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

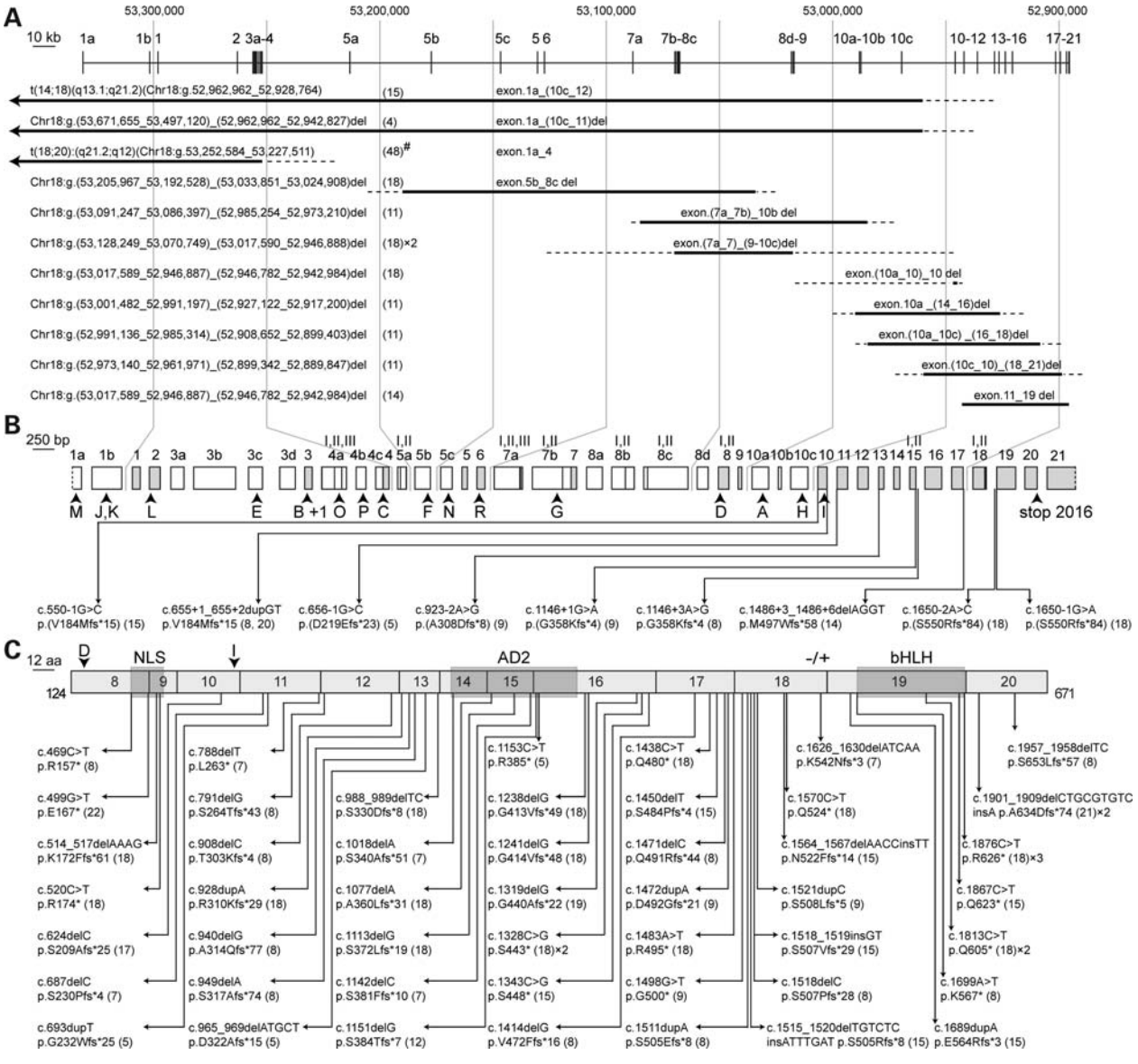

Figure 1. Schematic depiction of Pitt–Hopkins disease-causing mutations. (**A**) The published deletions and translocations that encompass part of *TCF4* gene. The minimum deleted or translocated regions are depicted as solid lines and the maximum regions as dotted lines in relation to the *TCF4* genomic 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

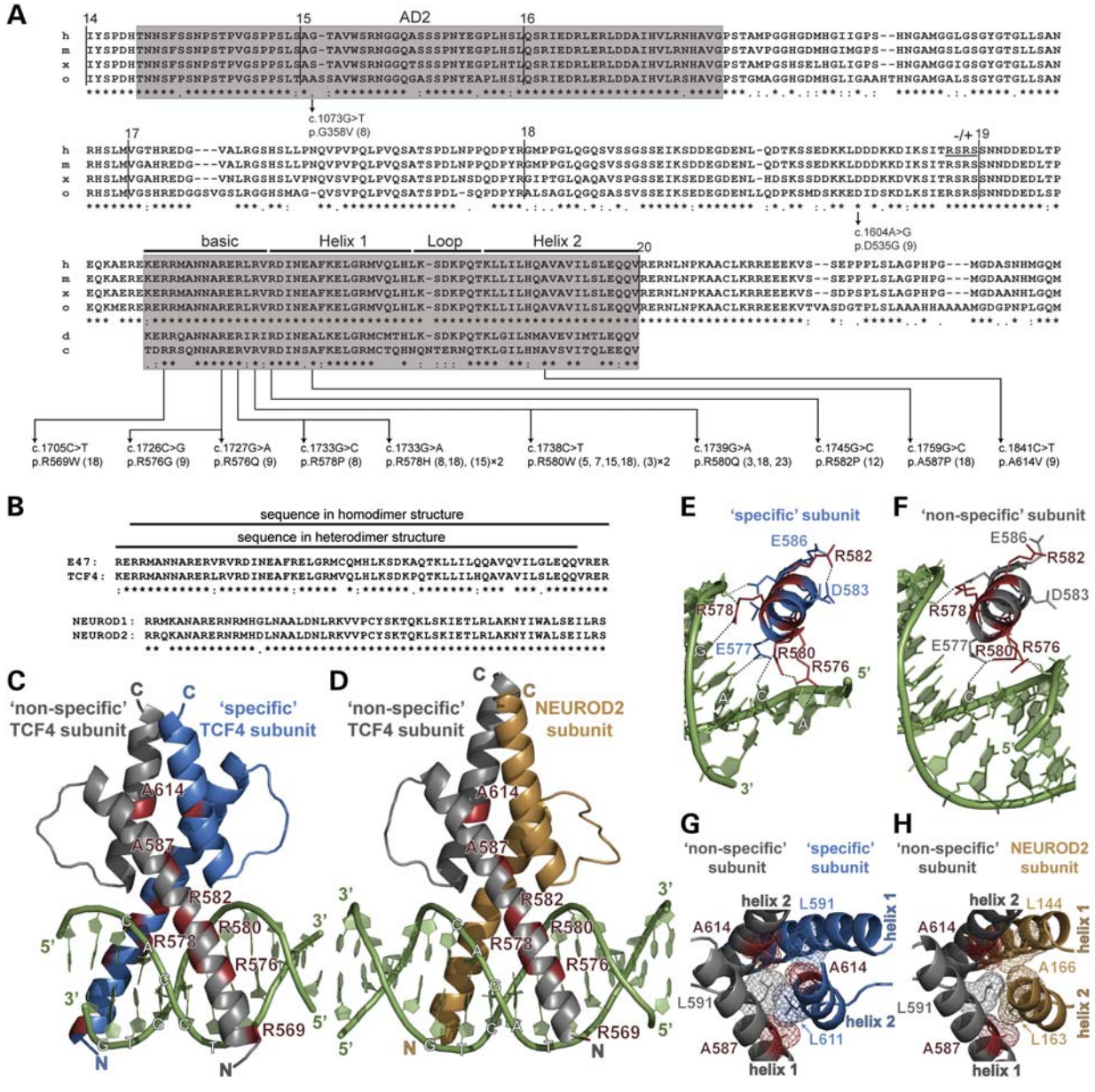

Figure 2. Mapping of Pitt–Hopkins disease-causing missense mutations. (**A**) The published *TCF4* missense mutations depicted in relation to the alignment of exons 14–20 encoded TCF4 protein sequences from *Homo sapiens* (h), *Mus musculus* (m), *Xenopus laevis* (x) and *Oryzias latipes* (o). The alignment 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

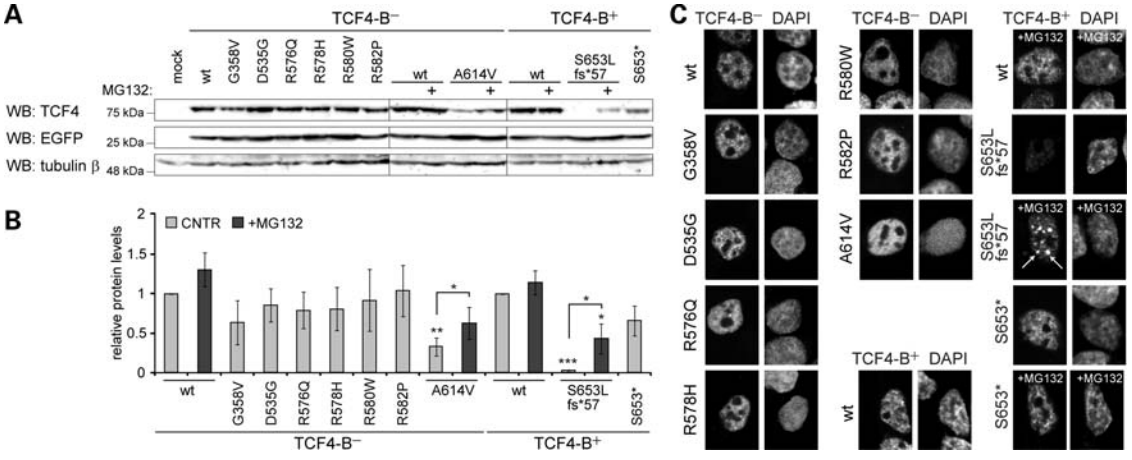

Figure 3. Expression and localization of TCF4-B mutants in cultured cells. (**A**) Western blot analysis of extracts from HEK293 cells co-transfected with pEGFP vector and wt or mutant TCF4-B-encoding constructs. Cells were treated with MG132 for 7 h where indicated. Signals of TCF4, EGFP (transfection 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

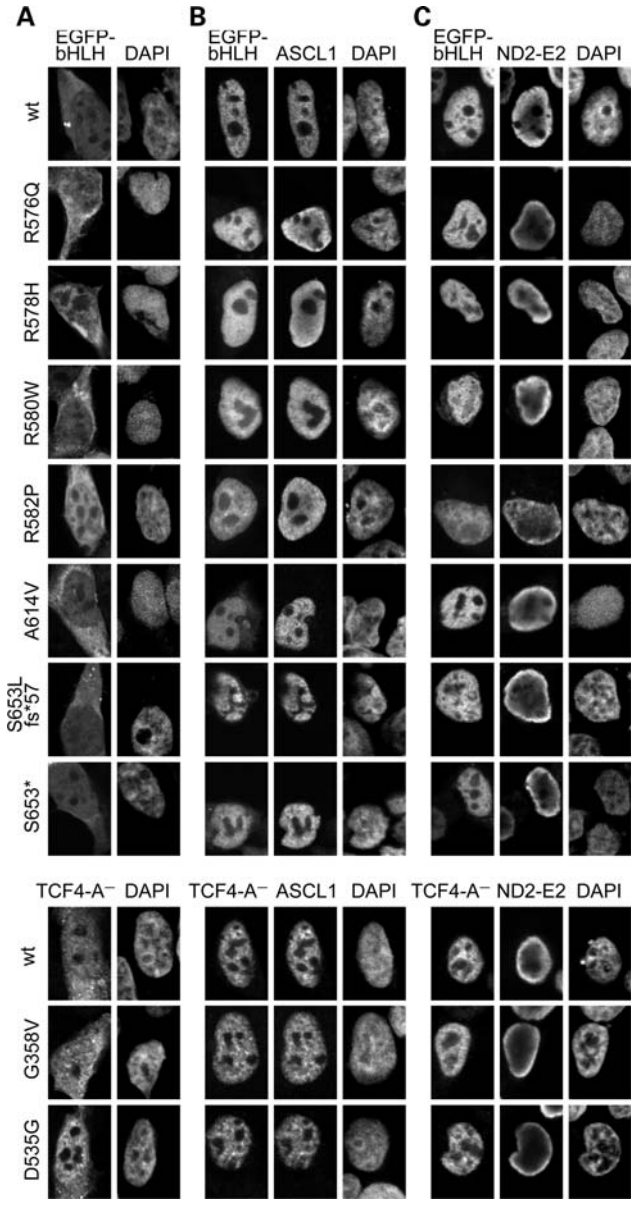

Figure 4. Heterodimerization of TCF4 mutants with NEUROD2 and ASCL1 in cultured cells. Nuclear redirection assay with wt and mutant EGFP-bHLH or TCF4-A proteins overexpressed in HEK293 cells (**A**) alone, (**B**) together with ASCL1 or (**C**) together with NEUROD2-E2 (ND2-E2). Immunocytochemical staining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

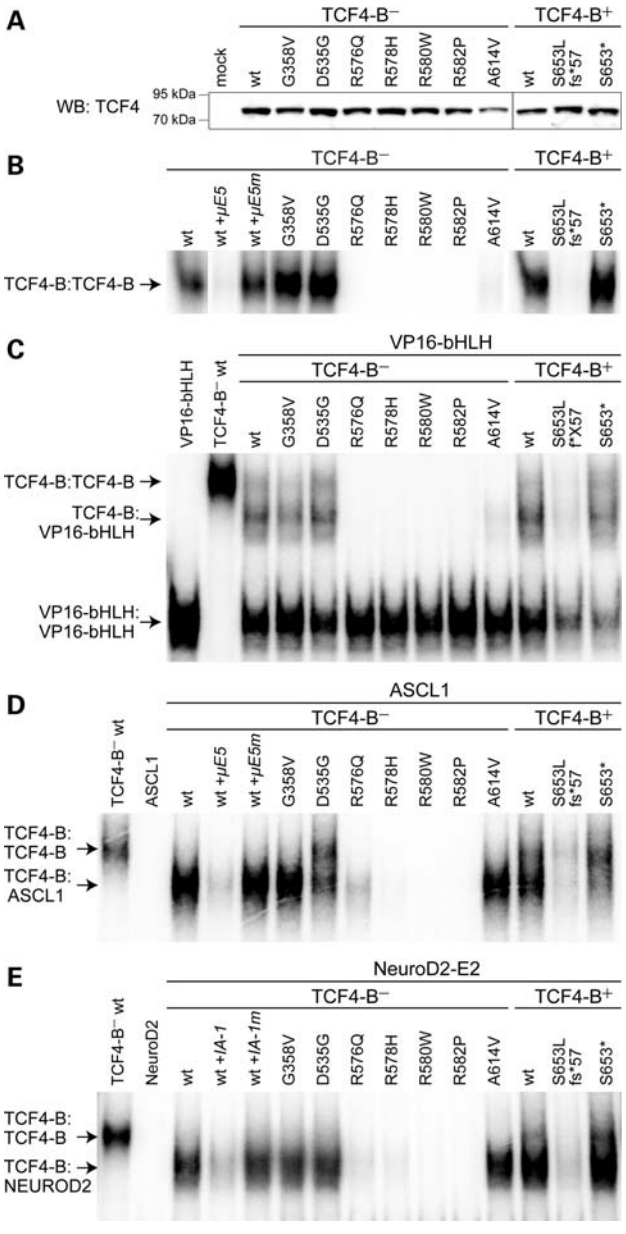

Figure 5. Ability of TCF4 mutants to bind DNA *in vitro*. (**A**) Western blot analysis of *in vitro*-translated wt and mutant TCF4-B proteins. Molecular mass markers are shown at the left. (**B**–**D**) Electrophoretic mobility shift analysis of *in vitro*-translated wt and mutant TCF4-B proteins' ability 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

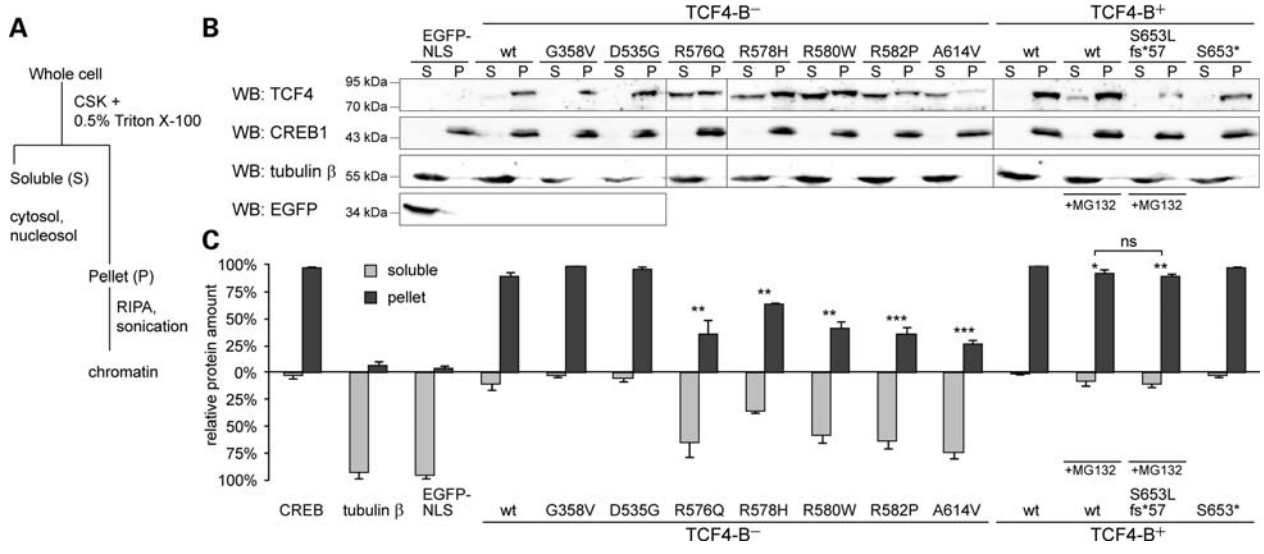

Figure 6. Co-fractionation of TCF4 mutants with cellular chromatin. (**A**) Schematic of cell extraction protocol. (**B**) Western blot analysis of nucleosol- and cytosol-containing soluble (S) and chromatin-containing pellet (P) fractions from HEK293 cells overexpressing EGFP-NLS, wt or mutant TCF4-B 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

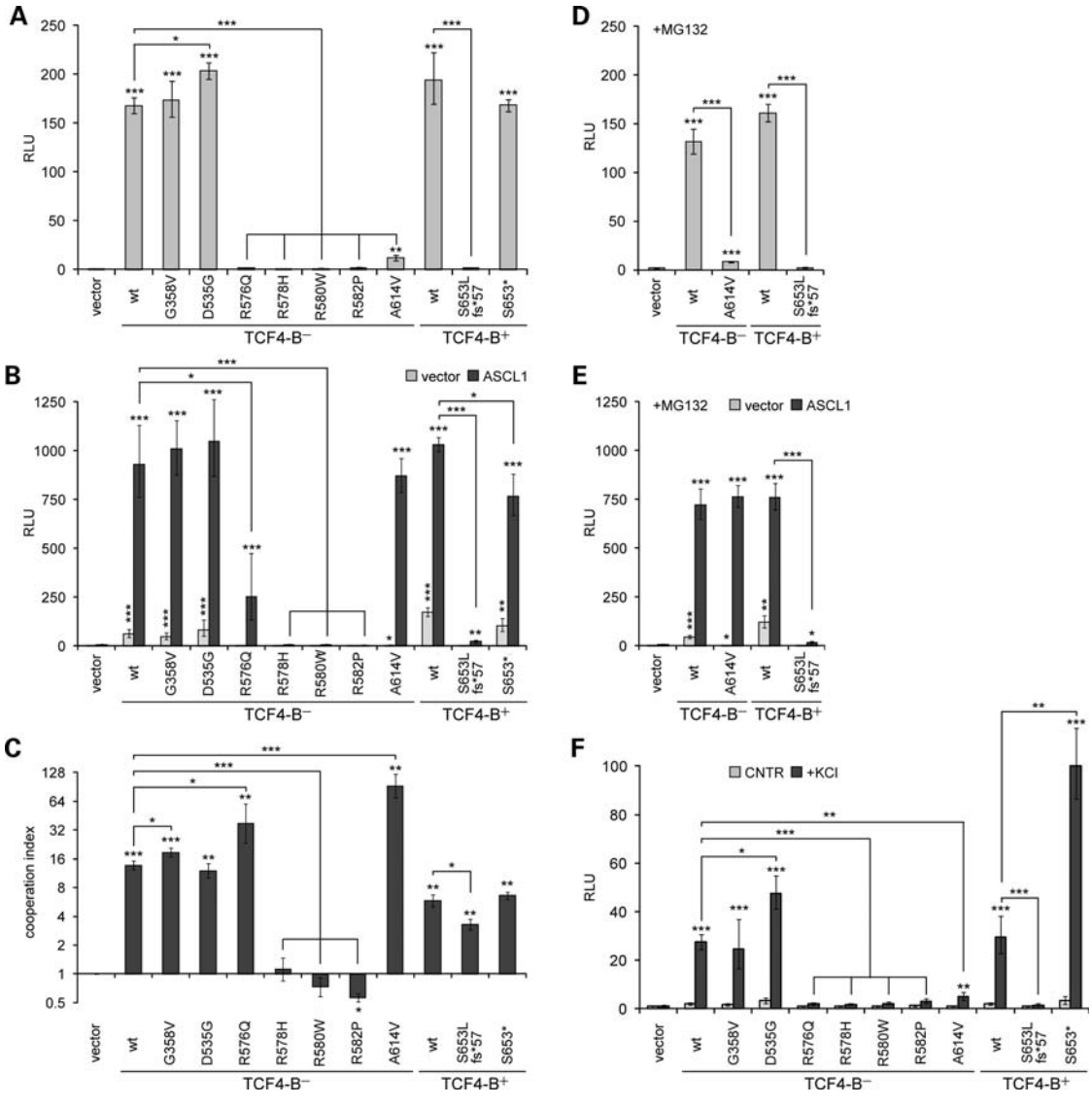

Figure 7. Activation of E-box-controlled transcription by TCF4 mutants in cultured cells. (**A**, **B**, **D** and **E**) Luciferase reporter assays with wt and mutant TCF4-B proteins overexpressed in HEK293 cells alone (A and D) or together with ASCL1 (B and E). Cells were co-transfected with firefly 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.