BMC Genomics

BioMed Central

Research article

Open Access

TMC and EVER genes belong to a larger novel family, the TMC gene family encoding transmembrane proteins

Gabor Keresztes[†1,2], Hideki Mutai[†1,2] and Stefan Heller[*1,2]

Address: [1]Department of Otolaryngology and Program in Neuroscience, Harvard Medical School and [2]Eaton Peabody Laboratory, Massachusetts Eye & Ear Infirmary, Boston, MA

Email: Gabor Keresztes - gabor@epl.meei.harvard.edu; Hideki Mutai - mutai@epl.meei.harvard.edu; Stefan Heller* - hellers@epl.meei.harvard.edu

* Corresponding author †Equal contributors

Published: 17 June 2003

BMC Genomics 2003, **4**:24

Received: 27 March 2003
Accepted: 17 June 2003

This article is available from: http://www.biomedcentral.com/1471-2164/4/24

Abstract

Background: Mutations in the transmembrane cochlear expressed gene 1 (*TMC1*) cause deafness in human and mouse. Mutations in two homologous genes, *EVER1* and *EVER2* increase the susceptibility to infection with certain human papillomaviruses resulting in high risk of skin carcinoma. Here we report that *TMC1*, *EVER1* and *EVER2* (now *TMC6* and *TMC8*) belong to a larger novel gene family, which is named TMC for trans membrane channel-like gene family.

Results: Using a combination of iterative database searches and reverse transcriptase-polymerase chain reaction (RT-PCR) experiments we assembled contigs for cDNA encoding human, murine, puffer fish, and invertebrate TMC proteins. TMC proteins of individual species can be grouped into three subfamilies A, B, and C. Vertebrates have eight TMC genes. The majority of murine TMC transcripts are expressed in most organs; some transcripts, however, in particular the three subfamily A members are rare and more restrictively expressed.

Conclusion: The eight vertebrate TMC genes are evolutionary conserved and encode proteins that form three subfamilies. Invertebrate TMC proteins can also be categorized into these three subfamilies. All TMC genes encode transmembrane proteins with intracellular amino- and carboxyl-termini and at least eight membrane-spanning domains. We speculate that the TMC proteins constitute a novel group of ion channels, transporters, or modifiers of such.

Background

The complete sequencing of the human genome led to the conclusion that the number of human genes is approximately 30,000 – 40,000 [1,2]. This initial estimate was proven to be incorrect because a comparison of the predicted genes that were identified in public and commercial sequencing projects and inclusion of a third cluster of known genes revealed little overlap among the three groups [3]. Although many of the discrepancies between the different human genome assemblies are caused by fundamental disparities in the compilation of the sequence information [4], it is still evident that the human genome harbors probably significantly more than 40,000 genes and that these genes are not easily identified. This is particularly true for genes that are either expressed transiently, at low levels, or in very few specific cell types. Such elusive genes are underrepresented in cDNA libraries and consequently also in expressed sequence tag (EST) databases. Computational annotation of such genes from genomic DNA becomes additionally challenging when the genes consist of many scattered small exons.

Two pairs of previously un-notated genes encoding transmembrane proteins have recently been identified based on their linkage with inherited disorders. The first set of genes, transmembrane cochlear expressed genes 1 and 2 (*TMC1*) and *TMC2* [5] were found in a search for mutations causing dominant and recessive deafness in human and mouse [5,6]. Mutations in the second pair of genes, *EVER1* and *EVER2*, are linked to epidermodysplasia verruciformis that is associated with an increased susceptibility to infection with some human papillomaviruses causing a high risk of skin carcinoma [7]. We noticed that the proteins encoded by these genes are homologous, which raised our interest in testing whether the genome accommodates additional related genes.

Results

The mammalian TMC gene and protein family

We used human TMC1 and TMC2, murine Tmc1 and Tmc2, and human EVER1 and EVER2 sequences in conjunction with database search algorithms to identify homologous sequences in genomic and expressed sequence tag (EST) databases. We iteratively assembled contiguous theoretical coding sequences for individual human and mouse proteins. We verified regions of individual murine cDNAs that were not unequivocally predictable with database information by RT-PCR and by sequencing. This strategy resulted in the identification of the murine orthologues of EVER1 and EVER2, and four hitherto uncharacterized human and mouse proteins (Figure 1, Table 1). All identified proteins have eight predicted membrane-spanning domains (TM1 – TM8) and share the completely conserved amino acid triplet C (cysteine) – W (tryptophan) – E (glutamic acid), predicted to be located in the extracellular loop upstream of TM6 (Figure 2). The presence of this hereby named *TMC signature sequence motif* CWETXVGQEly(K/R)LtvXD is our defining criterion for this novel TMC protein family (Figures 1, 2, additional file 2).

The eight mammalian TMC proteins can be grouped into three subfamilies A, B, and C, based on sequence homology and on similarities of the genomic structure of their respective genes (Figures 1, 3, and 4). The murine TMC protein subfamily A consists of three proteins, Tmc1, Tmc2, and the novel Tmc3. All TMC subfamily A proteins are between 757 and 1130 amino acids in length. Tmc3 bears a long carboxyl-terminal tail that is unlike all other TMC proteins (Figure 1). The overall identity within the TMC subfamily A is 36–56%; the positions of >73% of the genes' introns within the conserved core region are conserved (Figure 3A, 3C, additional file 1).

The murine TMC subfamily B consists of Tmc5 and Tmc6 (mouse orthologue of EVER1), proteins of 810 and 757 amino acid residues that are 31% identical and share a >92% conservation of the corresponding genes' intron locations within the conserved core region (Figure 3A, 3C, additional file 1). A significant structural difference between subfamily B and subfamily C proteins is that the long presumptive extracellular loop of TMC subfamily A proteins between TM5 and TM6 is much shorter in subfamily B proteins and mainly consists of the TMC signature sequence motif (Figures 1, 2).

Finally, the three members of the murine TMC subfamily C, Tmc4, Tmc7, and Tmc8 (mouse orthologue of EVER2) are the shortest TMC proteins with 694, 726, and 722 amino acid residues. The overall identity within the murine TMC protein subfamily C is 29–33% with a common gene structure of >92% conserved intron locations within the conserved core region (Figure 3A, 3C, additional file 1).

Analysis of the human TMC gene and protein family yielded principally identical results, likely because of the high degree of conservation between human and murine TMC proteins [5] (Figure 4). The new TMC classification now designates *TMC6* for *EVER1*, and *TMC8* for *EVER2* (Table 1).

The TMC genes map to six chromosomal locations in the human and mouse (Table 1). Two chromosomal locations in both species harbor two neighboring TMC genes, *Tmc5* and *Tmc7* on murine chromosome 7, and *Tmc6* and *Tmc8* on murine chromosome 11; the human orthologues are located on the syntenic regions of chromosomes 16 and 17, respectively [7]. An additional murine locus on chromosome 8 represents a partial gene fragment, likely a pseudogene of *Tmc2*.

Expression of murine TMC transcripts

To demonstrate that all predicted murine TMC genes are transcribed, we performed RT-PCR experiments. Primer pairs specific for each TMC transcript amplified products of predicted length and sequence (Figure 5).

Tmc1 and Tmc3 mRNAs were detectable in most neuronal organs and we also found expression in some non-neuronal organs. Tmc2 transcripts were only detectable in testis; we did not reveal by RT-PCR Tmc2 expression in cochlea. However, we were able to verify that Tmc2 is expressed in the cochlea by using an organ of Corti cDNA library as a template for PCR, which corroborates the results presented by Kurima et al. (2002).

We observed that mRNAs encoding Tmc5, Tmc6, Tmc4 and Tmc7 are expressed in most murine organs tested. Tmc8 mRNA is detectable in thymus and lung; we also found expression of Tmc8 mRNA in spleen (not shown).

Figure 1

The murine TMC protein family Murine (Mm) TMC amino acid sequences were aligned using the MultAlin sequence comparison algorithm. Positions that are conserved (identical or similar) in all three subfamilies are marked orange (present in at least 4 of the 8 sequences) and red (conserved in at least 6 of the 8 sequences). Conserved positions within individual TMC protein subfamilies are depicted in blue. Amino acids that were considered similar based on the physicochemical features of their side chains: L, I, V and M; F and Y; E, D, N and Q; T and S; K and R. The TMC motif is denoted with a green bar. Presumptive transmembrane regions are underlined. Indicated with bold letters is experimentally verified TMC sequence.

Table 1: TMC genes

Gene	Chromosome (strand) or framework	GenBank Accession number
Mm *Tmc1* 19(-)	AF417579, AY263155	
Mm *Tmc2*	2 (+)	AF417581, AY263156
Mm *Tmc3*	7 (+)	AY4263157
Mm *Tmc5*	7 (+)	AY4263159
Mm *Tmc6*	11 (-)	AY4263158
Mm *Tmc4*	7 (-)	AY4263162
Mm *Tmc7*	7 (-)	AY4263161
Mm *Tmc8*	11 (+)	AY4263160
Hs *TMC1*	9q21.13 (+)	AF417578
Hs *TMC2*	20p13 (+)	AF417580
Hs *TMC3*	15q23 (+)	AY4263163
Hs *TMC5*	16p12.3 (+)	AY4263164
Hs *TMC6* (*EVER1*)	17q25.3 (-)	AY057379, AY099356
Hs *TMC4*	19q13.42 (-)	AY4263166
Hs *TMC7*	16p12.3 (+)	AY4263165
Hs *TMC8* (*EVER2*)	17q25.3 (+)	AY057380, AY099358
Fr *Tmc2-rs1*	S2870	AY4263167
Fr *Tmc2-rs2*	S1411	AY4263168
Fr *Tmc3*	S2699	AY4263169
Fr *Tmc5*	S1110	AY4263172
Fr *Tmc6-rs1*	S1355	AY4263170
Fr *Tmc6-rs2*	S2029	AY4263171
Fr *Tmc4*	S3252	AY4263173
Fr *Tmc7*	S1110	AY4263174
Ce *TmcAh1*	B0416.1	NM_077168
Ce *TmcAh2*	T13G4.3	NM_075820
Ag *TmcAh*	2	AY4263175
Ag *TmcBh*	X	AY4263176
Ag *TmcCh*	3	AY4263177
Dm *TmcAh*	3L (CG3280)	NM_140079

We did not detect Tmc8 transcripts in any other organs investigated.

Our expression analysis results are corroborated by an analysis of applicable TMC ESTs obtained from GenBank http://www.ncbi.nlm.nih.gov/dbEST/ (Figure 5).

Non-mammalian vertebrate and invertebrate TMC genes and proteins

The high degree of similarity between the corresponding human and murine TMCs suggests a conserved function of these proteins. This role of TMC proteins may also be conserved among other species. We therefore decided to investigate the TMC genes of other vertebrates and invertebrates.

We identified eight TMC loci in the Japanese pufferfish (Torafugu, *Fugu rubripes*, Fr) genomic database http://fugu.hgmp.mrc.ac.uk/. Whereas the genome of *Fugu rubripes* is one order of magnitude shorter than the length of the human genome, the total number of genes is estimated to be approximately the same [8,9]. Because of the low homology of amino- and carboxyl-terminal sequences, we were not able to determine the complete coding sequences of the eight pufferfish TMC proteins unequivocally; nevertheless, we obtained sufficient sequence information of the central parts of the proteins bearing the transmembrane domains to classify the eight pufferfish TMCs into the three subfamilies. This subfamily assignment of individual TMC proteins was further substantiated by an analysis of the degree of conservation of intron positions within the *Fugu rubripes* TMC gene family (Figure 3B). The *Fugu rubripes* genome contains three TMC subfamilyA genes *Tmc2-rs1* (*Tmc2-related sequence1*), *Tmc2-rs2* (*Tmc2-related sequence2*), *Tmc3*, three TMC subfamilyB genes *Tmc5*, *Tmc6-rs1* (*Tmc6-related sequence1*), *Tmc6-rs2* (*Tmc6-related sequence2*), and two subfamilyC genes *Tmc4* and *Tmc7* (Table 1). The nomenclature of the pufferfish TMC genes and proteins is derived from the phylogenetic relation of the corresponding sequences with the mammalian TMCs (Figure 4). It is interesting that the pufferfish genome lacks orthologues of mammalian TMC1 and TMC8. *Fugu rubripes Tmc5* and *Tmc7* are

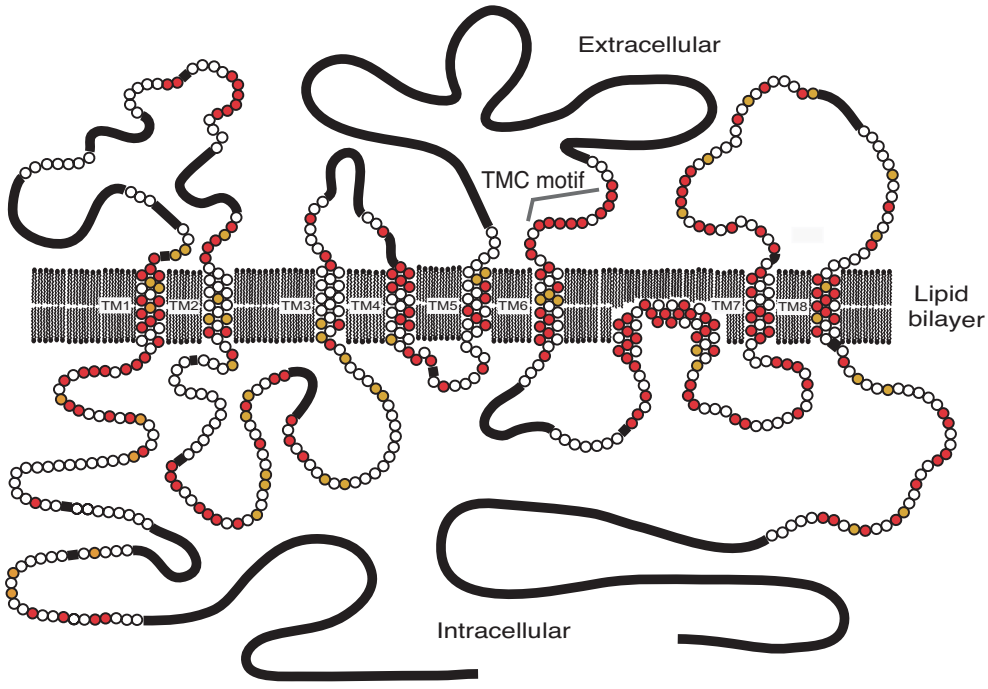

Figure 2
Proposed structure of TMC proteins Schematic illustration of the presumptive TMC protein topology. Each circle represents an amino acid. Amino acid positions conserved in at least 50% and 75% of murine TMC proteins are shaded orange and red, respectively. Black lines represent regions of variable length within the TMC family. The position of the TMC signature sequence motif is indicated.

clustered, equivalent to the clustering of their mammalian orthologues [7].

TMC genes also exist in invertebrates. In GenBank, we identified two mRNA sequences encoding *Caenorhabditis elegans* (Ce) TMC proteins. These mRNAs are transcribed from *TmcAh1* (*Tmc subfamily A homologue1*) and *TmcAh2* (*Tmc subfamily A homologue2*) (Table 1).

Whereas the *C. elegans* genome appears to lack TMC subfamily B and C genes, some insects have genes for all three TMC subfamilies. For example, the mosquito *Anopheles gambiae* has three TMC genes, *TmcAh* (*Tmc subfamily A homologue*), *TmcBh* (*Tmc subfamily B homologue*), and *TmcCh* (*Tmc subfamily C homologue*) (Figure 4, Table 1). We did not find deposited cDNA *TmcAh* and *TmcCh*, but *TmcBh* appears to be transcribed (ESTs BM645887, BM621478, BM605758, and BM636384). A search of the *Drosophila melanogaster* genome database revealed only a single TMC gene, *TmcAh*, (Figure 4, Table 1).

We did not find evidence for TMC genes in genomes and cDNA databases of yeast and plants.

Discussion

In an effort to define a novel gene family, we set out to identify genes related to *TMC1*, *TMC2*, *EVER1*, and *EVER2* [5–7]. We obtained the coding sequences of additional homologues in human and mouse, which form the TMC protein family (Figures 1, 2). We subdivided the TMC protein family into three subcategories A, B, and C. This subfamily-classification is based on two major observations. First, the sequence homologies among the different TMC protein sequences of individual species' do cluster into three groups (Figures 1, 4, and additional file 1). Second, our analysis of the organization of vertebrate TMC

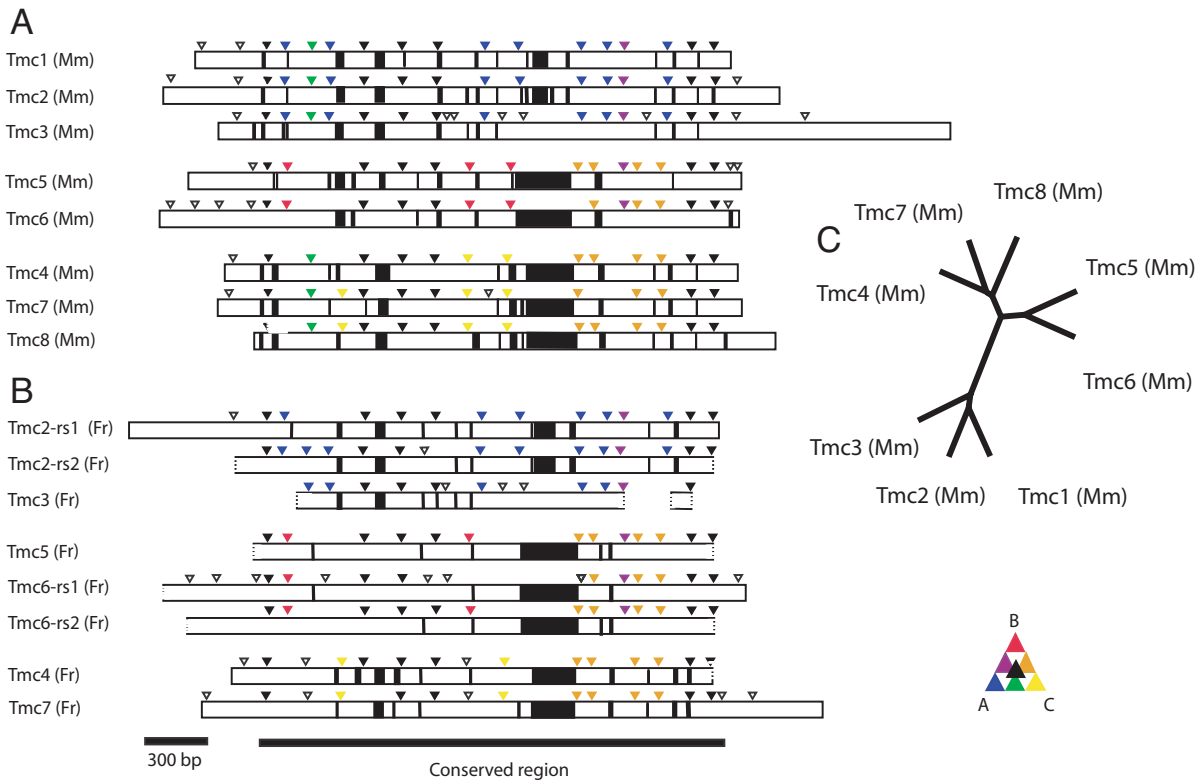

Figure 3

Comparison of murine and pufferfish TMC gene structures Comparison of murine (Mm)(A) and pufferfish (Fr)(B) TMC gene structures. To compare and visualize the gene structures of TMC genes, the predicted coding sequences (depicted here with white boxes) were aligned. Gaps were introduced when required for appropriate comparison (illustrated as black boxes). Triangles mark positions where the coding sequence is interrupted by introns in the genome (intron locations). The triangles are colour-coded to reflect the degree of conservation of the intron locations among TMC subfamilies A, B, and C. Positions of introns conserved in all members of the TMC family are denoted by black triangles; blue, red and yellow triangles label positions of introns conserved within individual TMC subfamilies; orange, magenta and green triangles denote intron positions shared by members of two subfamilies; intron positions unique to a single family member are indicated by open black triangles. (C) Phylogenetic tree drawn based on the similarity of the gene structures of murine TMC genes restricted to the conserved region of the TMC proteins. The corresponding matrix of distances is shown in additional file 1.

genes implied identical allocation of the individual TMCs (Figure 3 and additional file 1).

On the basis of comparative structural predictions of all mammalian TMCs we propose a basic topology of the generic TMC protein with variable intracellular amino- and carboxyl-termini, a conserved core with eight membrane-spanning domains, and some variability in the length of several intra- and extracellular loops (Figure 2). This prediction further refines the previously published TMC protein features [5,7]. The structural presumptions for the region flanked by TM6 and TM7 were somewhat ambiguous, despite the region's high conservation among all TMC proteins with a high proportion of apolar amino acids (Figure 1). We hypothesize that it is unlikely that some members of the TMC family display an atypical topology, thus we propose that the lipophilic intracellular loop between TM6 and TM7 bears some flexibility, which may enable this domain to integrate into the inner surface of the phospholipid bilayer or even to pass through the plasmamembrane (Figure 2).

The high amino acid sequence conservation of 75–96% identity among the individual human and mouse TMC proteins implies that mutations that alter the coding sequence in the corresponding genes are subjected to significant selective pressure; thus advocating that TMC proteins have important cellular roles. Mutations in *TMC1* are responsible for the autosomal dominant human hearing disorder DFNA36 and for the recessive

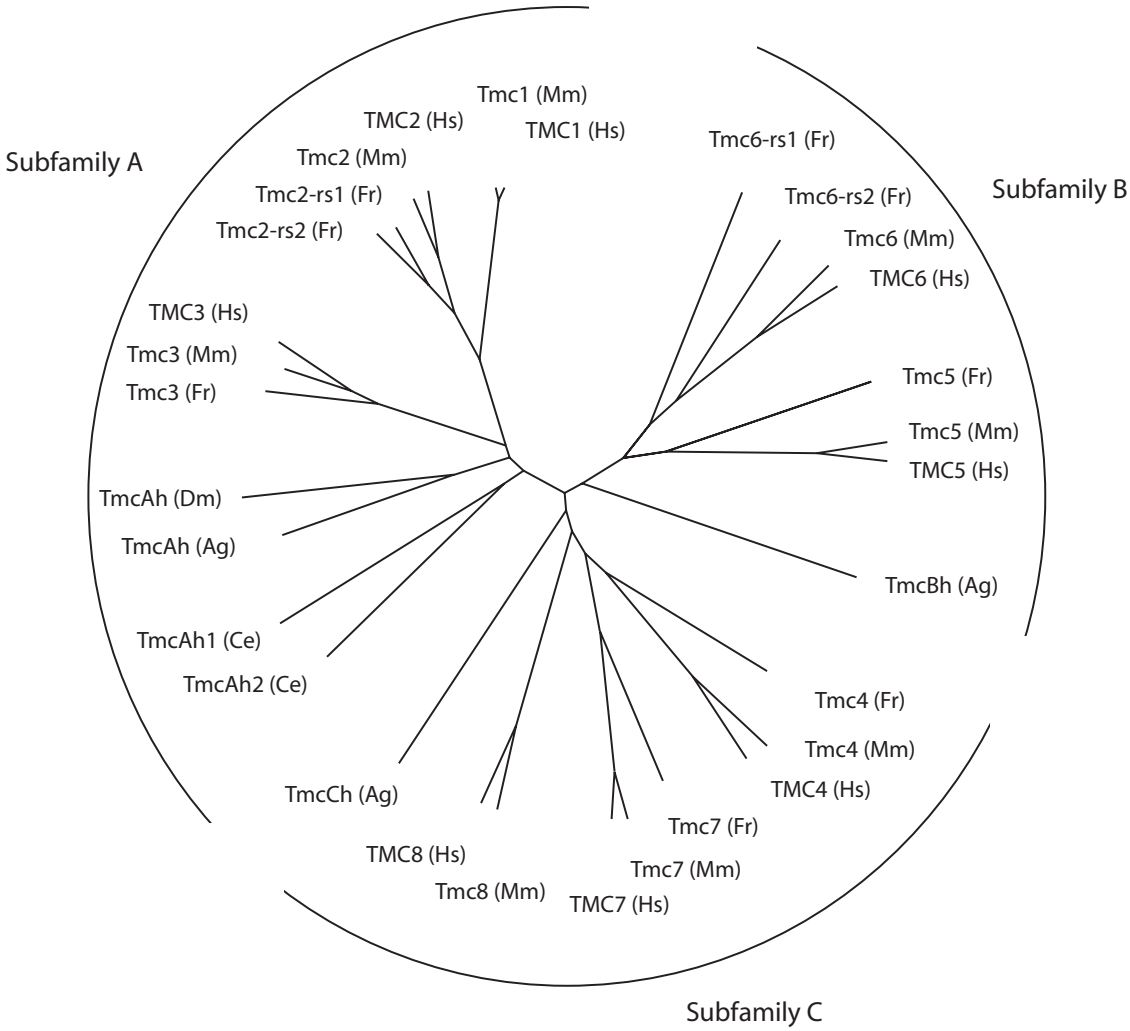

Figure 4
Evolutionary relationship of vertebrate and invertebrate TMC proteins Phylogenetic tree, based on comparison of TMC protein sequences, illustrating the evolutionary relationship of TMC proteins of *Homo sapiens* (Hs), *Mus musculus* (Mm), *Fugu rubripes* (Fr), *Drosophila melanogaster* (Dm), *Anopheles gambiae* (Ag), and *Caenorhabditis elegans* (Ce).

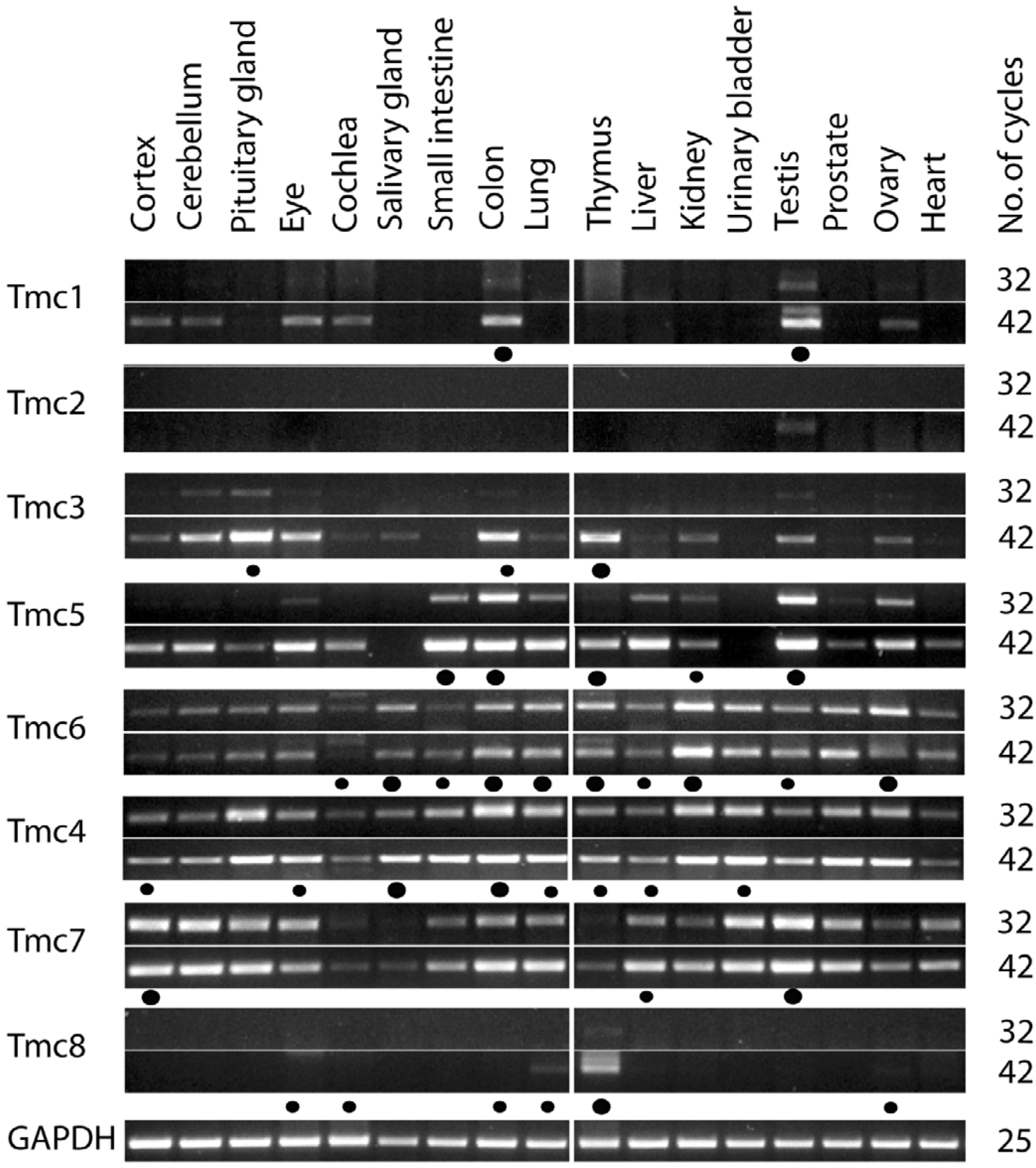

Figure 5
TMC expression analysis RT-PCR analysis of expression of TMC family members in mouse organs as judged by 32 and 42 cycles of amplification. Expression analysis of the ubiquitously expressed glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is shown for reference. Dots indicate that one or two (small dot) or three and more (large dots) ESTs representing the respective organ are present in public databases.

deafness DFNB7/11 [5]; the corresponding murine gene *Tmc1* also causes deafness in the mutant mouse strains *Beethoven* (*Bth*) and *deafness* (*dn*) [6]. *Bth*/*Bth* mice display altered potassium currents in early postnatal cochlear hair cells, in particular Ik,n and Ik,f currents appear to be depressed [10]. This observation led to the hypothesis that Tmc1 may participate directly or indirectly in regulating the permeability of potassium channels [10]. Because of the high level of conservation of the amino acid sequences among the eight mammalian TMC proteins (Figure 1, Additional file 1), we speculate that other TMC proteins may as well be modifiers of ion channels or transporters.

Our assembly of TMC protein sequences allowed us to analyze their phylogenetic relationships (Figure 4). Because our analysis did not reveal any TMC genes in yeast and plants, we conclude that TMC proteins are a specific trait of animals. The *C. elegans* genome harbors two genes encoding presumptive TMC subfamily A-like proteins, both of which branch off close to the center of the phylogenetic tree, which may indicate similarity with the primordial TMC protein sequence. It is interesting that some of the intron locations of the two *C. elegans* genes are conserved when compared with vertebrate TMC genes, in particular with the TMC subfamily A (Additional file 3).

It is likely that the TMC family has diversified into three subfamilies before the *Protostomia* and the *Deuterostomia* diverged because the genome of the mosquito *Anopheles gambiae* contains one putative member of each TMC subfamily. Interesting in this regard is that the fruit fly's genome only has a single TMC subfamily A-like gene. The vertebrate TMC gene family is diversified, with each subfamily represented by multiple members.

Conclusions

The recently identified genes encoding the cochlear transcripts for TMC1 and TMC2, and encoding the proteins EVER1 and EVER2, belong to a novel gene family that is conserved in animals. The TMC protein family has eight members in vertebrates and forms three subfamilies, A, B, and C. TMCs are proteins with a conserved core of eight membrane-spanning domains. Most murine TMC genes are widely expressed at relatively low transcription levels; some TMC genes, however, display more restricted expression patterns.

Methods

Database-aided and experimental assessment of TMC cDNA and protein sequences

We created and continuously refined contigs of members of novel TMC cDNAs by aligning fragmented sequence information obtained from public and commercial EST databases (GenBank/NCBI – http://www.ncbi.nlm.nih.gov, Celera Discovery System – http://www.celeradiscoverysystem.com). We started this procedure by using sequence fragments homologous to the known mammalian TMC proteins. Individual fragments were assembled to longer contigs using multiple alignment tools such as ClustalW [11] http://www.es.embnet.org/Doc/phylodendron/clustal-form.html or Multalin [12] http://prodes.toulouse.inra.fr/multalin/multalin.html. Our sequence determination was continuously refined with the results of repeated BlastN and tBlastN database searches with the updated contigs, and finally we extended our search by using genomic databases. The majority of Blast searches were conducted with the default parameter settings. With this strategy were able to generate presumptive "framework" contigs for the in databases less abundantly represented TMC genes. Because of the inter-species conservation of protein-coding sequences we were able to predict the presumptive DNA sequence encoding amino- and carboxyl-termini of various TMC proteins by comparing human and murine genomic TMC sequence.

We constantly refined individual TMC sequences by using genomic sequence information and comparing gene structures. In addition to iterative database searches, we also took into account the results of gene prediction algorithms using genomic DNA harboring TMC genes as inputs (GeneScan [13], GenomeScan [14] http://genes.mit.edu). Some invertebrate TMC genes were already annotated as hypothetical proteins. We used human and murine genome databases (GenBank/NCBI – http://www.ncbi.nlm.nih.gov) to determine the chromosomal locations of all mammalian TMC genes.

In a second phase, we used polymerase chain reaction and 5' and 3' rapid amplification of cDNA ends (RACE) experiments to obtain experimental proof for our cDNA predictions and to determine murine TMC sequence that was only ambiguously predicted.

Secondary structure predictions were done with PSORTII [15] http://psort.ims.u-tokyo.ac.jp. The phylogenetic distances between various TMC protein sequences were calculated with ClustalW [11] http://www.es.embnet.org/Doc/phylodendron/clustal-form.html and the resulting distance matrix Additional file: 1 (A) was visualized as a tree diagram with centered node position using the Phylodendron phylogenetic tree printer available at http://iubio.bio.indiana.edu/treeapp.

Evolutionary distance between two gene structures was defined as the reciprocal of the ratio of conserved intron locations with respect to the coding sequences of the TMC genes. The distances were determined for each pair of murine TMC genes (additional file 1 (B)). We compared

the relationship of the TMC gene structures by performing cluster analysis of distances using the kitsch algorithm of the PHYLIP Phylogeny Inference Package [16] http://bioweb.pasteur.fr/seqanal/phylogeny/intro-uk.html. The results of the cluster analysis were visualized as a tree diagram with centred node position using the Phylodendron phylogenetic tree printer.

Accession numbers for all TMC cDNAs reported in this paper are listed in Table 1.

mRNA expression analysis

Organs we dissected from eight-week-old mice and quickly frozen in liquid nitrogen. Specifically, we used cortex, cerebellum, eye, cochlea, pituitary gland, salivary gland, small intestine, colon, lung, thymus, liver, kidney, heart, urinary bladder, testis, prostate, and ovary. We also used in a control experiment aimed to verify Tmc2 expression in the cochlea, plasmid DNA obtained from an organ of Corti cDNA library that was kindly provided by Dr. B. Kachar, NIDCD/NIH. Total RNA was extracted using RNeasy Mini and Midi kits (Qiagen, Valencia, CA). For RT-PCR, 1 μg of total RNA was reverse-transcribed into cDNA using SuperscriptII reverse transcriptase (Invitrogen, Carlsbad, CA). One tenth of the resulting cDNA was used as template for each PCR amplification with primers specific for each member of the TMC gene family (Tmc1: 5'-GAAACAATGGTGGGGCAGGAA and 5'-GATGGCGGGAGGGAGACGAT, Tmc2: 5'-CCCCGGCCACCACCACAC and 5'-GAAGGGGCCACAGTCAAACGAG, Tmc3: 5'-TCTGGACAAAGTGAACAGCATG and 5'-GAACATTGAATGCTGGCAGACA, Tmc5: 5'-GGGCGGGTTATTTTGGGGACAC and 5'-ACAACAAAGGGCAGCAACAACACC, Tmc6: 5'-CTATCACTGGCTGGGCCGCAGG and 5'-GCAGCCGGAGAATCTGTACTGC, Tmc4: 5'-GTGGAGTCAGATCACGTGTGGG and 5'-CAGGCAGAACAGTATGAGGAAC, Tmc7: 5'-GGGCTCCAAGATCACATCCTGT and 5'-CGAATTTCAAGGTGGCAATGGC, and Tmc8: 5'-GGGTCAGACGGTGCTGTGCATA and 5'-AGAGGAAGACACTGTTGAGTAG). Care was taken to select primer pairs that amplify TMC transcripts encoded in multiple exons to rule out the amplification of genomic DNA contamination. Specific products obtained with each primer pair from various organs and in several independent experiments were verified by sequencing. To estimate the quality of the template and to compare expression levels, we amplified mouse glyceraldehyde 3-phosphate dehydrogenase (GAPDH), a ubiquitously expressed abundant transcript with specific primers (5'-AACGGGAAGCCCATCACC and 5'-CAGCCTTGGCAGCACCA). Cycling parameters were 94°C for 3 min, 8 cycles of (94°C20 s, 60°C20 s -1°C/cycle, 68°C45 s), followed by 17, 24 or 34 additional cycles of (94°C30 s, 52°C30 s, 68°C45 s) and 68°C for 5 min.

Authors' contributions

KG carried out iterative database searches to identify mammalian members of the TMC family, he did PCR and RACE experiments to obtain TMC cDNA sequence, and he performed part of the transcript expression analysis. HM identified non-mammalian and invertebrate TMC genes, he assembled their coding sequences from public databases, and he performed part of the transcript expression analysis. KG and HM drafted the manuscript. All authors conceived of the study as result of discussions during the group's laboratory meeting. SH participated in design and coordination of the study and finalized the manuscript, which has been read and approved by all authors.

Additional material

Additional File 1

Identity matrices of murine TMCs. (A) Percentage of amino acid residue identity among individual murine TMC proteins. (B) Evolutionary distance ratios of the gene structures of individual murine TMC genes. We calculated the ratio of conserved intron locations within the conserved region of the eight murine Tmc*s (see Figure 2). The reciprocal values of the distance ratios represent the degree of conserved intron locations in percent. The three subfamilies are shaded in green (TMC Subfamily A), red (TMC Subfamily B), and blue (TMC Subfamily C).*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-4-24-S1.pdf]

Additional File 2

TMC signature sequence motif consensus sequence. Comparison of the conserved TMC signature sequence motif among vertebrate and invertebrate TMC proteins. The consensus sequence is based on a 100% conservation of the three leading amino acid residues. All other capitalized amino acids are conserved in more than 70% of all TMCs, and lower case letters indicate the most probable amino acid residue.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-4-24-S2.pdf]

Additional File 3

Comparison of invertebrate TMC gene structures. Comparison of the intron locations projected onto a schematic comparison of invertebrate TMC proteins. A representation of the conserved intron locations of the vertebrate TMC subfamilyA is displayed as a reference. For details on the color-coding of intron locations, see figure 3, with the addition that open blue triangles refer to intron positions conserved within invertebrate cladi*. Each black box indicates a gap to adjust the alignment.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-4-24-S3.pdf]

Acknowledgements

We thank Drs. E. Corrales and W.F. Sewell for comments on the manuscript. In particular, we thank Drs. K. Kurima and A. Griffith for sharing the HUGO and mouse nomenclature committee-approved TMC nomenclature, which has been submitted elsewhere by their research team in a related context. This work was supported by grant DC04563 to SH from

the NIH/NIDCD and in part by Basil O'Connor Starter Scholar Research Award Grant No. 5-FY00-528 from the March of Dimes Birth Defects Foundation.

References

1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA and Holt RA *et al.*: **The sequence of the human genome.** *Science* 2001, **291:**1304-51.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M and FitzHugh W *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.
3. Hogenesch JB, Ching KA, Batalov S, Su AI, Walker JR, Zhou Y, Kay SA, Schultz PG and Cooke MP: **A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes.** *Cell* 2001, **106:**413-5.
4. Li S, Liao J, Cutler G, Hoey T, Hogenesch JB, Cooke MP, Schultz PG and Ling XB: **Comparative analysis of human genome assemblies reveals genome-level differences.** *Genomics* 2002, **80:**138-9.
5. Kurima K, Peters LM, Yang Y, Riazuddin S, Ahmed ZM, Naz S, Arnaud D, Drury S, Mo J and Makishima T *et al.*: **Dominant and recessive deafness caused by mutations of a novel gene, TMC1, required for cochlear hair-cell function.** *Nat Genet* 2002, **30:**277-84.
6. Vreugde S, Erven A, Kros CJ, Marcotti W, Fuchs H, Kurima K, Wilcox ER, Friedman TB, Griffith AJ and Balling R *et al.*: **Beethoven, a mouse model for dominant, progressive hearing loss DFNA36.** *Nat Genet* 2002, **30:**257-8.
7. Ramoz N, Rueda LA, Bouadjar B, Montoya LS, Orth G and Favre M: **Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis.** *Nat Genet* 2002, **32:**579-81.
8. Hedges SB and Kumar S: **Genomics. Vertebrate genomes compared.** *Science* 2002, **297:**1283-5.
9. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S and Smit A *et al.*: **Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.** *Science* 2002, **297:**1301-10.
10. Steel KP, Marcotti W, Erven A, Johnson SL and Kros CJ: **Cochlear function in Beethoven mouse mutants.** In: *Association of Research of Otolaryngology Winter Meeting Daytona Beach, Florida*; 2003:1063.
11. Thompson JD, Higgins DG and Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-80.
12. Corpet F: **Multiple sequence alignment with hierarchical clustering.** *Nucleic Acids Res* 1988, **16:**10881-90.
13. Burge C and Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268:**78-94.
14. Yeh RF, Lim LP and Burge CB: **Computational inference of homologous gene structures in the human genome.** *Genome Res* 2001, **11:**803-16.
15. Nakai K and Horton P: **PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization.** *Trends Biochem Sci* 1999, **24:**34-6.
16. Felsenstein J: **PHYLIP – Phylogeny Interference Package (Version 3.2).** *Cladistics* 1989, **5:**164-6.