, M. and Munoz-Canoves, P. (2005) E47 phosphorylation by p38 MAPK promotes MyoD/E47 association and muscle-specific gene transcription. *EMBO J.*, **24**, 974–984.
67. Fiser, A. and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.*, **374**, 461–491.
68. Benkert, P., Kunzli, M. and Schwede, T. (2009) QMEAN server for protein model quality estimation. *Nucleic Acids Res.*, **37**, W510–W514.
69. Pruunsild, P., Sepp, M., Orav, E., Koppel, I. and Timmusk, T. (2011) Identification of cis-elements and transcription factors regulating neuronal activity-dependent transcription of human BDNF gene. *J. Neurosci.*, **31**, 3295–3308.
70. Okuno, Y., McNairn, A.J., den Elzen, N., Pines, J. and Gilbert, D.M. (2001) Stability, chromatin association and functional activity of mammalian pre-replication complex proteins during the cell cycle. *EMBO J.*, **20**, 4263–4277.
71. Tamberg, N., Lulla, V., Fragkoudis, R., Lulla, A., Fazakerley, J.K. and Merits, A. (2007) Insertion of EGFP into the replicase gene of Semliki Forest virus results in a novel, genetically stable marker virus. *J. Gen. Virol.*, **88**, 1225–1230.
72. 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.

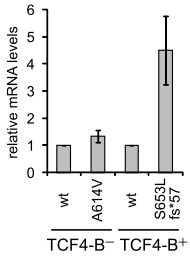


Figure S1. Quantitative RT-PCR analysis of overexpressed *TCF4-B* mRNA levels. HEK293 cells were cotransfected with wt or mutant TCF4-B encoding pCDNA constructs and pGL4.83[hRlucP/EF1α/Puro] vector. Levels of overexpressed *TCF4-B* were detected with primers specific for T7 sequence in pCDNA vector and exon 5 sequence in *TCF4*. For normalization of transfection efficiencies and RNA quantities, levels of *hRLuc* transcripts were determined with intron spanning primers in EF1α promoter and hRluc encoding sequence. Shown are the means from two independent experiments, error bars indicate standard deviations. Recombinant *TCF4-B* mRNA level measured in wt TCF4-B overexpressing cells was arbitrarily set as 1. Non-reverse-transcribed control samples showed negligible amplification with cycle threshold values more than 10 cycles greater than in positive samples.

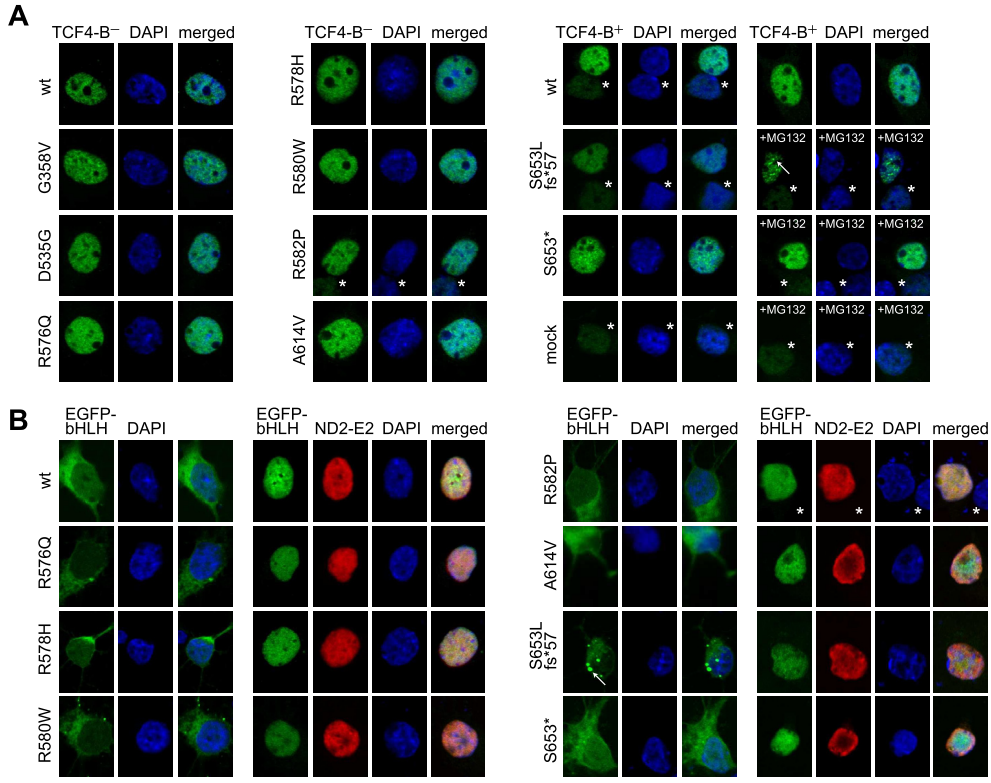


Figure S2. Localization of wt and PTHS-associated mutant TCF4 proteins in rat primary neurons. (**A**) Distribution of wt and mutant TCF4-B proteins. Cells were treated with MG132 for 7 hours where indicated. (**B**) Heterodimerization of TCF4 mutants with NEUROD2. Nuclear redirection assay with wt and mutant EGFP-bHLH proteins overexpressed in primary neurons alone or together with NEUROD2-E2 (ND2-E2). (**A, B**) Immunocytochemical staining was carried out with TCF4 and E2 antibodies. EGFP-fused proteins were visualized by direct fluorescence and nuclei were counterstained with DAPI. Examples of protein aggregates are indicated with arrows and untransfected cells are marked with asterisks. Of note, TCF4 antibodies also detect endogenous protein.

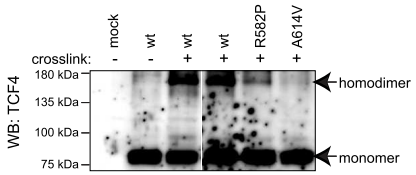


Figure S3. Crosslinking analysis of TCF4 homodimers. *In vitro* translated wt, R582P and A614V TCF4-B$^-$ proteins were diluted 10-fold in 20 mM HEPES (pH 7.6), 50 mM $MgCl_2$, 10% glycerol, 1 mM EDTA, 0.1% NP-40 and 1 mM dithiothreitol, and treated with 0.001% glutaraldehyde for 45 min at room temperature. Proteins were fractionated in 7.5% SDS-PAGE, and TCF4 signals were detected by immunoblotting. Molecular mass markers are shown at the left and the location of TCF4 homodimer and monomer at the right.

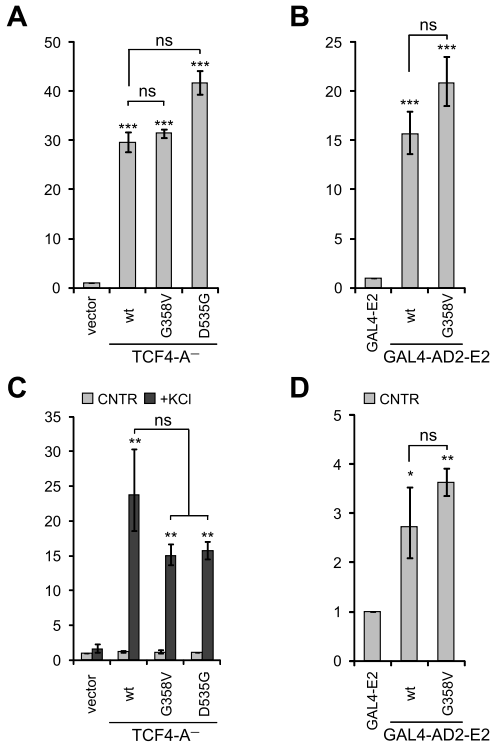


Figure S4. Activation of transcription by G358V and D535G mutants in cultured cells. (**A, B**) Luciferase reporter assays with wt and mutant TCF4-A$^-$ or GAL4-AD2-E2 proteins overexpressed in HEK293 cells. Cells were cotransfected with firefly luciferase construct carrying 12 μE5 E-boxes in front of minimal promoter (A) or 5 GAL4 binding sites in front of adenovirus major late promoter (B) and Renilla luciferase construct with minimal promoter for normalization. (**C, D**) Luciferase reporter assays with wt and mutant TCF4-A$^-$ or GAL4-AD2-E2 proteins overexpressed in primary neurons. Cells were cotransfected with firefly luciferase construct carrying 12 μE5 E-boxes in front of TK promoter (C) or 5 GAL4 binding sites in front of adenovirus major late promoter (D) and Renilla luciferase construct with EF1α promoter for normalization. Neurons were left untreated (CNTR) or treated with KCl for 8 hours. Luciferase activities were measured and data are presented as fold induced levels above the signals obtained from vector transfected cells in A-D. Shown are the mean results from at least three independent experiments performed in duplicates. Error bars indicate standard deviations, statistical significance shown with asterisks is relative to the vector transfected cells or between the bars connected with lines. RLU-relative luciferase unit. *, $p<0.05$; **,$p<0.01$; ***, $p<0.001$; ns, not significant; t-test.

Table S1. Oligonucleotides used for cloning of EF1α promoter, ASCL1 and TCF4 sequences, quantitative polymerase chain reaction (qPCR), site-directed mutagenesis of TCF4 coding sequence, and electrophoretic mobility shift assay (EMSA). Nucleotides in lowercase indicate adapter sequences (CACC Kozak consensus and restiction sites) or mutated positions.

	Specificity	Forward primer (5′-3′)	Reverse primer (5′-3′)
CLONING	ASCL1	ATGGAAAGCTCTGCCAAGATGG	GAAGCAGGGTGATCGCACAA
	TCF4-B$^+$ with 3′UTR	caccATGCATCACCAACAGCGAATGG	ccactcgaGCGTCTGCGATTCATAACTAC
	TCF4-B$^+$ S653*	caccATGCATCACCAACAGCGAATGG	atactcgagttAGAGAGGGGGAGGCTCCGA
	EF1α promoter	aaggatccGTGGAGAAGAGCATGCGTGAG	gtagaagcttGGGGTAGTTTTCACGACACCTG
qPCR	T7+TCF4	TAATACGACTCACTATAGGG	GGACCCTGAGCTACTTCTGTCTTC
	EF1α+hRluc	GGTTTGCCGCCAGAACACAG	CACGTTCATTTGCTTGCAGCGA
MUTAGENESIS	G358V	CCATCTCTCTCAGCAGtCACAGCTGTTTGG	CCAAACAGCTGTGaCTGCTGAGAGAGATGG
	D535G	GGACAAGAAATTAGgTGACGACAAGAAGG	CCTTCTTGTCGTCAcCTAATTTCTTGTCC
	R576Q	CCAACAATGCCCaAGAGCGTCTGCGGGTCC	GGACCCGCAGACGCTCTtGGGCATTGTTGG
	R578H	CAATGCCCGAGAGCaTCTGCGGGTCCGTG	CACGGACCCGCAGAtGCTCTCGGGCATTG
	R580W	GCCCGAGAGCGTCTGtGGGTCCGTGACATC	GATGTCACGGACCCaCAGACGCTCTCGGGC
	R582P	GCGTCTGCGGGTCCcTGACATCAACGAGGC	GCCTCGTTGATGTCAgGGACCCGCAGACGC
	A614V	GATCCTCCACCAGGtGGTGGCCGTCATCC	GGATGACGGCCACCaCCTGGTGGAGGATC
	S653Lfs*57	CTCgGAGCCTCCCccTCTCCTTGGC	GCCAAGGAGAggGGGAGGCTCcGAG
EMSA	µE5	TGCAAGAACACCTGCAAACA	TGTTTGCAGGTGTTCTTGCA
	µE5 mutant	TGCAAGAAttCCcaCAAACA	TGTTTGtgGGaaTTCTTGCA
	I-A1	TCAGGTACATCTGCCGCACC	GGTGCGGCAGATGTACCTGA
	I-A1 mutant	TCAGGTAacTCTtCCGCACC	GGTGCGGaAGAgtTACCTGA

PAPER III

Sepp M, Vihma H, Kannike K, Pruunsild P, Timmusk T
Basic helix-loop-helix transcription factor TCF4 is a calcium-dependent transcriptional activator in neurons.
Manuscript.

CURRICULUM VITAE

General information

Name: Mari Sepp
Date and place of birth: 06.07.1982, Põltsamaa
Address: Kapi 2-5, Tallinn 10136
E-mail: mari.sepp@ttu.ee

Education

2006-2012: Tallinn University of Technology, doctoral studies
2004-2006: Tallinn University of Technology, *M.Sc* in natural sciences
2000-2004: University of Tartu; diploma *cum laude*, *B.Sc* in transgenic technology
1997-2000: Miina Härma Secondary School, Secondary School Leaving Certificate with honours
1988-1997: Puhja Secondary School, Basic School Leaving Certificate

Work Experience

- January 2005 – Tallinn University of Technology, Department of Gene Technology; engineer, researcher
- October 2004 – December 2005: National Institute of Chemical Physics and Biophysics, Department of Molecular Genetics; engineer
- Summers 2003 and 2004: University of Helsinki, Institute of Biotechnology
- October 2001 – May 2004: University of Tartu, Institute of Molecular and Cell Biology

Awards and stipends

2012: The World Congress of Psychiatric Genetics Poster Presentation Award
2009: Tiina Mõis stipend; Development Foundation of Tallinn University of Technology
2005: Artur Lind stipend; Estonian Genome Foundation
2003: Gerhard Treuberg stipend; Rotalia Foundation

Courses and conferences

- October 2012: The World Congress of Psychiatric Genetics; Hamburg, Germany. Presentation "*Pitt-Hopkins syndrome associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant negative effects*"
- June 2011: Gordon Research Conference "Neurotrophic Factors"; Newport, USA. Presentation "*TCF4/ITF2/E2-2 is a calcium-dependent transcriptional activator in neurons*"

- June 2010: NGF meeting "Neurotrophic Factors in Health and Disease"; Helsinki, Finland. Presentation "*Neuronal activity dependent regulation of human BDNF transcription*"
- August 2008: European Molecular Biology Laboratory, Transcription Meeting; Heidelberg, Germany. Presentation "*A vast repertoire of human basic helix-loop-helix transcription factor TCF4 isoforms generated by multiple promoter usage and alternative splicing*"
- November 2007: Society for Neuroscience, Annual Meeting; San Diego, USA. presentation "*Reduced nuclear levels of transcription factor ITF2 lead to BDNF downregulation in cell models of Huntington's disease*"
- June 2007: Graduate School of Biomedicine and Biotechnology, summer school in molecular biology "Chromosomes, chromatin and epigenetics"; Palmse, Estonia
- May 2007: European Molecular Imaging Laboratories, practical course "Introduction to the Use of Bioluminescent and Fluorescent Imaging"; Leiden, Netherlands
- June 2005: International Brain Research Organization, Course in Neuroscience; Tallinn, Estonia
- September 2004: Society for Histochemistry, 46th Symposium "The Cell Nucleus"; Prague, Czech Republic. Presentation „*Subcellular localisation of bovine papilloma virus type 1 E2 proteins*"
- September 2004: Society for Histochemistry, 46th pre-symposium workshop „Fluorescence Detection of Antigens"; Prague, Czech Republic
- May 2002: University of Rostock, students' workshop "The Hanseatic Students' Days of Science"; Rostock, Germany

Theses supervised

- Kati Taal, Master thesis "*Nuclear Transport of Neuralized-1 is Mediated via Interaction with Importin alpha 3 (Kpna3)*". Tallinn University of Technology, Department of Gene Technology, 2011.
- Kaja Kannike, Master thesis "*Increased activity of Foxo3a in Huntington's disease cells*". Tallinn University of Technology, Department of Gene Technology, 2010.
- Ave Eesmaa, Master thesis "*The Role of Transcription Factor ITF2 in the Regulation of Neurotrophin BDNF Gene Expression in Huntington's Disease Cell Models*". Tallinn University of Technology, Department of Gene Technology, 2009.
- Mari Urb, Bachelor thesis "*Silencing of Basic-Helix-Loop-Helix Transcription Factor TCF4 Expression by RNA Interference*". Tallinn University of Technology, Department of Gene Technology, 2011.
- Kaja Kannike, Bachelor thesis "*Analysis of Expression of Transcription Factor FoxO3a Target Genes in Huntington's Disease Model Cell Lines*". Tallinn University of Technology, Department of Gene Technology, 2007.

- Ave Eesmaa, Bachelor thesis "*The Role of Transcription Factor ITF2 in the Regulation of Neurotrophin BDNF Gene Expression*". Tallinn University of Technology, Department of Gene Technology, 2007.
- Jaan Toots, case study "*Study on Nuclear Transport Mechanisms of Neurl1 Protein*". Tallinn University of Technology, Department of Gene Technology and Tallinn Secondary Science School, 2012.

Publications

- **Sepp M**, Pruunsild P, Timmusk T. 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. 2012; 21(13):2873-2888
- **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. 2011;6(7):e22138.
- Pruunsild P, **Sepp M**, Orav E, Koppel I, Timmusk T. Identification of cis-elements and transcription factors regulating neuronal activity-dependent transcription of human BDNF gene. J Neurosci. 2011 Mar 2;31(9):3295-308.
- Koppel I, Aid-Pavlidis T, Jaanson K, **Sepp M**, Palm K, Timmusk T. BAC transgenic mice reveal distal cis-regulatory elements governing BDNF gene expression. Genesis. 2010 Apr;48(4):214-9.
- Kairisalo M, Korhonen L, **Sepp M**, Pruunsild P, Kukkonen JP, Kivinen J, Timmusk T, Blomgren K, Lindholm D. NF-kappaB-dependent regulation of brain-derived neurotrophic factor in hippocampal neurons by X-linked inhibitor of apoptosis protein. Eur J Neurosci. 2009 Sep;30(6):958-66.
- Rullinkov G, Tamme R, Sarapuu A, Laurén J, **Sepp M**, Palm K, Timmusk T. Neuralized-2: expression in human and rodents and interaction with Delta-like ligands. Biochem Biophys Res Commun. 2009 Nov 20;389(3):420-5.
- Koppel I, Aid-Pavlidis T, Jaanson K, **Sepp M**, Pruunsild P, Palm K, Timmusk T. Tissue-specific and neural activity-regulated expression of human BDNF gene in BAC transgenic mice. BMC Neurosci. 2009 Jun 25;10:68.
- Kazantseva A*, **Sepp M***, Kazantseva J, Sadam H, Pruunsild P, Timmusk T, Neuman T, Palm K. N-terminally truncated BAF57 isoforms contribute to the diversity of SWI/SNF complexes in neurons. J Neurochem. 2009 Feb 20.
- Kurg R, Uusen P, Sepp T, **Sepp M**, Abroi A, Ustav M. Bovine papillomavirus type 1 E2 protein heterodimer is functional in papillomavirus DNA replication in vivo. Virology. 2009 Feb 19.
- Timmusk T, Koppel I, Pruunsild P, **Sepp M**, Tamme R. Neurotroofsed tegurid. Eesti Arst. 2007; 86:614 - 621.
- Kurg R, Sild K, Ilves A, **Sepp M**, Ustav M. Association of bovine papillomavirus E2 protein with nuclear structures in vivo. J Virol. 2005 Aug;79(16):10528-39.

* equal contribution

ELULOOKIRJELDUS

Üldine info

Nimi: Mari Sepp
Sünniaeg ja -koht: 06.07.1982, Põltsamaa
Aadress: Kapi 2-5, Tallinn 10136
E-post: mari.sepp@ttu.ee

Haridustee

2006-2012: Tallinna Tehnikaülikool, doktoriõpe geenitehnoloogia erialal
2004-2006: Tallinna Tehnikaülikool, loodusteaduste magistri kraad (*M.Sc*)
2000-2004: Tartu Ülikool; diplom *cum laude*, *B.Sc* transgeense tehnoloogia erialal
1997-2000: Miina Härma Gümnaasium, gümnaasiumi lõputunnistus, kuldmedal
1988-1997: Puhja Gümnaasium, põhikooli lõputunnistus

Erialane töökogemus

- Jaanuar 2005 – Tallinna Tehnikaülikool, Geenitehnoloogia instituut; insener, teadur;
- Oktoober 2004 – detsember 2005: Keemilise ja Bioloogilise Füüsika instituut; insener;
- Suvi 2003 ja 2004: Helsingi Ülikool, Biotehnoloogia instituut
- Oktoober 2001 – mai 2004: Tartu Ülikool, Molekulaar- ja Rakubioloogia instituut.

Auhinnad ja stipendiumid

2012: Poster-ettekande auhind Psühhiaatrilise Geneetika Kongressil
2009: Tiina Mõisa stipendium; Sihtasutus Tallinna Tehnikaülikooli Arengufond
2005: Artur Linnu stipendium; Eesti Geenikeskus
2003: Gerhard Treubergi stipendium; Rotalia Fond

Kursused ja konverentsid

- Oktoober 2012: Psühhiaatrilise Geneetika Kongress; Hamburg, Saksamaa. Ettekanne „*Pitt-Hopkins syndrome associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant negative effects*"
- Juuni 2011: Gordon Teaduskonverents „Neurotroofsed tegurid"; Newport, USA. Ettekanne „*TCF4/ITF2/E2-2 is a calcium-dependent transcriptional activator in neurons*"

- Juuni 2010: NGF konverents „Neurotroofsed tegurid normis ja haiguses"; Helsingi, Soome. Ettekanne „*Neuronal activity dependent regulation of human BDNF transcription*"
- August 2008: EMBL (*European Molecular Biology Laboratory*) Transkriptsiooni konverents; Heidelberg, Saksamaa. Ettekanne „*A vast repertoire of human basic helix-loop-helix transcription factor TCF4 isoforms generated by multiple promoter usage and alternative splicing*"
- November 2007: Neuroteaduste ühingu (*Society for Neuroscience*) aastakonverents; San Diego, USA. Ettekanne „*Reduced nuclear levels of transcription factor ITF2 lead to BDNF downregulation in cell models of Huntington's disease*"
- Juuni 2007: Biomeditsiini ja Biotehnoloogia doktorikooli teoreetiline kursus „Kromatiin ja epigeneetika"; Eesti, Palmse
- Mai 2007: EMIL (*European Molecular Imaging Laboratories*) parktiline kursus „Bioluminestsentsi ja fluorestsentsi kuvamine"; Holland, Leiden
- Juuni 2005: IBRO (*International Brain Research Organization*) Neuroteaduste kursus; Tallinn, Eesti
- September 2004: Histokeemia Ühingu 46. Sümpoosium „Rakutuum"; Praha, Tsehhi. Ettekanne „*Subcellular localisation of bovine papilloma virus type 1 E2 proteins*"
- September 2004: Histokeemia Ühingu 46. Sümpoosiumi töötuba „Antigeenide detekteerimine kudedes"; Praha, Tšehhi
- Mai 2002: Hansa Tudengite Teaduspäevad, teoreetiline ja praktiline kursus molekulaar- ja rakubioloogia meetoditest; Rostock, Saksamaa

Juhendatud diplomitööd

- Kati Taal, magistritöö „Neuralized-1 tuumatransporti vahendab interaktsioon Importiiniga alfa 3 (Kpna3)". Tallinna Tehnikaülikool, Geenitehnoloogia instituut, 2011.
- Kaja Kannike, magistritöö „Foxo3a suurenenud aktiivsus Huntingtoni tõve rakkudes" Tallinna Tehnikaülikool, Geenitehnoloogia instituut, 2010.
- Ave Eesmaa, magistritöö „Transkriptsioonifaktori ITF2 roll neurotrofiini BDNF geeniekspressiooni regulatsioonis Huntingtoni tõve rakumudelites" Tallinna Tehnikaülikool, Geenitehnoloogia instituut, 2009.
- Mari Urb, bakalaureusetöö „Aluselise heeliks-ling-heeliks transkriptsioonifaktori TCF4 ekspressiooni vaigistamine RNA interferentsi meetodil". Tallinna Tehnikaülikool, Geenitehnoloogia instituut, 2011.
- Kaja Kannike, bakalaureusetöö „Transkriptsioonifaktori FoxO3a sihtmärkgeenide ekspressiooni analüüs Huntingtoni tõve mudelrakuliinides". Tallinna Tehnikaülikool, Geenitehnoloogia instituut, 2007.
- Ave Eesmaa, bakalaureusetöö „Transkriptsioonifaktori ITF2 roll neurotrofiini BDNF geeniekspressiooni regulatsioonis". Tallinna Tehnikaülikool, Geenitehnoloogia instituut, 2007.

- Jaan Toots, uurimistöö „Neurl1 valgu tuumatranspordi mehhanismide uurimine“ Tallinna Tehnikaülikool, Geenitehnoloogia instituut ja Tallinna Reaalkool, 2012.

Publikatsioonid

- **Sepp M**, Pruunsild P, Timmusk T. 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. 2012;21(13):2873-2888
- **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. 2011;6(7):e22138.
- Pruunsild P, **Sepp M**, Orav E, Koppel I, Timmusk T. Identification of cis-elements and transcription factors regulating neuronal activity-dependent transcription of human BDNF gene. J Neurosci. 2011 Mar 2;31(9):3295-308.
- Koppel I, Aid-Pavlidis T, Jaanson K, **Sepp M**, Palm K, Timmusk T. BAC transgenic mice reveal distal cis-regulatory elements governing BDNF gene expression. Genesis. 2010 Apr;48(4):214-9.
- Kairisalo M, Korhonen L, **Sepp M**, Pruunsild P, Kukkonen JP, Kivinen J, Timmusk T, Blomgren K, Lindholm D. NF-kappaB-dependent regulation of brain-derived neurotrophic factor in hippocampal neurons by X-linked inhibitor of apoptosis protein. Eur J Neurosci. 2009 Sep;30(6):958-66.
- Rullinkov G, Tamme R, Sarapuu A, Laurén J, **Sepp M**, Palm K, Timmusk T. Neuralized-2: expression in human and rodents and interaction with Delta-like ligands. Biochem Biophys Res Commun. 2009 Nov 20;389(3):420-5.
- Koppel I, Aid-Pavlidis T, Jaanson K, **Sepp M**, Pruunsild P, Palm K, Timmusk T. Tissue-specific and neural activity-regulated expression of human BDNF gene in BAC transgenic mice. BMC Neurosci. 2009 Jun 25;10:68.
- Kazantseva A*, **Sepp M***, Kazantseva J, Sadam H, Pruunsild P, Timmusk T, Neuman T, Palm K. N-terminally truncated BAF57 isoforms contribute to the diversity of SWI/SNF complexes in neurons. J Neurochem. 2009 Feb 20.
- Kurg R, Uusen P, Sepp T, **Sepp M**, Abroi A, Ustav M. Bovine papillomavirus type 1 E2 protein heterodimer is functional in papillomavirus DNA replication in vivo. Virology. 2009 Feb 19.
- Timmusk T, Koppel I, Pruunsild P, **Sepp M**, Tamme R. Neurotroofsed tegurid. Eesti Arst. 2007; 86:614 - 621.
- Kurg R, Sild K, Ilves A, **Sepp M**, Ustav M. Association of bovine papillomavirus E2 protein with nuclear structures in vivo. J Virol. 2005 Aug;79(16):10528-39.

* võrdsed autorid