TM6SF2 and MAC30, new enzyme homologs in sterol metabolism and common metabolic disease

Luis Sanchez-Pulido * and Chris P. Ponting *

Medical Research Council Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK

Carriers of the Glu167Lys coding variant in the TM6SF2 gene have recently been identified as being more susceptible to non-alcoholic fatty liver disease (NAFLD), yet exhibit lower levels of circulating lipids and hence are protected against cardiovascular disease. Despite the physiological importance of these observations, the molecular function of TM6SF2 remains unknown, and no sequence similarity with functionally characterized proteins has been identified. In order to trace its evolutionary history and to identify functional domains, we embarked on a computational protein sequence analysis of TM6SF2. We identified a new domain, the EXPERA domain, which is conserved among TM6SF, MAC30/TMEM97 and EBP (D8, D7 sterol isomerase) protein families. EBP mutations are the cause of chondrodysplasia punctata 2 X-linked dominant (CDPX2), also known as Conradi-Hünermann-Happle syndrome, a defective cholesterol biosynthesis disorder. Our analysis of evolutionary conservation among EXPERA domain-containing families and the previously suggested catalytic mechanism for the EBP enzyme, indicate that TM6SF and MAC30/TMEM97 families are both highly likely to possess, as for the EBP family, catalytic activity as sterol isomerases. This unexpected prediction of enzymatic functions for TM6SF and MAC30/TMEM97 is important because it now permits detailed experiments to investigate the function of these key proteins in various human pathologies, from cardiovascular disease to cancer.

Keywords: TM6SF2, MAC30, TMEM97, EBP, NAFLD, CDPX2, cancer, cholesterol

INTRODUCTION

Exome resequencing studies have shown great success in identifying variants that cause rare Mendelian disease (Bamshad et al., 2011). More recently, exome association studies have begun to reveal coding variants that contribute to complex disease risk (Do et al., 2012; Kiezun et al., 2012). To fully understand disease pathoetiology the identification of these variants should be followed by experimental studies that seek to reveal their effects on biochemical pathways and cellular processes. This can be straightforward when much is already known about the mutated gene and its encoded protein. Nevertheless, disease-associated coding variants often lie within sequence or genes that are devoid of annotated functions or features, such as currently defined motifs or domains (Gollery et al., 2006), or they occur within human genes whose proteins have not yet been experimentally characterized. Most human proteins currently have no well-defined molecular function (Lee et al., 2007). Even when proteins contain features such as catalytic amino acids that are indicative of enzymatic activity their substrates may remain unknown (Bartlett et al., 2003; Galperin and Koonin, 2004; Addou et al., 2009). A productive approach to determining the molecular functions of

newly-assigned disease genes is to identify, using in-depth protein sequence analyses, homology relationships that reveal evolutionary relationships and domain architectures and, on occasion, explain the molecular and cellular deficits in disease (Goodstadt and Ponting, 2001).

Recently, a coding variant (p.Glu167Lys) in a human gene TM6SF2 (Transmembrane 6 Superfamily Member 2) was found to exceed genome-wide significance for association with total cholesterol and liver fat levels (Dongiovanni et al., 2014; Holmen et al., 2014; Kozlitina et al., 2014; Liu et al., 2014; Sookoian et al., 2014). This amino acid substitution also explains the genome-wide association study signals on chromosome 19p12 for plasma triglyceride or total cholesterol levels, and for increased myocardial infarction risk and non-alcoholic fatty liver disease (NAFLD) susceptibility (Dongiovanni et al., 2014; Holmen et al., 2014; Kozlitina et al., 2014; Mahdessian et al., 2014; Sookoian et al., 2014). These studies showed that TM6SF2 is expressed highly in the liver, with the Glu167Lys variant being expressed at greatly reduced levels, and that alteration of its transcript's levels in the mouse results in changes in liver triglyceride, cholesterol, low-density and high-density lipoprotein levels and in very-low-density lipoprotein (VLDL) secretion.

Nevertheless, the molecular function of TM6SF2 remains unknown, and its only distinguishing features are its predicted 10 transmembrane helices, and its localisation to the endoplasmic

Abbreviations: TM6SF2, Transmembrane 6 Superfamily Member 2; TM6SF1, Transmembrane 6 Superfamily Member 1; MAC30, Meningioma-Associated protein 30; TMEM97, Transmembrane Protein 97; EBP, Emopamil Binding Protein; NAFLD, Non-alcoholic Fatty Liver Disease; CDPX2, Chondro-Dysplasia Punctata 2, X-linked dominant; MEND, Male EBP Disorder with Neurological Defects.

reticulum (ER) and the ER-Golgi intermediate compartment in liver cells (Mahdessian et al., 2014). In particular, it is currently unknown whether this protein has value of being a potential drug target (Holmen et al., 2014), and how its variant contributes to liver triglyceride metabolism, coronary artery disease and type 2 diabetes mellitus (Mahdessian et al., 2014). In-depth analysis of protein sequences, so successful previously in explaining the molecular bases of disease (for example, Sanchez-Pulido et al., 2012; Zhang et al., 2012; Babbs et al., 2013), may shed light on its function.

Consequently, we embarked on a computational sequence analysis of the TM6SF protein family and identified a novel domain (termed "EXPERA," see below) that is present twice in TM6SF proteins and that is also conserved among MAC30 (Meningioma-associated protein 30; also known as Transmembrane protein 97 [TMEM97]), and EBP (Emopamil binding protein) protein families (**Figures 1–4**). These observations provide evidence that these previously uncharacterised human proteins have, in common with EBP, an isomerase enzymatic activity contributing to sterol metabolism.

MATERIAL AND METHODS

Multiple sequence alignments for each EXPERA domain-containing families were generated independently with the program T-Coffee using default parameters (Notredame et al., 2000), slightly refined manually and visualized with the Belvu program (Sonnhammer and Hollich, 2005). Profiles of the alignment as global hidden Markov models (HMMs) were generated using HMMer (Eddy, 1996; Finn et al., 2011). Profile-based sequence searches were performed against the Uniref50 protein sequence database (Wu et al., 2006) using HMMsearch (Eddy, 1996; Finn et al., 2011). We used NAIL to view and analyse the HMMsearch results, which provided a formatted view with hyper-links to related web resources and coloring related to taxonomic information, thus facilitating interpretation of the results (Sánchez-Pulido et al., 2000). Remote homology analyses were performed using
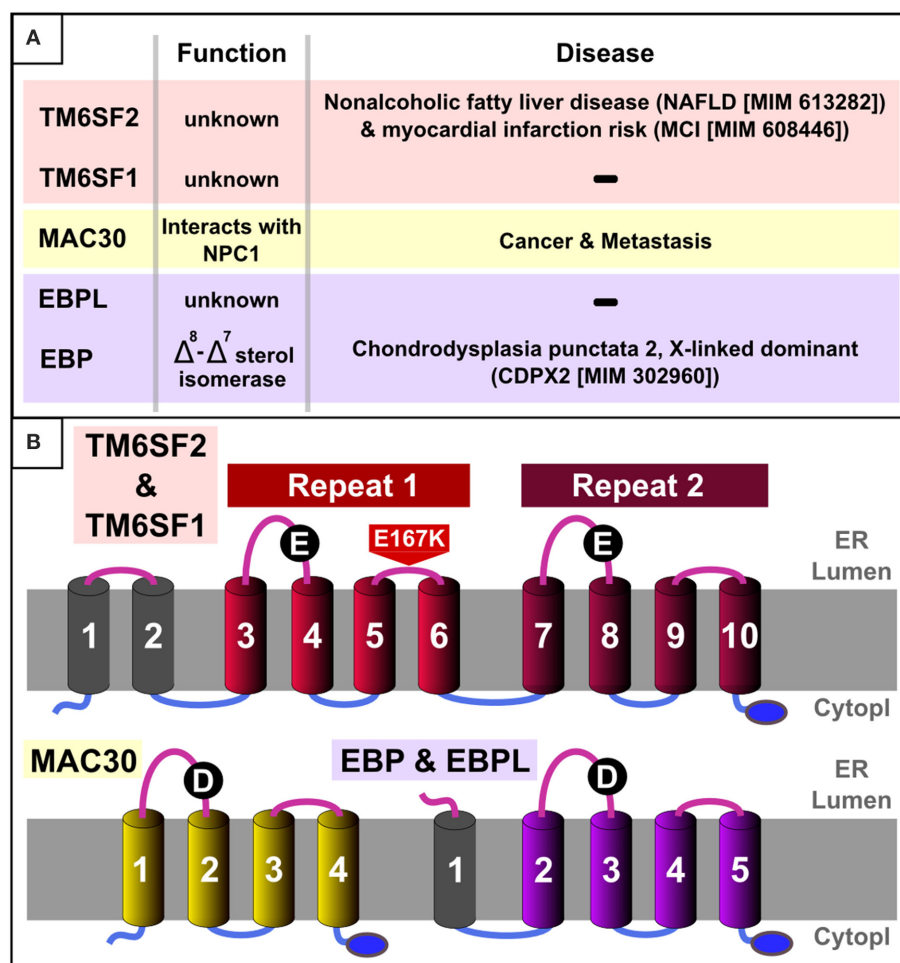

FIGURE 1 | EXPERA domain-containing proteins in humans. (A) Summary table of known function/disease relationships for all EXPERA domain-containing proteins in humans. **(B)** Transmembrane topology and common features in human members of the EXPERA domain-containing families. Common topology for the core four transmembrane regions (colored cylinders) present in all members of the EXPERA superfamily (**Figures 3**, **S1–S3**). The blue oval labels the putative C-terminal ER retention signal (lysine-rich sequence). Black circles label the most conserved position of the EXPERA superfamily (E101 and E255 in TM6SF2, D56 in MAC30 and D108 in EBP).

profile-to-profile comparisons (Söding et al., 2005). The significance of sequence-to-sequence, profile-to-sequence, and profile-to-profile matches were evaluated in terms of an E-value, which reflects the number of observations of better sequence matches expected by chance. Transmembrane predictions were performed using the TMHMM Server (Krogh et al., 2001). Figures were generated using Inkscape (http://inkscape.org/).

RESULTS AND DISCUSSION
SEQUENCE ANALYSIS
We started by considering the number of transmembrane regions present in TM6SF protein family members. A few members of a second family, that of MAC30/TMEM97 proteins, are

annotated by Pfam as being homologous to TM6SF proteins (Pfam entry: DUF2781—Domain of Unknown Function 2781) (Bateman et al., 2010; Punta et al., 2012). This was puzzling because ten transmembrane regions are consistently predicted for TM6SF proteins (**Figures 1, S1**) whereas only four such helices are evident for MAC30/TMEM97 proteins (**Figures 1, S2**). As transmembrane proteins often contain internal duplications (Shimizu et al., 2004), we considered whether TM6SF proteins contain tandem repeats of multiple transmembrane regions. Indeed, using the HHpred profile-profile comparison approach (Söding et al., 2005), and a sequence profile of the last four transmembrane regions of TM6SF (corresponding to human TM6SF2 amino acids 217–351), we identified statistically significant sequence
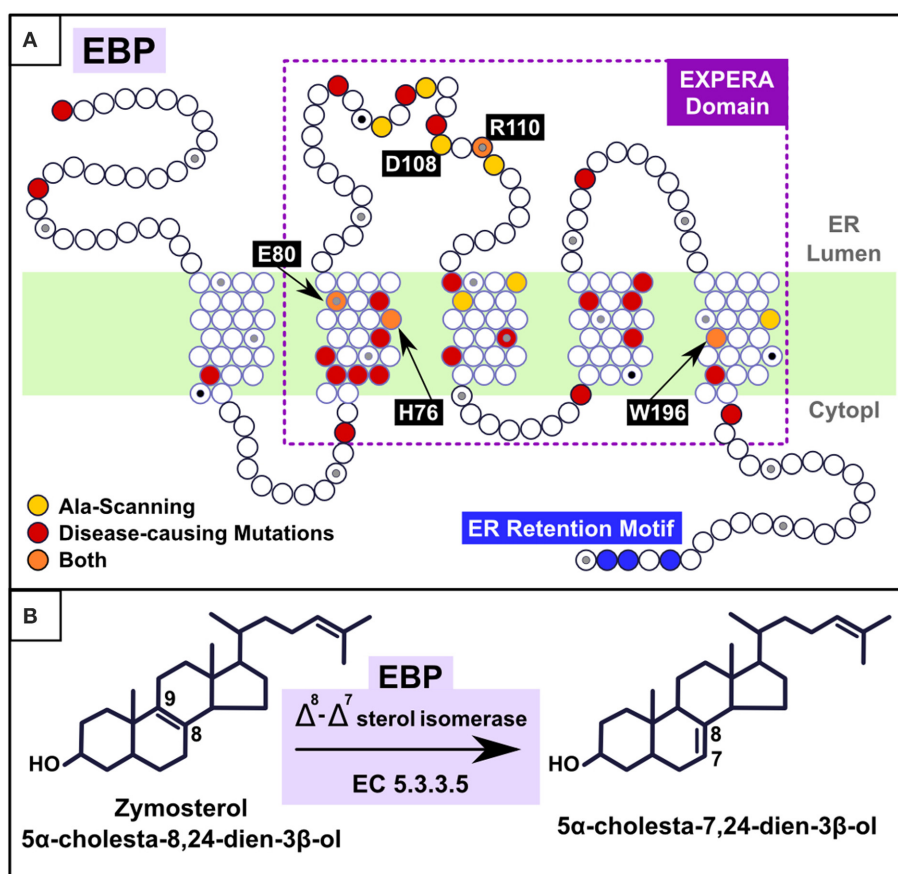

FIGURE 2 | (A) Mapping alanine-scanning mutagenesis and known disease causing missense mutations in the EBP model. Alanine-scanning (Moebius et al., 1999) identified 11 residues as major determinants of EBP catalytic activity (His77, Glu81, Trp102, Tyr105, Asp109, Arg111, Tyr112, Glu123, Thr126, Asn194, and Trp197; here renumbered to the current EBP_HUMAN SwissProt entry numbering by subtracting one to the number of each position). Four (orange points) are present in exactly the same positions as are disease associated mutations (H76Y, E80K, R110Q, and W196S) and the remaining seven (W101, Y104, D108, Y111, E122, T125, and N193) (yellow points) are located in the vicinity of disease associated mutations (fewer than five residues-distant). Mapped CDPX2 disease causing missense mutations (red points), derived from Human Gene Mutation Database (HGMD) and PubMed analysis (Stenson et al., 2003), are: M1I (Steijlen et al., 2007), M1V (Hello et al., 2010), R62W (Herman et al., 2002), L66P (Whittock et al., 2003), C67R (Morice-Picard et al., 2011), W68C (Lambrecht et al., 2014), C72Y

(Herman et al., 2002), I75N (Barboza-Cerda et al., 2014), H76Y (Umekoji et al., 2008), E80K (Braverman et al., 1999; Ikegawa et al., 2000; Aughton et al., 2003), W82C (Has et al., 2002; Shirahama et al., 2003), S98F (Tysoe et al., 2008), S98P (Tysoe et al., 2008), E103K (Kolb-Mäurer et al., 2008), G107R (Derry et al., 1999), R110Q (Derry et al., 1999; Hou, 2013), V119G (Non-lethal) (Cañueto et al., 2012; Bode et al., 2013), G130V (Herman et al., 2002), S133R (Braverman et al., 1999; Derry et al., 1999), R147G (Becker et al., 2001), R147H (Braverman et al., 1999; Has et al., 2000; Ikegawa et al., 2000; Shirahama et al., 2003), G157S (Herman et al., 2002), D162H (Whittock et al., 2003), L164P (Cañueto et al., 2012), Y165C (Shirahama et al., 2003), G173R (Herman et al., 2002), W196S (Herman et al., 2002), L203P (Has et al., 2002), D206Y (Ausavarat et al., 2008). L18P and W47C (Milunsky et al., 2003; Furtado et al., 2010) present a less severe phenotype called MEND (Male EBP Disorder with Neurological Defects) syndrome (Arnold et al., 2012). **(B)** Reaction catalyzed by EBP. Cholesterol carbon atoms C7, C8, and C9 are label.

similarity with a profile generated from TM6SF transmembrane regions three to six (corresponding to human TM6SF2 amino acids 61–186; $E = 0.03$). In addition, this approach revealed significant sequence similarity between each of these repeats and a single repeat in the MAC30/TMEM97 family (corresponding to human MAC30/TMEM97 amino acids 10–157; $E = 6 \times 10^{-7}$ and 0.03; **Figure 4**).

By iteratively improving the phyletic coverage in each protein family using HMMer database searches (Eddy, 1996), we obtained statistical significance from profile-profile comparisons that link

these three sequence families (specifically, the two TM6SF repeats and the single MAC30/TMEM97 repeat) to the Emopamil binding protein (EBP) family (**Figures 3**, **4**). The significance of these sequence similarities, their common transmembrane helix configuration, and their shared predicted C-terminal ER retention signal (**Figures 1**, **2**) (Jackson et al., 1990) imply that these domains are homologous, having derived from a common evolutionary ancestor. We name this four transmembrane region the EXPERA (EXPanded EBP superfamily) domain.
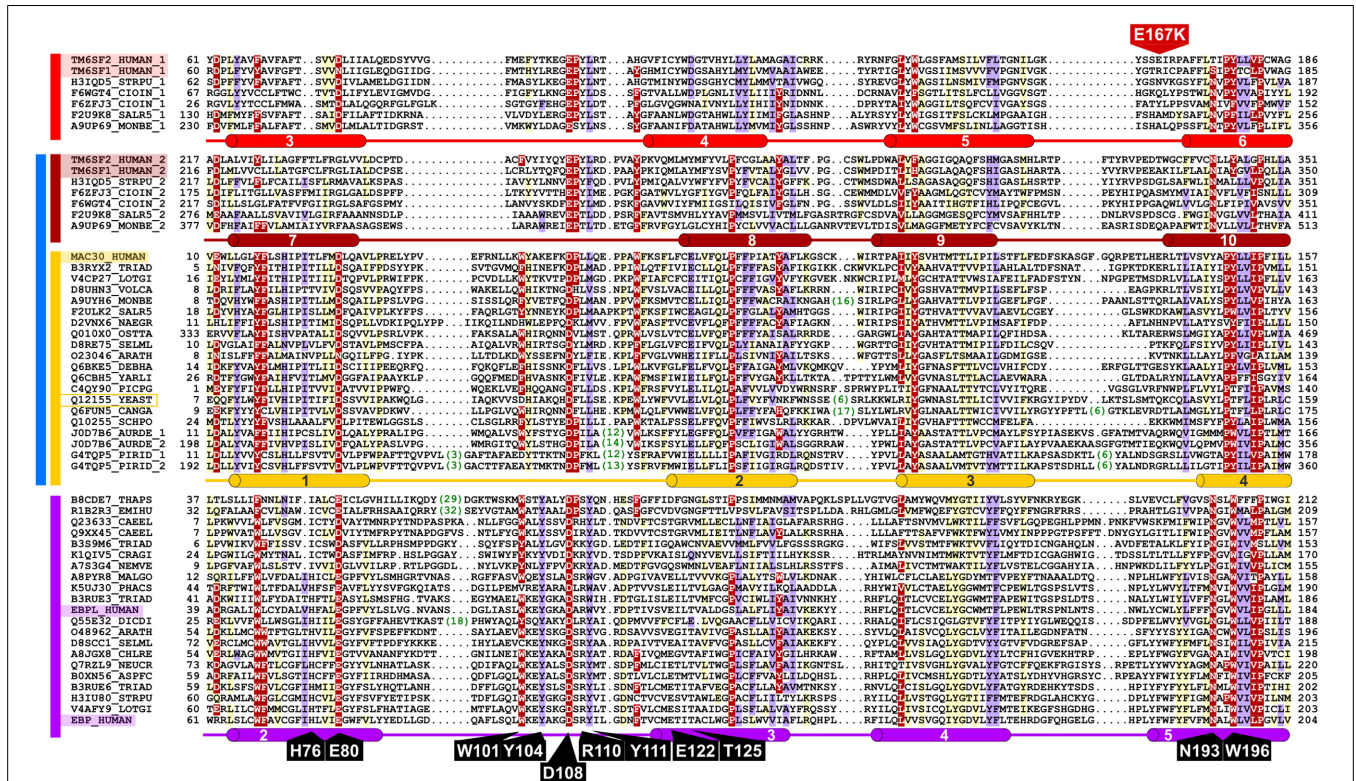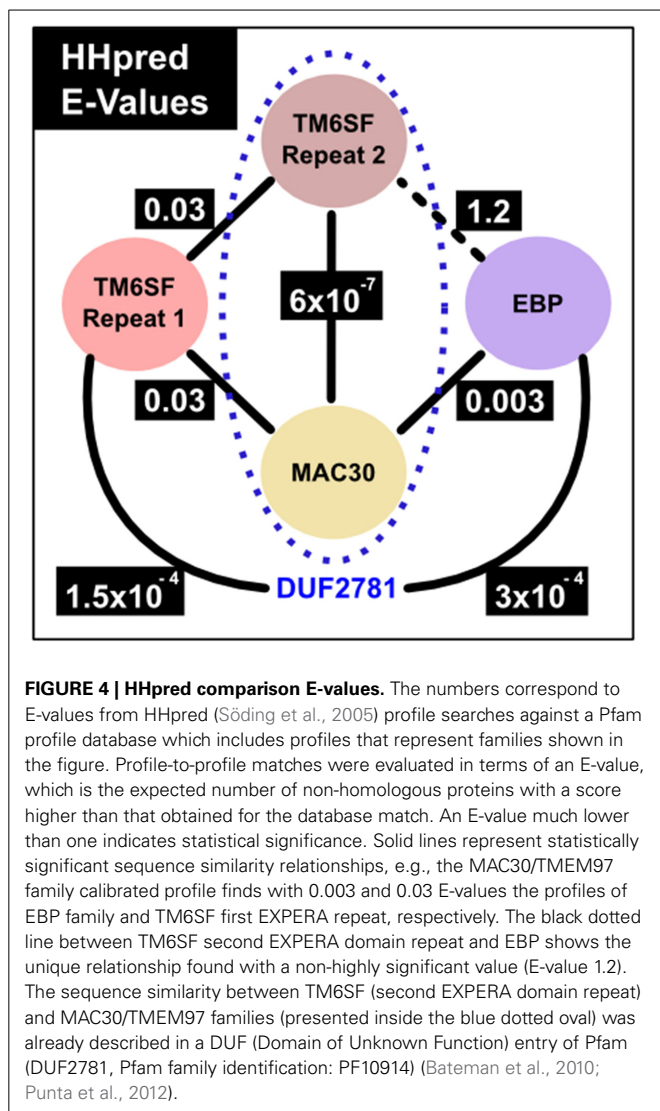

FIGURE 3 | Representative multiple sequence alignment of the EXPERA domain. Putative EBP catalytic residues (identified by alanine-scanning) described by Moebius et al. are label in black (Moebius et al., 1999). A mutation identified in TM6SF2 is label in red (Holmen et al., 2014; Kozlitina et al., 2014; Sookoian et al., 2014). Human sequence names are highlighted and the only member of the EXPERA superfamily in *Saccharomyces cerevisiae*, part of the MAC30/TMEM97 family, is indicated by a yellow box. Numbers shown in green represent inserted amino acids that have been removed from the alignment. Different groups of the EXPERA sequences identified by sequence similarity are shown by colored lines to the left of the alignment: light red, TM6SF family first repeat; dark red, TM6SF family second repeat; yellow, MAC30/TMEM97 family; purple, EBP family. DUF2781 (in blue), previously defined in Pfam (includes TM6SF second repeat and MAC30 family). The TMHMM helix transmembrane (Krogh et al., 2001) consensus prediction are shown below the alignment for each family, in red, yellow, and violet cylinders for TM6SF (repeats 1 and 2), MAC30/TMEM97, and EBP families, respectively (see **Figures S1–S3**). The limits of the protein sequences included in the alignment are indicated by flanking residue positions. Alignments were produced with T-Coffee, HMMer, and HHpred (Eddy, 1996; Notredame et al., 2000; Söding et al., 2005; Finn et al., 2011) using default parameters and slightly refined manually. The alignment was presented with the program Belvu (Sonnhammer and Hollich, 2005) using a coloring scheme

indicating the average BLOSUM62 scores (which are correlated with amino acid conservation) of each alignment column: red (>0.7), violet (between 0.7 and 0.4) and light yellow (between 0.4 and 0.2). Sequences are named according to their UniProt identifications (Wu et al., 2006). Species abbreviations: ARATH, *Arabidopsis thaliana* (Mouse-ear cress); ASPFC, *Neosartorya fumigata* (Fungus); AURDE, *Auricularia delicata* (White-rot fungus); CAEEL, *Caenorhabditis elegans*; CANGA, *Candida glabrata* (Yeast); CHLRE, A8JGX8_CHLRE, *Chlamydomonas reinhardtii* (Green alga); CIOIN, *Ciona intestinalis*; CRAGI, *Crassostrea gigas* (Pacific oyster); DEBHA, *Debaryomyces hansenii* (Yeast); DICDI, *Dictyostelium discoideum* (Slime mold); EMIHU, *Emiliania huxleyi* (Chromalveolata); HUMAN, *Homo sapiens*; LOTGI, *Lottia gigantea* (Giant owl limpet); MALGO, *Malassezia globosa* (Fungus); MONBE, *Monosiga brevicollis* (Choanoflagellate); NAEGR, *Naegleria gruberi* (Amoeba); NEMVE, *Nematostella vectensis* (Starlet sea anemone); NEUCR, *Neurospora crassa* (Fungus); OSTTA, *Ostreococcus tauri* (Green alga); PHACS, *Phanerochaete carnosa* (Fungus); PICPG, *Komagataella pastoris* (Yeast); PIRID, *Piriformospora indica* (Fungus); SALR5, *Salpingoeca rosetta* (Choanoflagellate); SCHPO, *Schizosaccharomyces pombe* (Fission yeast); SELML, *Selaginella moellendorffii* (Spikemoss); STRPU, *Strongylocentrotus purpuratus* (Purple sea urchin); THAPS, *Thalassiosira pseudonana* (Marine diatom); TRIAD, *Trichoplax adhaerens* (Trichoplax); VOLCA, *Volvox carteri* (Green alga); YARLI, *Yarrowia lipolytica* (Yeast); YEAST, *Saccharomyces cerevisiae* (Baker's yeast).

FIGURE 4 | HHpred comparison E-values. The numbers correspond to E-values from HHpred (Söding et al., 2005) profile searches against a Pfam profile database which includes profiles that represent families shown in the figure. Profile-to-profile matches were evaluated in terms of an E-value, which is the expected number of non-homologous proteins with a score higher than that obtained for the database match. An E-value much lower than one indicates statistical significance. Solid lines represent statistically significant sequence similarity relationships, e.g., the MAC30/TMEM97 family calibrated profile finds with 0.003 and 0.03 E-values the profiles of EBP family and TM6SF first EXPERA repeat, respectively. The black dotted line between TM6SF second EXPERA domain repeat and EBP shows the unique relationship found with a non-highly significant value (E-value 1.2). The sequence similarity between TM6SF (second EXPERA domain repeat) and MAC30/TMEM97 families (presented inside the blue dotted oval) was already described in a DUF (Domain of Unknown Function) entry of Pfam (DUF2781, Pfam family identification: PF10914) (Bateman et al., 2010; Punta et al., 2012).

FUNCTION PREDICTION

As for the TM6SF family, the molecular function of MAC30/TMEM97 is currently poorly understood. Nevertheless, because of its wide phyletic distribution in eukaryotes (in plants, metazoa and fungi) it is likely to have a fundamental cellular function. As expected from our sequence analysis, it is mainly localized in the ER (Huh et al., 2003; Matsuyama et al., 2006). Under sterol-depleted conditions, however, it becomes enriched in the endo-lysosomal compartment where it interacts with NPC1 (Niemann-Pick disease, type C1 Protein) and regulates cellular cholesterol levels (Wilcox et al., 2007; Bartz et al., 2009). In a variety of cancers, elevated MAC30/TMEM97 expression has been directly related to unfavorable prognosis, and its down-regulation inhibits the proliferation of gastric cancer cells (Kayed et al., 2004; Zhang et al., 2006; Moparthi et al., 2007; Yan et al., 2010; Zhao et al., 2011; Han et al., 2013; Xiao et al., 2013; Yang et al., 2013; Xu et al., 2014).

The only member of the EXPERA superfamily with known molecular function is EBP, an enzyme with a $\Delta 8$, $\Delta 7$ sterol isomerase activity that catalyzes the transposition of a double bond from C8 $=$ C9 to C7 $=$ C8 in the sterol B-ring (**Figure 2**) (Wilton et al., 1969; Akhtar et al., 1970; Silve et al., 1996; Bae et al., 2001; Nes et al., 2002; Rahier et al., 2008).

EBP forms homotetramers (Nes et al., 2002) and higher-order protein complexes with sterol $\Delta 7$ reductase (DHCR7), catalyzing and regulating key steps in the final cholesterol biosynthesis pathway (Kedjouar et al., 2004; de Medina et al., 2010; Silvente-Poirot and Poirot, 2012). Mutations in EBP cause Conradi-Hünermann-Happle syndrome (also known as Chondrodysplasia punctata type II disease), a rare X-linked dominant disorder characterized by skeletal malformations, skin abnormalities, cataracts, and short stature (Braverman et al., 1999; Derry et al., 1999; Has et al., 2000, 2002; Cañueto et al., 2012, 2014).

Based on consensus transmembrane predictions over the expanded EBP family and superfamily (EXPERA domain-containing proteins) (**Figures 3, S3**) we generated a five trans-membrane EBP model (**Figures 1, 2**), that adds to and reconciles the main features of previously proposed four transmembrane EBP models (Hanner et al., 1995; Moebius et al., 1997, 2000; Dussossoy et al., 1999). Our model is in agreement with a lumenal localisation of the EBP N-terminal region (Dussossoy et al., 1999), and with a cytoplasmic localisation of its predicted C-terminal ER retention motif (Jackson et al., 1990; Hanner et al., 1995; Moebius et al., 1997, 2000). Mapping alanine-scanning mutagenesis variants (Moebius et al., 1999) and known disease-causing missense mutations (**Figure 2**) onto the EBP model, predict that the luminal, and transmembrane portions of the EBP EXPERA domain contain the greater fraction of key functional sites (**Figures 2, 3**).

Several of these sites, previously proposed contain catalytic residues (Moebius et al., 1999), are conserved not just in EBP orthologues but also across the EXPERA domain superfamily (**Figure 3**). Conservation of acidic amino acids at positions 80 and 108 (human EBP numbering) strongly suggests their involvement in catalysis involving sterols across all members of this superfamily. How these residues might catalyze sterol isomerisation is unclear, but may be similar to the enzymatic action of ketosteroid isomerases for which acidic residues (Asp or Glu) act as a proton donor or acceptor (Pollack, 2004; Sharma et al., 2006). EBP's proposed catalytic mechanism initially involves C-9 protonation of the steroid molecule, with the subsequent generation of a carbonium ion at C-8, and finally the elimination of a proton from C-7 (Wilton et al., 1969; Nes et al., 2002; Rahier et al., 2008) (**Figure 1**). It is possible that the conserved acidic residues in other EXPERA domain proteins, including MAC30/TMEM97 and TM6SF1/2, catalyze a similar sterol isomerisation reaction as proton donors and acceptors.

The homologous relationships described here between TM6SF2 and EBP could also explain the reported side-effects of tamoxifen (Oien et al., 1999; Hackshaw et al., 2011), which is an antagonist of the estrogen receptor commonly used in breast cancer therapy (Jordan, 2000). Drug cross-reactivity among homologous proteins frequently underlies undesired pleiotropic effects (Searls, 2003; Campillos et al., 2008).

Tamoxifen is a known inhibitor of EBP (Moebius et al., 1998; Kedjouar et al., 2004; de Medina et al., 2010; Silvente-Poirot and Poirot, 2012), and if its homolog TM6SF is similarly inhibited by tamoxifen then this would result in a reduction of total cholesterol level and the induction of NAFLD (Holmen et al., 2014; Kozlitina et al., 2014; Liu et al., 2014; Sookoian et al., 2014) which are precisely the known side-effects of tamoxifen administration (Oien et al., 1999; Hackshaw et al., 2011).

MAC30/TMEM97 is expressed at high levels in breast, esophagus, stomach, and colon cancers (Kayed et al., 2004). Human gastric cancer cells are known to have reduced cellular proliferation and mobility when MAC30/TMEM97 transcript levels are down-regulated (Xu et al., 2014). This implies that the inhibition of MAC30/TMEM97 catalytic activity by tamoxifen may also lead to reduced proliferation of cancer cells. Identification of the EXPERA domain family may thus help to elucidate the complex interplay between cancer and cholesterol metabolism (Silvente-Poirot and Poirot, 2014).

In summary, our analyses have identified TM6SF1, TM6SF2, and MAC30/TMEM97 as EBP homologs. This indicates that these proteins are all likely to possess similar catalytic activities, potentially as sterol isomerases. These results provide new opportunities for their experimental characterization, and for the development of drugs that would inhibit members of the EXPERA superfamily.

AUTHOR CONTRIBUTIONS

Luis Sanchez-Pulido and Chris P. Ponting designed the research and wrote the paper.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00439/abstract

We include as Supplementary Material three full-length multiple sequence alignments of representative sequences and transmembrane predictions for the EXPERA domain-containing families: TM6SF, MAC30/TMEM97 and EBP (**Figures S1–S3**).

Figure S1 | TM6SF family representative full-length alignment and transmembrane prediction. The sequence variant identified in TM6SF2 is labeled in red (Holmen et al., 2014; Kozlitina et al., 2014; Sookoian et al., 2014). EXPERA domains limits are marked with red bars (light red for the first repeat and dark red for the second repeat) above the alignment. The most conserved positions of both EXPERA domains (E101 and E255) are labeled. TMHMM helix transmembrane (Krogh et al., 2001) predictions are shown below each input sequence (consensus of these predictions is shown in **Figure 3**). The alignment was presented with the program Belvu (Sonnhammer and Hollich, 2005) using a coloring scheme indicating the average BLOSUM62 scores (which are correlated with amino acid conservation) of each alignment column: black (>3), gray (between 3 and 1.5) and light gray (between 1.5 and 0.5). Sequences are named according

to their UniProt identifications (Wu et al., 2006). Human proteins identifications are underlined in red (TM6SF2, Q9BZW4_HUMAN and TM6SF1, Q9BZW5_HUMAN). Species abbreviations: CALMI, *Callorhynchus milii* (Australian ghost shark); CIOIN, *Ciona intestinalis*; CRAGI, *Crassostrea gigas* (Pacific oyster); HUMAN, *Homo sapiens*; MONBE, *Monosiga brevicollis* (Choanoflagellate); ORENI, *Oreochromis niloticus* (Nile tilapia); SALR5, *Salpingoeca rosetta* (Choanoflagellate); STRPU, *Strongylocentrotus purpuratus* (Purple sea urchin); TETUR, *Tetranychus urticae* (Chelicerata); XENTR, *Xenopus tropicalis*.

Figure S2 | MAC30 family representative full-length alignment and transmembrane prediction. The extent of the EXPERA domain is marked with a yellow bar above the alignment. The most conserved position of the EXPERA superfamily (D56) is labeled. TMHMM helix transmembrane (Krogh et al., 2001) predictions are shown below each input sequence (consensus of these predictions is shown in **Figure 3**). The alignment was presented with the program Belvu (Sonnhammer and Hollich, 2005) using a coloring scheme indicating the average BLOSUM62 scores (which are correlated with amino acid conservation) of each alignment column: black (>3), gray (between 3 and 1.5) and light gray (between 1.5 and 0.5). Sequences are named according to their UniProt identifications (Wu et al., 2006). MAC30/TMEM97 human protein identification is underlined in yellow (Q5BJF2_HUMAN). Species abbreviations: 9CHLO, *Coccomyxa subellipsoidea*; ACACA, *Acanthamoeba castellanii*; ARATH, *Arabidopsis thaliana* (Mouse-ear cress); BYSSP, *Byssochlamys spectabilis*; CANAW, *Candida albicans*; CLAL4, *Clavispora lusitaniae*; COCIM, *Coccidioides immitis*; DEBHA, *Debaryomyces hansenii* (Yeast); HUMAN, *Homo sapiens*; LOTGI, *Lottia gigantea* (Giant owl limpet); MARBU, *Marssonina brunnea*; MICGM, *Mycosphaerella graminicola*; MICPC, *Micromonas pusilla*; MUSAM, *Musa acuminata*; PHATC, *Phaeodactylum tricornutum*; PICPG, *Komagataella pastoris* (Yeast); PICSI, *Picea sitchensis*; PICST, *Scheffersomyces stipitis*; PRUPE, *Prunus persica*; PSED2, *Pseudogymnoascus destructans*; SCHPO, *Schizosaccharomyces pombe* (Fission yeast); SELML, *Selaginella moellendorffii* (Spikemoss); TRIAD, *Trichoplax adhaerens*; VOLCA, *Volvox carteri* (Green alga); XENTR, *Xenopus tropicalis*; YEAST, *Saccharomyces cerevisiae* (Baker's yeast).

Figure S3 | EBP family representative full-length alignment and transmembrane prediction. The extent of the EXPERA domain is marked with a violet bar above the alignment. The most conserved position of the EXPERA superfamily (D108) is labeled. TMHMM helix transmembrane (Krogh et al., 2001) predictions are shown below each input sequence (consensus of these predictions is shown in **Figure 3**). The alignment was presented with the program Belvu (Sonnhammer and Hollich, 2005) using a coloring scheme indicating the average BLOSUM62 scores (which are correlated with amino acid conservation) of each alignment column: black (>3), gray (between 3 and 1.5) and light gray (between 1.5 and 0.5). Sequences are named according to their UniProt identifications (Wu et al., 2006). Human proteins identifications are underlined in violet (EBP, Q15125_HUMAN and EBPL, Q9BY08_HUMAN). Species abbreviations: 9AGAR, *Moniliophthora roreri*; 9GLOM, *Rhizophagus irregularis*; ACACA, *Acanthamoeba castellanii*; ARATH, *Arabidopsis thaliana* (Mouse-ear cress); ASPO3, *Aspergillus oryzae*; BRAFL, *Branchiostoma floridae* (Amphioxus); CAEEL, *Caenorhabditis elegans*; CAPO3, *Capsaspora owczarzaki*; COPC7, *Coprinopsis cinerea*; CRAGI, *Crassostrea gigas* (Pacific oyster); HUMAN, *Homo sapiens*; LOTGI, *Lottia gigantea* (Giant owl limpet); MAGO7, *Magnaporthe oryzae*; MONDO, *Monodelphis domestica* (opossum); NAEGR, *Naegleria gruberi* (Amoeba); PHACS, *Phanerochaete carnosa*; RHOT1, *Rhodosporidium toruloides*; SELML, *Selaginella*

moellendorffii (Spikemoss); STRPU, *Strongylocentrotus purpuratus* (Purple sea urchin); TALSN, *Talaromyces stipitatus*; TRIAD, *Trichoplax adhaerens*; VERA1, *Verticillium alfalfae*.

REFERENCES

Addou, S., Rentzsch, R., Lee, D., and Orengo, C. A. (2009). Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *J. Mol. Biol.* 387, 416–430. doi: 10.1016/j.jmb.2008.12.045

Akhtar, M., Rahimtula, A. D., and Wilton, D. C. (1970). The stereochemistry of hydrogen elimination from C-7 in cholesterol and ergosterol biosynthesis. *Biochem. J.* 117, 539–542.

Arnold, A. W., Bruckner-Tuderman, L., Has, C., and Happle, R. (2012). Conradi-Hünermann-Happle syndrome in males vs. MEND syndrome (male EBP disorder with neurological defects). *Br. J. Dermatol.* 166, 1309–1313. doi: 10.1111/j.1365-2133.2012.10808.x

Aughton, D. J., Kelley, R. I., Metzenberg, A., Pureza, V., and Pauli, R. M. (2003). X-linked dominant chondrodysplasia punctata (CDPX2) caused by single gene mosaicism in a male. *Am. J. Med. Genet. A* 116A, 255–260. doi: 10.1002/ajmg.a.10852

Ausavarat, S., Tanpaiboon, P., Tongkobpetch, S., Suphapeetiporn, K., and Shotelersuk, V. (2008). Two novel EBP mutations in Conradi-Hünermann-Happle syndrome. *Eur. J. Dermatol.* 18, 391–393. doi: 10.1684/ejd.2008.0433

Babbs, C., Roberts, N. A., Sanchez-Pulido, L., McGowan, S. J., Ahmed, M. R., Brown, J. M., et al. (2013). Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type I. *Haematologica* 98, 1383–1387. doi: 10.3324/haematol.2013.089490

Bae, S., Seong, J., and Paik, Y. (2001). Cholesterol biosynthesis from lanosterol: molecular cloning, chromosomal localization, functional expression and liver-specific gene regulation of rat sterol delta8-isomerase, a cholesterogenic enzyme with multiple functions. *Biochem. J.* 353, 689–699. doi: 10.1042/0264-6021:3530689

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., et al. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 12, 745–755. doi: 10.1038/nrg3031

Barboza-Cerda, M. C., Wong, L.-J., Martínez-de-Villarreal, L. E., Zhang, V. W., and Déctor, M. A. (2014). A novel EBP c.224T>A mutation supports the existence of a male-specific disorder independent of CDPX2. *Am. J. Med. Genet. A* 164, 1642–1647. doi: 10.1002/ajmg.a.36508

Bartlett, G. J., Borkakoti, N., and Thornton, J. M. (2003). Catalysing new reactions during evolution: economy of residues and mechanism. *J. Mol. Biol.* 331, 829–860. doi: 10.1016/S0022-2836(03)00734-4

Bartz, F., Kern, L., Erz, D., Zhu, M., Gilbert, D., Meinhof, T., et al. (2009). Identification of cholesterol-regulating genes by targeted RNAi screening. *Cell Metab.* 10, 63–75. doi: 10.1016/j.cmet.2009.05.009

Bateman, A., Coggill, P., and Finn, R. D. (2010). DUFs: families in search of function. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 66, 1148–1152. doi: 10.1107/S1744309110001685

Becker, K., Csikós, M., Horváth, A., and Kárpáti, S. (2001). Identification of a novel mutation in 3beta-hydroxysteroid-Delta8-Delta7-isomerase in a case of Conradi-Hünermann-Happle syndrome. *Exp. Dermatol.* 10, 286–289. doi: 10.1034/j.1600-0625.2001.100409.x

Bode, H., Galm, C., Hummler, H., Teller, C., Haas, D., and Gencik, M. (2013). Non-lethal non-mosaic male with Conradi-Hunermann syndrome caused by a novel EBP c.356T>G mutation. *Am. J. Med. Genet. A* 161A, 2385–2388. doi: 10.1002/ajmg.a.35985

Braverman, N., Lin, P., Moebius, F. F., Obie, C., Moser, A., Glossmann, H., et al. (1999). Mutations in the gene encoding 3 beta-hydroxysteroid-delta 8, delta 7-isomerase cause X-linked dominant Conradi-Hünermann syndrome. *Nat. Genet.* 22, 291–294. doi: 10.1038/10357

Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J., and Bork, P. (2008). Drug target identification using side-effect similarity. *Science* 321, 263–266. doi: 10.1126/science.1158140

Cañueto, J., Girós, M., Ciria, S., Pi-Castán, G., Artigas, M., García-Dorado, J., et al. (2012). Clinical, molecular and biochemical characterization of nine Spanish families with Conradi-Hünermann-Happle syndrome: new insights into X-linked dominant chondrodysplasia punctata with a comprehensive review of the literature. *Br. J. Dermatol.* 166, 830–838. doi: 10.1111/j.1365-2133.2011.10756.x

Cañueto, J., Girós, M., and González-Sarmiento, R. (2014). The role of the abnormalities in the distal pathway of cholesterol biosynthesis in the Conradi-Hünermann-Happle syndrome. *Biochim. Biophys. Acta* 1841, 336–344. doi: 10.1016/j.bbalip.2013.09.002

de Medina, P., Paillasse, M. R., Segala, G., Poirot, M., and Silvente-Poirot, S. (2010). Identification and pharmacological characterization of cholesterol-5,6-epoxide hydrolase as a target for tamoxifen and AEBS ligands. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13520–13525. doi: 10.1073/pnas.1002922107

Derry, J. M., Gormally, E., Means, G. D., Zhao, W., Meindl, A., Kelley, R. I., et al. (1999). Mutations in a delta 8-delta 7 sterol isomerase in the tattered mouse and X-linked dominant chondrodysplasia punctata. jderry@immunex.com. *Nat. Genet.* 22, 286–290. doi: 10.1038/10350

Do, R., Kathiresan, S., and Abecasis, G. R. (2012). Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum. Mol. Genet.* 21, R1–R9. doi: 10.1093/hmg/dds387

Dongiovanni, P., Petta, S., Maglio, C., Fracanzani, A. L., Pipitone, R., Mozzi, E., et al. (2014). TM6SF2 gene variant disentangles nonalcoholic steatohepatitis from cardiovascular disease. *Hepatology.* doi: 10.1002/hep.27490. [Epub ahead of print].

Dussossoy, D., Carayon, P., Belugou, S., Feraut, D., Bord, A., Goubet, C., et al. (1999). Colocalization of sterol isomerase and sigma(1) receptor at endoplasmic reticulum and nuclear envelope level. *Eur. J. Biochem.* 263, 377–386. doi: 10.1046/j.1432-1327.1999.00500.x

Eddy, S. R. (1996). Hidden Markov models. *Curr. Opin. Struct. Biol.* 6, 361–365. doi: 10.1016/S0959-440X(96)80056-X

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Furtado, L. V., Bayrak-Toydemir, P., Hulinsky, B., Damjanovich, K., Carey, J. C., and Rope, A. F. (2010). A novel X-linked multiple congenital anomaly syndrome associated with an EBP mutation. *Am. J. Med. Genet. A* 152A, 2838–2844. doi: 10.1002/ajmg.a.33674

Galperin, M. Y., and Koonin, E., V (2004). "Conserved hypothetical" proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* 32, 5452–5463. doi: 10.1093/nar/gkh885

Gollery, M., Harper, J., Cushman, J., Mittler, T., Girke, T., Zhu, J.-K., et al. (2006). What makes species unique? The contribution of proteins with obscure features. *Genome Biol.* 7:R57. doi: 10.1186/gb-2006-7-7-r57

Goodstadt, L., and Ponting, C. P. (2001). Sequence variation and disease in the wake of the draft human genome. *Hum. Mol. Genet.* 10, 2209–2214. doi: 10.1093/hmg/10.20.2209

Hackshaw, A., Roughton, M., Forsyth, S., Monson, K., Reczko, K., Sainsbury, R., et al. (2011). Long-term benefits of 5 years of tamoxifen: 10-year follow-up of a large randomized trial in women at least 50 years of age with early breast cancer. *J. Clin. Oncol.* 29, 1657–1663. doi: 10.1200/JCO.2010.32.2933

Han, K.-Y., Gu, X., Wang, H.-R., Liu, D., Lv, F.-Z., and Li, J.-N. (2013). Overexpression of MAC30 is associated with poor clinical outcome in human non-small-cell lung cancer. *Tumour Biol.* 34, 821–825. doi: 10.1007/s13277-012-0612-z

Hanner, M., Moebius, F. F., Weber, F., Grabner, M., Striessnig, J., and Glossmann, H. (1995). Phenylalkylamine Ca2+ antagonist binding protein. Molecular cloning, tissue distribution, and heterologous expression. *J. Biol. Chem.* 270, 7551–7557.

Has, C., Bruckner-Tuderman, L., Müller, D., Floeth, M., Folkers, E., Donnai, D., et al. (2000). The Conradi-Hünermann-Happle syndrome (CDPX2) and emopamil binding protein: novel mutations, and somatic and gonadal mosaicism. *Hum. Mol. Genet.* 9, 1951–1955. doi: 10.1093/hmg/9.13.1951

Has, C., Seedorf, U., Kannenberg, F., Bruckner-Tuderman, L., Folkers, E., Fölster-Holst, R., et al. (2002). Gas chromatography-mass spectrometry and molecular genetic studies in families with the Conradi-Hünermann-Happle syndrome. *J. Invest. Dermatol.* 118, 851–858. doi: 10.1046/j.1523-1747.2002.01761.x

Hello, M., David, A., and Barbarot, S. (2010). [Conradi-Hünermann-Happle syndrome with unilateral distribution]. *Ann. Dermatol. Venereol.* 137, 44–47. doi: 10.1016/j.annder.2009.11.006

Herman, G. E., Kelley, R. I., Pureza, V., Smith, D., Kopacz, K., Pitt, J., et al. (2002). Characterization of mutations in 22 females with X-linked dominant chondrodysplasia punctata (Happle syndrome). *Genet. Med.* 4, 434–438. doi: 10.1097/00125817-200211000-00006

Holmen, O. L., Zhang, H., Fan, Y., Hovelson, D. H., Schmidt, E. M., Zhou, W., et al. (2014). Systematic evaluation of coding variation identifies a candidate causal

variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat. Genet.* 46, 345–351. doi: 10.1038/ng.2926

Hou, J.-W. (2013). Conradi-Hünermann-Happle syndrome with cervical stenosis. *Pediatr. Neurol.* 49, 513–514. doi: 10.1016/j.pediatrneurol.2013.06.017

Huh, W. -K., Falvo, J. V, Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., et al. (2003). Global analysis of protein localization in budding yeast. *Nature* 425, 686–691. doi: 10.1038/nature02026

Ikegawa, S., Ohashi, H., Ogata, T., Honda, A., Tsukahara, M., Kubo, T., et al. (2000). Novel and recurrent EBP mutations in X-linked dominant chondrodysplasia punctata. *Am. J. Med. Genet.* 94, 300–305.

Jackson, M. R., Nilsson, T., and Peterson, P. A. (1990). Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *EMBO J.* 9, 3153–3162.

Jordan, V. C. (2000). Tamoxifen: a personal retrospective. *Lancet. Oncol.* 1, 43–49. doi: 10.1016/S1470-2045(00)00009-7

Kayed, H., Kleeff, J., Ding, J., Hammer, J., Giese, T., Zentgraf, H., et al. (2004). Expression analysis of MAC30 in human pancreatic cancer and tumors of the gastrointestinal tract. *Histol. Histopathol.* 19, 1021–1031.

Kedjouar, B., de Médina, P., Oulad-Abdelghani, M., Payré, B., Silvente-Poirot, S., Favre, G., et al. (2004). Molecular characterization of the microsomal tamoxifen binding site. *J. Biol. Chem.* 279, 34048–34061. doi: 10.1074/jbc.M405230200

Kiezun, A., Garimella, K., Do, R., Stitziel, N. O., Neale, B. M., McLaren, P. J., et al. (2012). Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* 44, 623–630. doi: 10.1038/ng.2303

Kolb-Mäurer, A., Grzeschik, K.-H., Haas, D., Bröcker, E.-B., and Hamm, H. (2008). Conradi-Hünermann-Happle syndrome (X-linked dominant chondrodysplasia punctata) confirmed by plasma sterol and mutation analysis. *Acta Derm. Venereol.* 88, 47–51. doi: 10.2340/00015555-0337

Kozlitina, J., Smagris, E., Stender, S., Nordestgaard, B. G., Zhou, H. H., Tybjærg-Hansen, A., et al. (2014). Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* 46, 352–356. doi: 10.1038/ng.2901

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Lambrecht, C., Wouters, C., Van Esch, H., Moens, P., Casteels, I., and Morren, M.-A., (2014). Conradi-Hünermann-Happle syndrome: a novel heterozygous missense mutation, c.204G>T (p.W68C). *Pediatr. Dermatol.* 31, 493–496. doi: 10.1111/pde.12336

Lee, D., Redfern, O., and Orengo, C. (2007). Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell Biol.* 8, 995–1005. doi: 10.1038/nrm2281

Liu, Y.-L., Reeves, H. L., Burt, A. D., Tiniakos, D., McPherson, S., Leathart, J. B. S., et al. (2014). TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nat. Commun.* 5, 4309. doi: 10.1038/ncomms5309

Mahdessian, H., Taxiarchis, A., Popov, S., Silveira, A., Franco-Cereceda, A., Hamsten, A., et al. (2014). TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8913–8918. doi: 10.1073/pnas.1323785111

Matsuyama, A., Arai, R., Yashiroda, Y., Shirai, A., Kamata, A., Sekido, S., et al. (2006). ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. *Nat. Biotechnol.* 24, 841–847. doi: 10.1038/nbt1222

Milunsky, J. M., Maher, T. A., and Metzenberg, A. B. (2003). Molecular, biochemical, and phenotypic analysis of a hemizygous male with a severe atypical phenotype for X-linked dominant Conradi-Hunerman-Happle syndrome and a mutation in EBP. *Am. J. Med. Genet. A* 116A, 249–254. doi: 10.1002/ajmg.a.10849

Moebius, F. F., Fitzky, B. U., and Glossmann, H. (2000). Genetic defects in post-squalene cholesterol biosynthesis. *Trends Endocrinol. Metab.* 11, 106–114. doi: 10.1016/S1043-2760(00)00235-6

Moebius, F. F., Reiter, R. J., Bermoser, K., Glossmann, H., Cho, S. Y., and Paik, Y. K. (1998). Pharmacological analysis of sterol delta8-delta7 isomerase proteins with [3H]ifenprodil. *Mol. Pharmacol.* 54, 591–598.

Moebius, F. F., Soellner, K. E., Fiechtner, B., Huck, C. W., Bonn, G., and Glossmann, H. (1999). Histidine77, glutamic acid81, glutamic acid123, threonine126, asparagine194, and tryptophan197 of the human emopamil binding protein

are required for *in vivo* sterol delta 8-delta 7 isomerization. *Biochemistry* 38, 1119–1127. doi: 10.1021/bi981804i

Moebius, F. F., Striessnig, J., and Glossmann, H. (1997). The mysteries of sigma receptors: new family members reveal a role in cholesterol synthesis. *Trends Pharmacol. Sci.* 18, 67–70. doi: 10.1016/S0165-6147(96)01037-1

Moparthi, S. B., Arbman, G., Wallin, A., Kayed, H., Kleeff, J., Zentgraf, H., et al. (2007). Expression of MAC30 protein is related to survival and biological variables in primary and metastatic colorectal cancers. *Int. J. Oncol.* 30, 91–95.

Morice-Picard, F., Kostrzewa, E., Wolf, C., Benlian, P., Taïeb, A., and Lacombe, D. (2011). Evidence of postzygotic mosaicism in a transmitted form of Conradi-Hunermann-Happle syndrome associated with a novel EBP mutation. *Arch. Dermatol.* 147, 1073–1076. doi: 10.1001/archdermatol.2011.230

Nes, W. D., Zhou, W., Dennis, A. L., Li, H., Jia, Z., Keith, R. A., et al. (2002). Purification, characterization and catalytic properties of human sterol 8-isomerase. *Biochem. J.* 367, 587–599. doi: 10.1042/BJ20020551

Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. doi: 10.1006/jmbi.2000.4042

Oien, K. A., Moffat, D., Curry, G. W., Dickson, J., Habeshaw, T., Mills, P. R., et al. (1999). Cirrhosis with steatohepatitis after adjuvant tamoxifen. *Lancet* 353, 36–37. doi: 10.1016/S0140-6736(05)74872-8

Pollack, R. M. (2004). Enzymatic mechanisms for catalysis of enolization: ketosteroid isomerase. *Bioorg. Chem.* 32, 341–353. doi: 10.1016/j.bioorg.2004.06.005

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301. doi: 10.1093/nar/gkr1065

Rahier, A., Pierre, S., Riveill, G., and Karst, F. (2008). Identification of essential amino acid residues in a sterol 8,7-isomerase from Zea mays reveals functional homology and diversity with the isomerases of animal and fungal origin. *Biochem. J.* 414, 247–259. doi: 10.1042/BJ20080292

Sanchez-Pulido, L., Kong, L., and Ponting, C. P. (2012). A common ancestry for BAP1 and Uch37 regulators. *Bioinformatics* 28, 1953–1956. doi: 10.1093/bioinformatics/bts319

Sánchez-Pulido, L., Yuan, Y. P., Andrade, M. A., and Bork, P. (2000). NAIL-Network Analysis Interface for Linking HMMER results. *Bioinformatics* 16, 656–657. doi: 10.1093/bioinformatics/16.7.656

Searls, D. B. (2003). Pharmacophylogenomics: genes, evolution and drug targets. *Nat. Rev. Drug Discov.* 2, 613–623. doi: 10.1038/nrd1152

Sharma, K., Vázquez-Ramírez, R., and Kubli-Garfias, C. (2006). A theoretical model of the catalytic mechanism of the Delta5-3-ketosteroid isomerase reaction. *Steroids* 71, 549–557. doi: 10.1016/j.steroids.2005.12.001

Shimizu, T., Mitsuke, H., Noto, K., and Arai, M. (2004). Internal gene duplication in the evolution of prokaryotic transmembrane proteins. *J. Mol. Biol.* 339, 1–15. doi: 10.1016/j.jmb.2004.03.048

Shirahama, S., Miyahara, A., Kitoh, H., Honda, A., Kawase, A., Yamada, K., et al. (2003). Skewed X-chromosome inactivation causes intra-familial phenotypic variation of an EBP mutation in a family with X-linked dominant chondrodysplasia punctata. *Hum. Genet.* 112, 78–83. doi: 10.1007/s00439-002-0844-x

Silve, S., Dupuy, P. H., Labit-Lebouteiller, C., Kaghad, M., Chalon, P., Rahier, A., et al. (1996). Emopamil-binding protein, a mammalian protein that binds a series of structurally diverse neuroprotective agents, exhibits delta8-delta7 sterol isomerase activity in yeast. *J. Biol. Chem.* 271, 22434–22440. doi: 10.1074/jbc.271.37.22434

Silvente-Poirot, S., and Poirot, M. (2012). Cholesterol epoxide hydrolase and cancer. *Curr. Opin. Pharmacol.* 12, 696–703. doi: 10.1016/j.coph.2012.07.007

Silvente-Poirot, S., and Poirot, M. (2014). Cancer. Cholesterol and cancer, in the balance. *Science* 343, 1445–1446. doi: 10.1126/science.1252787

Söding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248. doi: 10.1093/nar/gki408

Sonnhammer, E. L. L., and Hollich, V. (2005). Scoredist: a simple and robust protein sequence distance estimator. *BMC Bioinform.* 6:108. doi: 10.1186/1471-2105-6-108

Sookoian, S., Castaño, G. O., Scian, R., Mallardi, P., Fernández Gianotti, T., Burgueño, A. L., et al. (2014). Genetic variation in TM6SF2 and the risk of nonalcoholic fatty liver disease and histological disease severity. *Hepatology*. doi: 10.1002/hep.27556. [Epub ahead of print].

Steijlen, P. M., van Geel, M., Vreeburg, M., Marcus-Soekarman, D., Spaapen, L. J. M., Castelijns, F. C. M., et al. (2007). Novel EBP gene mutations in

Conradi-Hünermann-Happle syndrome. *Br. J. Dermatol.* 157, 1225–1229. doi: 10.1111/j.1365-2133.2007.08254.x

Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S. T., et al. (2003). Human gene mutation database (HGMD): 2003 update. *Hum. Mutat.* 21, 577–581. doi: 10.1002/humu.10212

Tysoe, C., Law, C. J., Caswell, R., Clayton, P., and Ellard, S. (2008). Prenatal testing for a novel EBP missense mutation causing X-linked dominant chondrodysplasia punctata. *Prenat. Diagn.* 28, 384–388. doi: 10.1002/pd.1980

Umekoji, A., Fukai, K., Kasama, T., Yokoi, T., Saito, M., Tsuruhara, A., et al. (2008). High 8-dehydrocholesterol level in a typical case of Conradi-Hunermann-Happle syndrome with a novel H76Y missense mutation. *J. Dermatol. Sci.* 51, 62–65. doi: 10.1016/j.jdermsci.2008.02.005

Whittock, N. V, Izatt, L., Mann, A., Homfray, T., Bennett, C., Mansour, S., et al. (2003). Novel mutations in X-linked dominant chondrodysplasia punctata (CDPX2). *J. Invest. Dermatol.* 121, 939–942. doi: 10.1046/j.1523-1747.2003.12489.x

Wilcox, C. B., Feddes, G. O., Willett-Brozick, J. E., Hsu, L.-C., DeLoia, J. A., and Baysal, B. E. (2007). Coordinate up-regulation of TMEM97 and cholesterol biosynthesis genes in normal ovarian surface epithelial cells treated with progesterone: implications for pathogenesis of ovarian cancer. *BMC Cancer* 7:223. doi: 10.1186/1471-2407-7-223

Wilton, D. C., Rahimtula, A. D., and Akhtar, M. (1969). The reversibility of the delta8-cholestenol-delta7-cholestenol isomerase reaction in cholesterol biosynthesis. *Biochem. J.* 114, 71–73.

Wu, C. H., Apweiler, R., Bairoch, A., Natale, D. A., Barker, W. C., Boeckmann, B., et al. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* 34, D187–D191. doi: 10.1093/nar/gkj161

Xiao, M., Li, H., Yang, S., Huang, Y., Jia, S., Wang, H., et al. (2013). Expression of MAC30 protein is related to survival and clinicopathological variables in breast cancer. *J. Surg. Oncol.* 107, 456–462. doi: 10.1002/jso.23269

Xu, X.-Y., Zhang, L.-J., Yu, Y.-Q., Zhang, X.-T., Huang, W., Nie, X.-C., et al. (2014). Down-regulated MAC30 expression inhibits proliferation and mobility of human gastric cancer cells. *Cell. Physiol. Biochem.* 33, 1359–1368. doi: 10.1159/000358703

Yan, B.-Y., Wang, D.-W., Zhu, Z.-L., Yang, Y.-H., Wang, M.-W., Cui, D.-S., et al. (2010). Overexpression of MAC30 in the cytoplasm of oral squamous cell carcinoma predicts nodal metastasis and poor differentiation. *Chemotherapy* 56, 424–428. doi: 10.1159/000317582

Yang, S., Li, H., Liu, Y., Ning, X., Meng, F., Xiao, M., et al. (2013). Elevated expression of MAC30 predicts lymph node metastasis and unfavorable prognosis in patients with epithelial ovarian cancer. *Med. Oncol.* 30:324. doi: 10.1007/s12032-012-0324-7

Zhang, D., Iyer, L. M., He, F., and Aravind, L. (2012). Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. *Front. Genet.* 3:283. doi: 10.3389/fgene.2012.00283

Zhang, Z.-Y., Zhao, Z.-R., Adell, G., Jarlsfelt, I., Cui, Y.-X., Kayed, H., et al. (2006). Expression of MAC30 in rectal cancers with or without preoperative radiotherapy. *Oncology* 71, 259–265. doi: 10.1159/000106449

Zhao, Z.-R., Zhang, L.-J., He, X.-Q., Zhang, Z.-Y., Zhang, F., Li, F., et al. (2011). Significance of mRNA and protein expression of MAC30 in progression of colorectal cancer. *Chemotherapy* 57, 394–401. doi: 10.1159/000331716