



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

TMEM132: an ancient architecture of cohesin and immunoglobulin domains define a new family of neural adhesion molecules

Citation for published version:

Sanchez-Pulido, L & Ponting, C 2018, 'TMEM132: an ancient architecture of cohesin and immunoglobulin domains define a new family of neural adhesion molecules' *Bioinformatics*. DOI: 10.1093/bioinformatics/btx689

Digital Object Identifier (DOI):

[10.1093/bioinformatics/btx689](https://doi.org/10.1093/bioinformatics/btx689)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Bioinformatics

Publisher Rights Statement:

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Sequence Analysis

TMEM132: an ancient architecture of cohesin and immunoglobulin domains define a new family of neural adhesion molecules.

Luis Sanchez-Pulido* and Chris P. Ponting

Medical Research Council Human Genetics Unit, IGMM,
University of Edinburgh, EH4 2XU, UK.

*To whom correspondence should be addressed.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: The molecular functions of TMEM132 genes remain poorly understood and under-investigated despite their mutations associated with non-syndromic hearing loss, panic disorder and cancer. Here we show the full domain architecture of human TMEM132 family proteins solved using in-depth sequence and structural analysis. We reveal them to be five previously unappreciated cell adhesion molecules whose domain architecture has an early holozoan origin prior to the emergence of choanoflagellates and metazoa. The extra-cellular portions of TMEM132 proteins contain five conserved domains including three tandem immunoglobulin domains, and a cohesin domain homologue, the first such domain found in animals. These findings strongly predict a cellular adhesion function for TMEM132 family, connecting the extracellular medium with the intracellular actin cytoskeleton.

Contact: Luis.Sanchez-Pulido@igmm.ed.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Many genes remain experimentally under-investigated not because they are functionally less important but because their discovery came relatively late (Pandey *et al.*, 2014). Our ignorance of aspects of basic biology and disease thus is perpetuated by the serendipitous order by which genes were first characterised. The need to experimentally determine proteins' normal molecular functions, and their molecular dysfunction in disease, becomes more critical when sequence variants within functionally enigmatic genes are robustly associated with Mendelian or complex disease, or with cancer progression. Determining the molecular functions of such poorly-characterised genes is all the more difficult when their protein sequences lack recognisable domains, because these otherwise can reliably provide structural and functional information through homology-based inference. Here, we shed much light on the previously-unknown domain structure

and functions of the 5 proteins of the human TMEM132 family (TMEM132A, B, C, D and E).

These are genes in which variants are enriched for individuals with hearing loss, panic disorder or cancer. A homozygous missense mutation in human *TMEM132E* (Arg420Gln) was confirmed using a zebrafish model to cause autosomal-recessive nonsyndromic hearing loss (Li *et al.*, 2015). Common variants within the *TMEM132E* gene are associated with insomnia symptoms (Lane *et al.*, 2017); common and rare variants near *TMEM132D* gene are robustly associated with panic disorder (Erhardt *et al.*, 2011; Erhardt *et al.*, 2012; Quast *et al.*, 2012; Inoue *et al.*, 2015; Howe *et al.*, 2016; Shimada-Sugimoto *et al.*, 2016; Wang *et al.*, 2016; Hodgson *et al.*, 2016); and variants near TMEM132B are associated with excessive daytime sleepiness (Lane *et al.*, 2017). In healthy individuals, some of the *TMEM132D* non-coding variants exhibit higher anxiety scores and larger volumetric estimates of

the amygdala and hippocampus, key neural structures associated with fear and anxiety (Haaker *et al.*, 2014). Furthermore, in cattle the TMEM132D locus appears to have undergone a selective sweep during domestication (Qanbari *et al.*, 2014), and in the mouse, anterior cingulate cortex TMEM132D expression correlates with anxiety-related behaviour (Erhardt *et al.*, 2011). Finally, mutations in *TMEM132D* are unusually frequent in small-cell lung cancer (Peifer *et al.*, 2012; Rudin *et al.*, 2012; Iwakawa *et al.*, 2015) and in pancreatic cancer (Forbes *et al.*, 2015). Disease mutations near, or functions of, *TMEM132A* or *C* have yet to be identified, although *TMEM132A* is thought to promote neuronal cell survival by regulating stress-related genes (Oh-hashii *et al.*, 2003; Oh-hashii *et al.*, 2006; Oh-hashii *et al.*, 2010; Oh-hashii *et al.*, 2012; Oh-hashii *et al.*, 2015). Functions of the single fruit fly (CG14446) or nematode (Y71H2AM.10) orthologous genes are yet to be described.

All five human paralogues encode cell-surface molecules expressed in the brain (Nomoto *et al.*, 2003; Oh-hashii, 2010; Walser *et al.*, 2011; Uhlen *et al.*, 2010) whose amino acid sequences contain no distinguishable features other than their N-terminal signal peptides and C-terminal proximal transmembrane sequences (Nomoto *et al.*, 2003; Oh-hashii *et al.*, 2010; Oh-hashii *et al.*, 2012), with intracellular C-terminal Ser/Thr phosphatase-1 (PP1) docking (Hendrickx *et al.*, 2009; Heroes *et al.*, 2012) and WIRS (WAVE regulatory complex interacting receptor sequence) cytoplasmic motifs (Figure 1 and Supplemental Figure S1) (Chen *et al.*, 2014). The latter is consistent with the reported co-localization of TMEM132D with actin filaments (Walser *et al.*, 2011).

Cell-cell junctions in the central nervous system (CNS) are maintained by a variety of transmembrane proteins that signal and physically link between the cytoskeletons of adjacent cells. Many such proteins (e.g. CNTN, LRIG, NCAM, SEMA3 and SIGLEC families) contain one or more immunoglobulin (IG) superfamily domains (Rougon and Hobert, 2003). Here we identify TMEM132 molecules as novel IG domain containing proteins of the CNS.

2 Results and discussion

2.1 Protein sequence analysis: three tandem immunoglobulin domains.

We initiated our analyses by performing a JackHMMER iterative search (Finn *et al.*, 2015) starting from the human TMEM132A protein sequence of the UniRef50 database (Wu *et al.*, 2006). Whilst characterising the TMEM132 family we identified full-length homologous proteins across essentially all of the animal kingdom, including nematodes (*Caenorhabditis elegans*) and hexapods (*Drosophila melanogaster*). As input for our analysis we used a full-length multiple sequence alignment generated with T-Coffee (Notredame *et al.*, 2000). Using extensive profile-to-sequence and profile-to-profile comparison analyses (Finn *et al.* 2015; Söding *et al.* 2005) we identified a repeated pattern of conserved amino-acids in the region lying between positions 400 and 767 for human TMEM132A, corresponding approximately to the conserved region used to define the

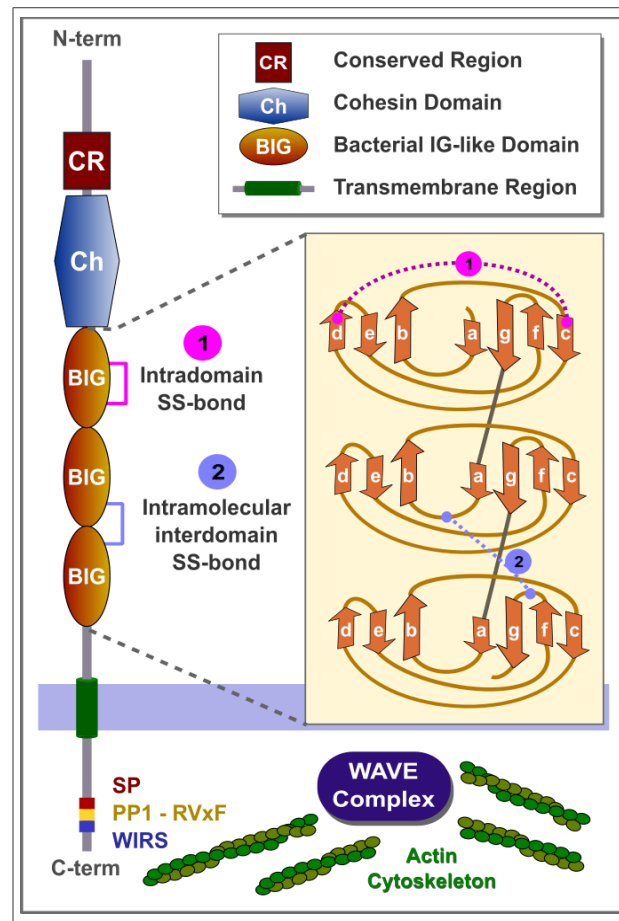


Fig. 1. Domain architecture and common features in the TMEM132 family. Schematic representation of conserved domains present in TMEM132 family members: a conserved region (CR; shown in red; see Supplemental Figure S6) precedes a cohesin domain (blue; see Supplemental Figure S5), and three adjacent BIG domains (orange; Supplemental Figures S3 and S4). The predicted intradomain and interdomain disulphide bridges of these BIG domains are indicated in the right panel (magenta and violet lines, respectively). The seven beta-strands, forming part of the immunoglobulin-like core of BIG domains, are labelled a-g following an established convention (Bork *et al.*, 1994) (Supplemental Figures S7). Evolutionarily conserved TMEM132 intracellular motifs putatively related with the control of actin cytoskeletal dynamics are: a putative serine phosphorylation motif (SP), a phosphatase-1 (PP1) interaction motif (RVxF) (Hendrickx *et al.*, 2009; Heroes *et al.*, 2012), and a WIRS (WAVE regulatory complex interacting receptor sequence) cytoplasmic motif (Chen *et al.*, 2014).

family in Pfam (Family TMEM132, accession: PF16070) (Punta *et al.*, 2012). This region is conserved among animals and some premetazoan proteins that are additionally rich in cadherin domains (Supplemental Figure S2) (Abedin and King, 2008; Nichols *et al.*, 2012). Profile-versus-sequence and profile-versus-profile comparisons of this conserved repeated pattern allowed the identification of three consecutive repeated regions, each of which independently yielded statistically significant E-values of sequence similarity with the same fold, the bacterial immunoglobulin-

like (BIG) domain (Mei *et al.*, 2015; Ptak *et al.*, 2014). HHpred searches against the PDB70 profile database. (Söding *et al.* 2005) using TMEM132 repeats 1, 2 and 3 as input (corresponding to amino acids 400-491, 495-630, and 641-767 of human TMEM132A) detected the BIG domain from *Leptospira interrogans* (PDB ID: 2mh4), for example, with E-values of 0.031, 0.018, and 0.022, respectively. Moreover in all three HHpred results, in support of the first match, the next most statistically significant matches corresponded to additional members of the immunoglobulin superfamily (Supplemental Figures S3 and S4). The PDB70 database contains profile hidden Markov models (HMMs) for representative sequences, clustered to 70% maximum pairwise sequence identity to reduce redundancy, drawn from the PDB database (Söding *et al.* 2005).

BIG domains are widely distributed among bacteria, archaea, and eukaryotes (Pfam family Big_2, accession: PF02368) (Punta *et al.*, 2012). This domain adopts a beta-sandwich fold composed of nine strands organised in three sheets. Two of these sheets (composed of seven strands) contribute the immunoglobulin-like core of BIG domains. These seven strands are labelled “a” to “g” in figure 1 and Supplemental Figure S3, following an established convention in the immunoglobulin fold (Bork *et al.*, 1994). BIG domains have been described with diverse functions, usually relating to matrix, protein-ligand, or protein-protein interactions and are mainly extracellular (Mei *et al.*, 2015; Ptak *et al.*, 2014).

Close 3D proximity and evolutionary conservation of four cysteines allow us to identify two putative disulphide bridges, one that is internal to BIG1 and another that is inter-domain between BIG2 and BIG3 domains (Figure 1; Supplemental Figures S3, S4, and S7). Disulphide bridges are commonly found in different IG folds and contribute to their structural stability (Bork *et al.*, 1994).

2.2 First animal Cohesin domain.

Identification of the three tandem BIG domains, then allowed our analyses to be focused on the TMEM132 family N-terminal region taking advantage of iterative profile-versus-sequence searches against the UniRef50 protein sequence database (Wu *et al.*, 2006). These resulted in the identification of two additional domains, each of which is present not only in animals but also among more diverse eukaryotes, including members of the Coherin family in choanoflagellates and sponges (Nichols *et al.*, 2012). The domain preceding the BIG domains was discovered as the first cohesin homology domain in vertebrates (HHpred E-value < 5×10^{-3}) (Supplemental Figure S5). Cohesin domains are found widely in prokaryotes but, in eukaryotes, were previously thought to be restricted to choanoflagellate and sponge proteins (Pfam accession: PF00963) (Peer *et al.*, 2009; Abedin and King, 2008; Nichols *et al.*, 2012). These are not to be confused with the cohesin complex that regulates the separation of sister chromatids. Rather, cohesin domains are highly specialised protein-protein interaction modules that bind dockerin domains together forming the core that glues together the Cellulosome complex, a multi-enzymatic complex present in cellulolytic bacteria specialised in degrading cellulose (Bras *et al.*, 2016; Artzi *et al.*, 2017; Pinheiro *et al.*, 2008; Tavares *et al.*, 1997;

Adams *et al.*, 2008). Bacterial cohesin-dockerin rupture forces (>120 pN) are among the highest ever reported for a receptor-ligand system (Nash *et al.*, 2016; Stahl *et al.*, 2012). It is unclear whether the TMEM132 cohesin domain mediates such a strong interaction, in part because dockerin domain homologues are not detectable in vertebrate proteins.

2.3 TMEM132 domain architecture is ancient

The conserved region (corresponding to amino acids 127-239 of human TMEM132A) preceding the cohesin domain in TMEM132 is also evident in choanoflagellate and sponge proteins (HMMER E-value < 0.005) (Supplemental Figure S6). Strikingly, despite each of the five TMEM132 domains (Figure 1) being identified independently in these choanoflagellate and sponge proteins, all five are both present and in the identical order in the three cadherin protein families - lefttyrins, coherins, and hedglings - that were contained in the last common ancestor of choanoflagellates and metazoans (Abedin and King, 2008; Nichols *et al.*, 2012) (Supplemental Figure S2). The TMEM132 domain architecture is thus ancient, preceding the emergence of early metazoans, and a repeated constituent of ancient cadherin domain-containing proteins with roles connecting the actin cytoskeleton with neighbouring cells and the extracellular matrix (Briehner and Yap, 2013; Ratheesh and Yap, 2012).

2.4 Disease and biological relevance

Eleven proteins are currently known to contain missense mutations within IG domains associated with 23 different disorders (Letunic *et al.*, 2015). To these now can be added a twelfth, TMEM132E, whose R420Q missense mutation, mapped to its second BIG domain (Supplemental Figure S1), has been validated using a zebrafish model to cause autosomal-recessive nonsyndromic hearing loss (Li *et al.*, 2015). A TMEM132B nonsynonymous variant that replaces a serine conserved in TMEM132B-E in a putative phosphorylation motif has been associated with intra-cranial aneurysm (Farlow *et al.*, 2015) although this variant also occurs rarely (frequency 1×10^{-4}) in the general population (Lek *et al.*, 2016). This lies adjacent to the Ser/Thr PP1 docking (Hendrickx *et al.*, 2009; Heroes *et al.*, 2012) and WIRS cytoplasmic motifs (Figure 1), the latter which is found in a variety of neurological and other proteins including protocadherins (Chen *et al.*, 2014). The newly identified domains, and conserved domain architecture, of TMEM132 proteins now should facilitate detailed experimental investigation of these proteins' domain and molecular functions and how these are modulated by sequence variants.

3 Conclusion

Their ancient ancestry and their associations with neurological disease suggest that TMEM132 genes have been underserving of their relative obscurity. Our identification of these proteins as CNS-expressed IG domain superfamily adhesion molecules now places them in a more appropriate perspective as a putative key connection between the extracellular matrix and the actin-based cell cytoskeleton, with major roles in reg-

ulating changes in neuronal cell morphology, motility, and migration. These findings should precipitate more detailed experimental and structural characterization of the TMEM132 family, and assist in formulating hypotheses concerning the cellular mechanisms by which sequence variants in these genes contribute to neurological disease.

Funding

This work has been supported by the Medical Research Council UK.

Conflict of Interest: none declared.

References

Abedin,M. and King,N. (2008) The premetazoan ancestry of cadherins. *Science*, **319**, 946-948.

Adams,J.J. *et al.* (2008) Structural basis of toxin complex formation. *Proc. Natl. Acad. Sci. USA*, **105**, 12194-12199.

Artzi,L. *et al.* (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.*, **15**, 83-95.

Bork,P. *et al.* (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.*, **242**, 309-320.

Briehner,W.M. and Yap,A.S. (2013) Cadherin junctions and their cytoskeleton(s). *Curr. Opin. Cell Biol.*, **25**, 39-46.

Bras,J.L. *et al.* (2016) Diverse specificity of cellulosome attachment to the bacterial cell surface. *Sci. Rep.*, **6**, 38292.

Chen,B. *et al.* (2014) The WAVE regulatory complex links diverse receptors to the actin cytoskeleton. *Cell*, **156**, 195-207.

Erhardt,A. *et al.* (2012) Replication and meta-analysis of TMEM132D gene variants in panic disorder. *Transl. Psychiatry*, **2**, e156.

Erhardt,A. *et al.* (2011) TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. *Mol. Psychiatry*, **16**, 647-63.

Farlow,J.L. *et al.* (2015) Lessons learned from whole exome sequencing in multiplex families affected by a complex genetic disorder, intracranial aneurysm. *PLoS One*, **10**, e0121104.

Finn,R.D. *et al.* (2015) HMMER web server: 2015 update. *Nucleic Acids Res.*, **43**, W30-W38.

Forbes,S.A. *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805-D811.

Haaker,J. *et al.* (2014) Higher anxiety and larger amygdala volumes in carriers of a TMEM132D risk variant for panic disorder. *Transl. Psychiatry*, **4**, e357.

Hendrickx,A. *et al.* (2009) Docking motif-guided mapping of the interactome of protein phosphatase-1. *Chem. Biol.*, **16**, 365-371.

Heroes,E. *et al.* (2013) The PP1 binding code: a molecular-lego strategy that governs specificity. *FEBS J.*, **280**, 584-595.

Hodgson,K. *et al.* (2016) Genome-wide significant loci for addiction and anxiety. *Eur. Psychiatry*, **36**, 47-54.

Howe,A.S. *et al.* (2016) Candidate genes in panic disorder: meta-analyses of 23 common variants in major anxiogenic pathways. *Mol. Psychiatry*, **21**, 665-679.

Inoue,A. *et al.* (2015) Association of TMEM132D, COMT, and GABRA6 genotypes with cingulate, frontal cortex and hippocampal emotional processing in panic and major depressive disorder. *Int. J. Psychiatry Clin. Pract.*, **19**, 192-200.

Iwakawa,R. *et al.* (2015) Expression and clinical significance of genes frequently mutated in small cell lung cancers defined by whole exome/RNA sequencing. *Carcinogenesis*, **36**, 616-621.

Lane,J.M. *et al.* (2017) Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.*, **49**, 274-281.

Lek,M. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285-291.

Letunic,I. *et al.* (2015) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.*, **43**, D257-D260.

Li,J. *et al.* (2015) Whole-exome Sequencing Identifies a Variant in TMEM132E Causing Autosomal-Recessive Nonsyndromic Hearing Loss DFNB99. *Hum. Mutat.*, **36**, 98-105.

Mei,S. *et al.* (2015) Solution structure of a bacterial immunoglobulin-like domain of the outer membrane protein (LigB) from *Leptospira*. *Proteins*, **83**, 195-200.

Nash,M.A. *et al.* (2016) Single versus dual-binding conformations in cellulosomal cohesin-dockerin complexes. *Curr. Opin. Struct. Biol.*, **40**, 89-96.

Nichols,S.A. *et al.* (2012) Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/beta-catenin complex. *Proc. Natl. Acad. Sci. USA*, **109**, 13046-13051.

Nomoto,H. *et al.* (2003) Molecular cloning of a novel transmembrane protein MOLT expressed by mature oligodendrocytes. *J. Biochem.*, **134**, 231-238.

Notredame,C. *et al.* (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205-217.

Oh-hashikawa,K. *et al.* (2006) GRP78-binding protein regulates cAMP-induced glial fibrillary acidic protein expression in rat C6 glioblastoma cells. *FEBS Lett.*, **580**, 3943-3947.

Oh-hashikawa,K. *et al.* (2010) Knockdown of transmembrane protein 132A by RNA interference facilitates serum starvation-induced cell death in Neuro2a cells. *Mol. Cell Biochem.*, **342**, 117-123.

Oh-hashikawa,K. *et al.* (2012) Characterization of the expression and cell-surface localization of transmembrane protein 132A. *Mol. Cell Biochem.*, **370**, 23-33.

Oh-hashikawa,K. *et al.* (2003) Cloning and characterization of a novel GRP78-binding protein in the rat brain. *J. Biol. Chem.*, **278**, 10531-10537.

Oh-Hashi,K. *et al.* (2015) Transcriptional and post-transcriptional regulation of transmembrane protein 132A. *Mol. Cell Biochem.*, **405**, 291-299.

Pandey,A.K. *et al.* (2014) Functionally enigmatic genes: a case study of the brain ignorome. *PLoS One*, **9**, e88889.

Peer,A. *et al.* (2009), Smith SP, Bayer EA, Lamed R, Borovok I. Non-cellulosomal cohesin- and dockerin-like modules in the three domains of life. *FEMS Microbiol. Lett.*, **291**, 1-16.

Peifer,M. *et al.* (2012) Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat. Genet.*, **44**, 1104-1110.

Pinheiro,B.A. *et al.* (2008) The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner. *J. Biol. Chem.*, **283**, 18422-18430.

Ptak,C.P. *et al.* (2014) NMR solution structure of the terminal immunoglobulin-like domain from the *leptospira* host-interacting outer membrane protein, LigB. *Biochemistry*, **53**, 5249-5260.

Punta,M. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res.*, **40**, 290-301.

Qanbari,S. *et al.* (2014) Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.*, **10**, e1004148.

Quast,C. *et al.* (2012) Rare variants in TMEM132D in a case-control sample for panic disorder. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.*, **159B**, 896-907.

Ratheesh,A. and Yap,A.S. (2012) A bigger picture: classical cadherins and the dynamic actin cytoskeleton. *Nat. Rev. Mol. Cell Biol.*, **13**, 673-679.

Rougon,G. and Hobert,O. (2003) New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. *Annu. Rev. Neurosci.*, **26**, 207-238.

Rudin,C.M. *et al.* (2012) Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat. Genet.*, **44**, 1111-1116.

Söding,J. *et al.* (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244-W248.

Shimada-Sugimoto,M. *et al.* (2016) Polymorphisms in the TMEM132D region are associated with panic disorder in HLA-DRB1*13:02-negative individuals of a Japanese population. *Hum. Genome Var.*, **3**, 16001.

Stahl,S.W. *et al.* (2012) Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl. Acad. Sci. USA*, **109**, 20431-20436.

TMEM132: a new family of neural adhesion molecules

- Tavares,G.A. *et al.* (1997) The crystal structure of a type I cohesin domain at 1.7 Å resolution. *J. Mol. Biol.*, **273**, 701-13.
- Uhlen,M. *et al.* (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.*, **28**, 1248-1250.
- Walser,S. *et al.* (2011) TMEM132D - a putative cell adhesion molecule involved in panic disorder. *Pharmacopsychiatry*, **21**, A117.
- Wang,M.H. *et al.* (2016) A fast and powerful W-test for pairwise epistasis testing. *Nucleic Acids Res.*, **44**, e115.
- Wu,C.H. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, **34**, D187-D191.