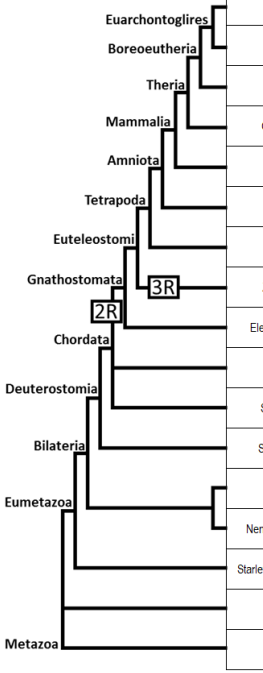
The American Journal of Human Genetics, Volume *104*

Supplemental Data

Pathogenic Variants in *GPC4* Cause Keipert Syndrome

David J. Amor, Sarah E.M. Stephenson, Mirna Mustapha, Martin A. Mensah, Charlotte W. Ockeloen, Wei Shern Lee, Rick M. Tankard, Dean G. Phelan, Marwan Shinawi, Arjan P.M. de Brouwer, Rolph Pfundt, Cari Dowling, Tomi L. Toler, V. Reid Sutton, Emanuele Agolini, Martina Rinelli, Rossella Capolino, Diego Martinelli, Giuseppe Zampino, Miroslav Dumić, William Reardon, Charles Shaw-Smith, Richard J. Leventer, Martin B. Delatycki, Tjitske Kleefstra, Stefan Mundlos, Geert Mortier, Melanie Bahlo, Nicola J. Allen, and Paul J. Lockhart



Common Name	Species Name	Species Abbreviation	GPC Subfamily I				GPC Subfamily II			
			GPC1	GPC2	GPC4	GPC6	GPC3	GPC5	GPC5-like	GPC6-like
Human	*Homo sapiens*	Hsap	NP_002072.2	NP_689955.1	NP_001439.2	NP_005699.1	NP_004475.1	NP_004457.1		
Mouse	*Mus musculus*	Mmus	NP_057905.1	NP_766000.1	NP_032176.2	NP_001073313.1	NP_057906.2	NP_780709.1		
Cow	*Bos taurus*	Btau	XP_010802238.1	NP_001180109.1	NP_001192713.1	NP_001075199.1*	NP_001030540.1	NP_001095540.1 & BF074246		
Opossum	*Monodelphis domestica*	Mdom	XP_001377925.1	XP_001372022.1	ENSMODT00000019446	XP_001366609.1	XP_007507267.1	XP_007501459.1^		
Platypus	*Ornithorhynchus anatinus*	Oana	XP_003430024.1**	XP_007658157.1**	XP_001514704.1	ENSOANT00000023495*	NP_057906.2	XP_007666192.1 & ENSOANT00000023499		
Chicken	*Gallus gallus*	Ggal	NP_001291989.1	NA	ENSGALT00000009812^	XP_416989.4^	XP_001232892.1	XP_004938641.1	XP_004937034.1	
Frog	*Xenopus (Silurana) tropicalis*	Xtro	XP_001377925.1	XP_012808830.1	ENSMODT00000019446	XP_012812967.1 & CX814199	NP_057906.2	NA	XP_004917861.1	
Zebrafish	*Danio Rerio*	Drer	NP_001032450.1 (a) / NP_957314.1 (b)	XP_012808830.1	NP_571935.2	NP_001103189.1	NP_001292547.1	NP_001289165.1	NP_001289150.1 / XP_005171488.2	NP_001274470.1
Elephant shark	*Callorhinchus milii*	Cmil	XP_007907526.1	NA	XP_007892858.1^	XP_007889010.1* & XP_007889009.1* & XP_007889008.1*	XP_007890697.1	XP_007895606.1 & JW872419.1	XP_007907130.1	
Lancelet	*Branchiostoma floridae*	Bflo	XP_002593274.1*				XP_002593271.1 & XP_002593272.1*			
Sea squirt	*Ciona intestinalis*	Cint	XP_009861100.1				XP_002128087.1			
Sea Urchin	*Strongylocentrotus purpuratus*	Spur	NP_001138965.1				NP_001138966.1			
Fruit fly	*Drosophila melanogaster*	Dmel	NP_524071.2				NP_523983.1			
Nematode worm	*Caenorhabditis elegans*	Cele	NA				NP_510582.1			
Starlet sea anemone	*Nematostella vectensis*	Nvec	FC228760.1 & FC228201.1 & FC226842.1				XP_001624312.1		GPC-like	
NA	*Trichoplax adhaerens*	Tadh	XP_002117928.1^ & XP_002117927.1^						XP_002117929.1	
Sponge	*Amphimedon queenslandica*	Aque	XP_003386976.1						XP_003386974.1	

Figure S1: Glypican sequences and taxonomic relationship of species used in this study.

A common tree of species used in this study was generated using the Taxonomy Browser of National Center for Biotechnology Information (NCBI) taxonomy database (http://www.ncbi.nlm.nih.gov/taxonomy). The taxonomic classifications of transitional branches are shown on the left of the tree. The evolutionary position of the two rounds of whole genome duplication is indicated (boxed 2R), and the teleost specific third whole genome duplication is indicated (boxed 3R). The table shown in conjunction with the tree indicates the common name of organism, the full scientific name and the species abbreviation used in this study. The accession number of sequences used in phylogenetic analysis are indicated. Sequences have been assigned to glypican (GPC) subfamily I or subfamily II based on phylogenetic analysis. NA; interrogation of whole genome shotgun sequences (WGS) and Expressed sequence tag (EST) databases did not identify sequences, which may reflect incomplete coverage or gene loss. *database sequence was modified using genomic information. ^incomplete sequence. Sequences are available on request.

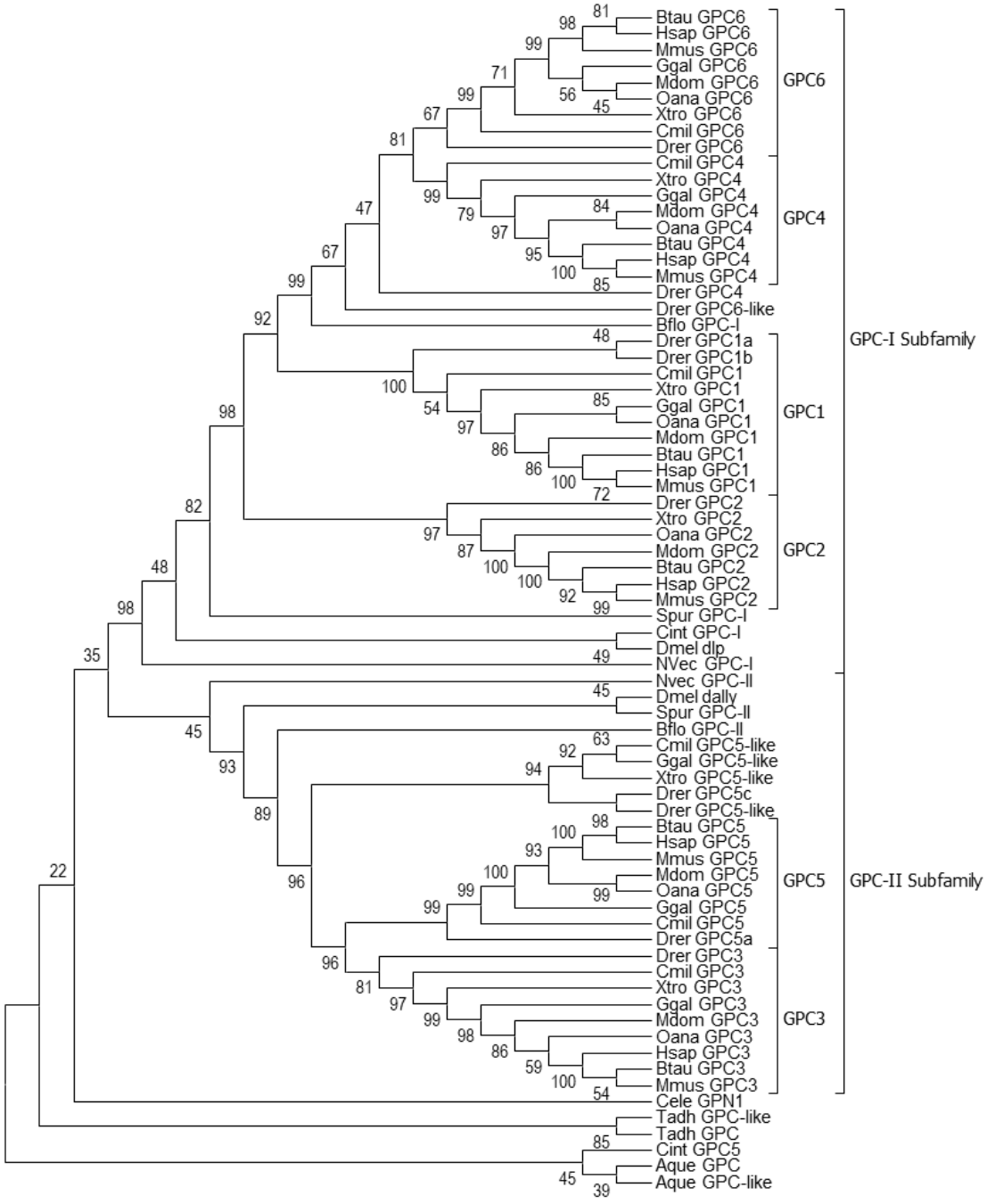


Figure S2: Molecular phylogenetic analysis of the glypican family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix model in MEGA6. The tree with the highest log likelihood is shown. The percentage of trees from 1000 replicates in which the associated taxa clustered together is shown next to the branches. Species abbreviations are used in this table, the full species and common name of organisms and the accession of sequences is found in Figure S1.

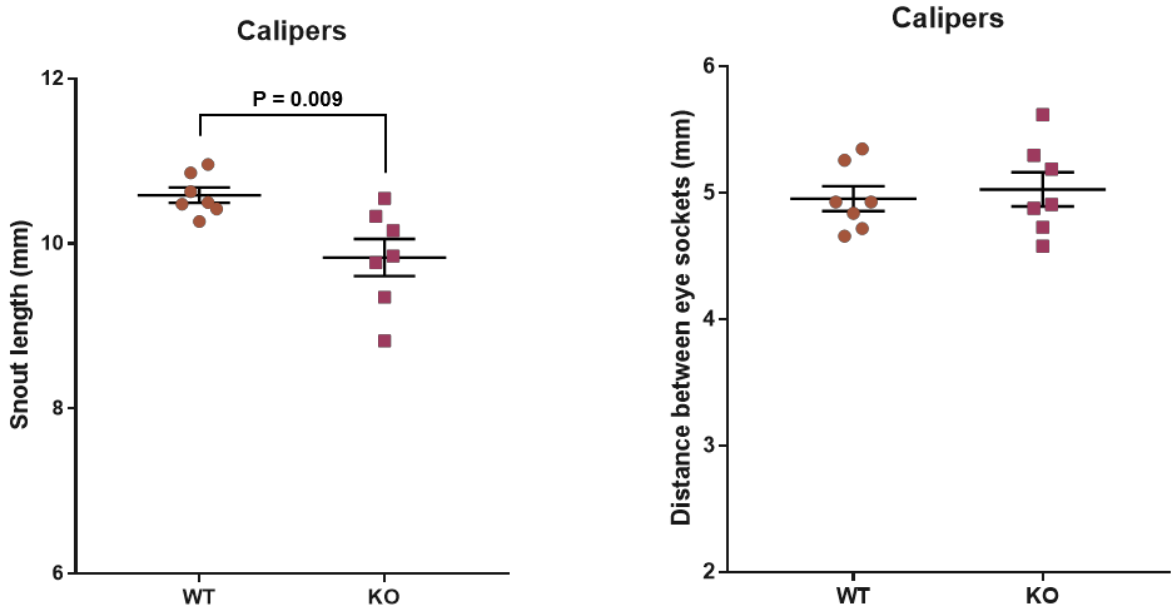


Figure S3: *Gpc*4 KO mice exhibit nasodigital deficits.

Manual measurement of skull morphology utilizing precision digital calipers demonstrated *Gpc4* KO mice have significantly shorter snouts, measured from the tip of the snout to the cranial aspect. However, there is no difference in the intra-orbital distance in *Gpc4* KO mice compared to WT. N= 7 animals per group, mean±SEM, statistical comparisons were made using a 2-tailed t-test.

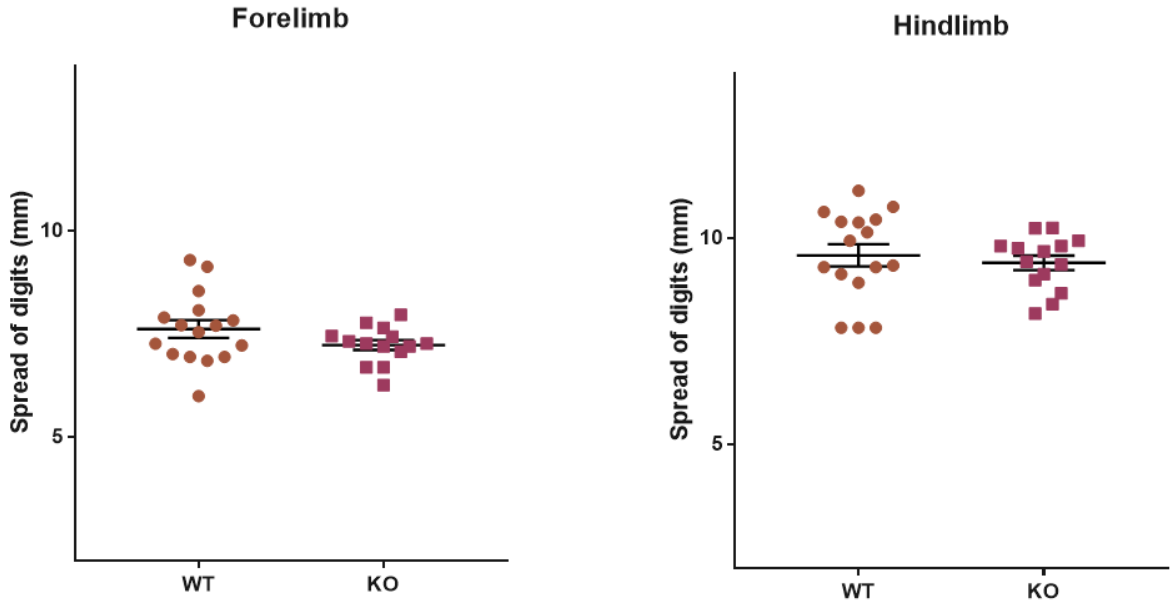


Figure S4: *Gpc4* KO mice do not display alterations in toe spread.

Images of the fore and hind paws of WT and KO mice were captured and analysed using GaitScan software. The tip of each phalanx was assigned a marker and the distance between the two outermost phalanges was determined. Left and right fore or hind paw measurements were combined, N=7 animals per group, mean±SEM, statistical comparisons were made using a 2-tailed t-test.

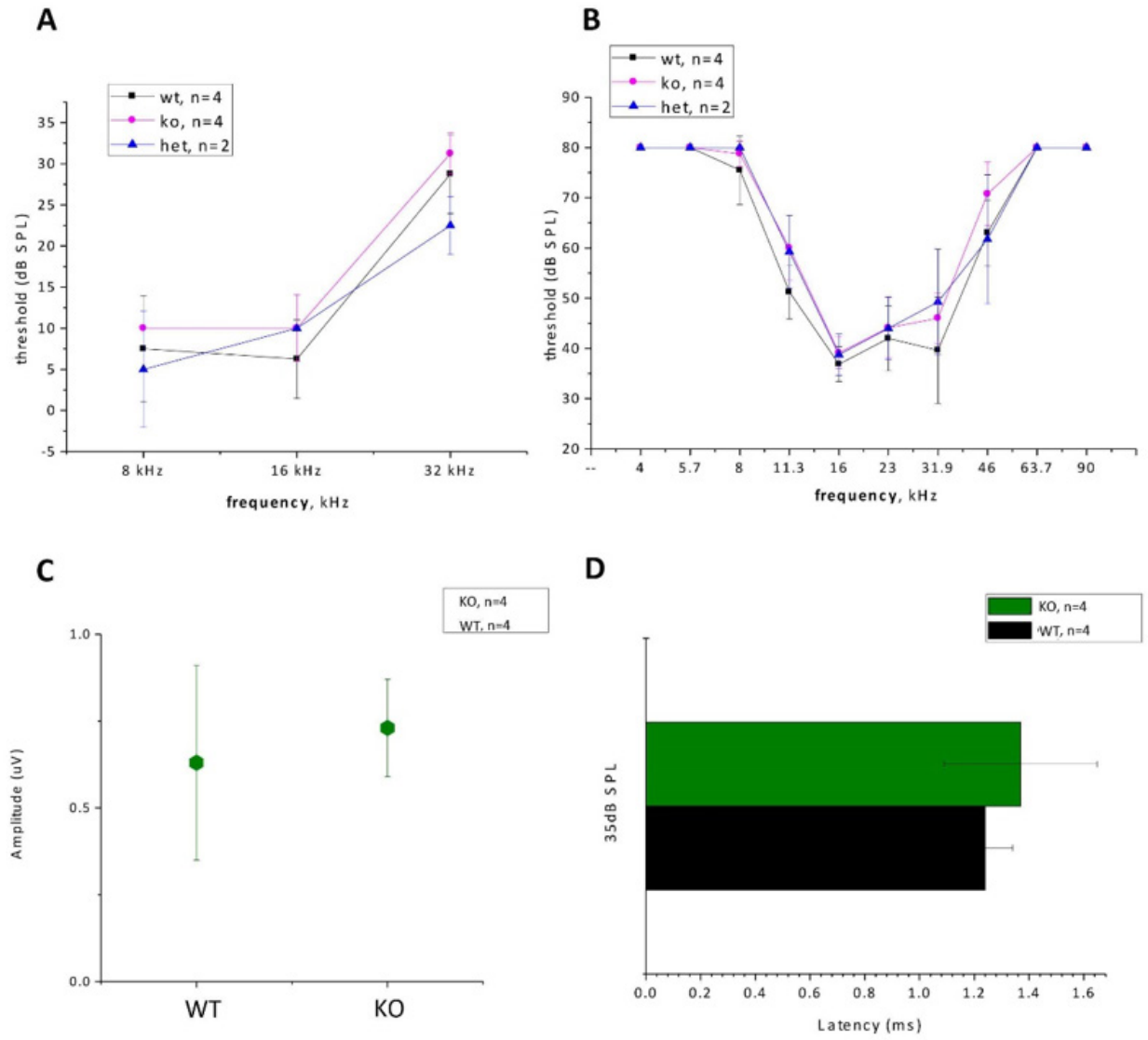


Figure S5: Assessment of auditory function in *Gpc4* KO mice by ABR and DPOAE. ABR (A) and DPOAE (B) thresholds are shown for mice at P18 of age. There were no significant differences in the thresholds at any frequency that were recorded in *Gpc4* KO mice compared to WT controls. ABR waveform amplitude of peak 1 (C) and latency (D) analysis measured at the 16 kHz frequency stimulus and at 35 dB demonstrated no significant differences between *Gpc4* KO and WT controls. N=4 animals per group, mean±SD, statistical comparisons were made using a 2-tailed t-test. Similar results were obtained at ages P21 and P30.

Gene	Position (hg19)	Variation	PolyPhen-2	SIFT	CADD score
MUM1L1	X:105450829	c.1404G>A p.(Trp468*)	-	-	11.8
GPC4	X:132437050	c.1516C>T p.(Gln506*)	-	-	12
AVPR2	X:153171703	c.743G>A p.(Arg248His)	benign	tolerated	0.3

Table S1: List of prioritized variants identified by exome sequencing within the linkage region of family KS1.

Exome sequencing, filtering and prioritization identified four potential candidate variants that were absent or present with a MAF<0.01 in population databases. Variants were subsequently ranked by potential functional impact using Polyphen2, SIFT and CADD. Gene reference sequences utilized were NM_152423.4 (*MUM1L1*), NM_001448.2 (*GPC4*) and NM_000054.5 (*AVPR2*).

ONLINE METHODS

Samples

Samples from the Keipert syndrome family KS1 were collected after receiving institutional ethics approval (Project 28097, Royal Children's Hospital, Melbourne, Australia and written informed consent from participants. The clinical description of family KS1 was reported previously and linkage to Xq22.2-Xq28 established.[1] Genomic DNA was extracted from whole blood using the BACC DNA extraction kit (GE Healthcare Life Sciences, Uppsala, Sweden) according to the manufacturer's protocols. Five additional families (KS2-KS6) including seven affected individuals were later enrolled with informed consent to this study (Project 2011/155, CMO Arnhem Nijmegen, Radboud University Medical Center, Nijmegen, the Netherlands; Project ID: EA2/087/15, Charité - Universitätsmedizin Berlin, the clinical diagnostic program at Washington University in St. Louis and the institutional scientific board of Bambino Gesù Children Hospital) through the web-based tool GeneMatcher or by direct contact with the managing clinician.[2] To assess the X-chromosome inactivation profile, we performed a methylation specific amplification/digestion test in the (CAG)n repeat of the androgen receptor (AR) gene, essentially as described previously.[3] Allelic ratios between 80-89% were considered as moderate skewing, between 90-100 % as strong skewing. Ratio values between 50 -79% are considered as a normal X-inactivation pattern.

Exome Sequencing and variant identification

To identify the underlying genetic cause of Keipert syndrome, gDNA from family KS1 was analyzed by whole exome capture and massively parallel sequencing (MPS) using the TruSeq capture kit and an Illumina HiSeq 2000 (Axeq Technologies, Macrogen, Seoul, Korea). Reads were aligned to the reference genome (UCSC hg19) by Novoalign (version 2.07.09), analysed with the SAMtools package (version 0.1.15) mpileup tool and variants were annotated with ANNOVAR[4] to the UCSC Known Genes[5] and any present in GnomAD with a MAF≥0.01 were excluded. The functional impact of variants on the protein was predicted with PolyPhen2,[6] SIFT[7] and CADD score[8] prior to validation by standard PCR amplification and Sanger sequence analysis. The reference sequences for *GPC4* utilized were NM_001448.2 and NP_001439.2. The ClinVar accession numbers for the GPC4 variants reported in this paper are

RCV000659264.1, RCV000659265.1, RCV000659266.1, RCV000659267.1 and searchable submission SUB5135546 .

X-inactivation analysis

To assess the X-chromosome inactivation profile of carrier females, we performed methylation specific analysis of the (CAG)n repeat of the androgen receptor gene. In this assay differential methylation of the X chromosomes is quantitated by methylation sensitive digestion of genomic DNA. Genomic DNA was digested with BamH1 or BamH1 and the methylation sensitive HpaII, amplified by PCR using a fluorescent labelled primer and analysed on a capillary array. The relative intensity of each allele is calculated as area under the peak and compared to the BamH1 digest alone. When analysed with BamH1 and the methylation sensitive HpaII, only one allele is digested, demonstrating skewed X-inactivation. Allelic ratios between 50 -79% are considered as a normal X-inactivation pattern and >90% as strong skewing.

Mouse studies

The mouse strain used for this research project, B6;129S5-Gpc4tm1Lex/Mmucd, identification number 032331-UCD, was obtained from the Mutant Mouse Regional Resource Center, a NCRR-NIH funded strain repository, and was donated to the MMRRC by Genentech, Inc. Glypican 4 null (male) mice have exon 3 removed, and lack Gpc4 mRNA and protein, were generated as previously described.[9; 10] For bone analysis mice were crossed to C57Bl6/J for 7+ generations. Experiments were conducted on littermate male mice (Gpc4 is on the X chromosome), *Gpc4* +/y (WT) and *Gpc4* -/y (KO) and were approved by the Salk Institute Institutional Animal Care and Use Committee.

Mouse skull measurement

Littermate male *Gpc4* KO and WT mice from 16 to 20 weeks were used for analysis. Mice were euthanized by overdose of carbon dioxide, and heads were removed and submerged in 10% neutral buffered formalin for one week at 4°C to ensure fixation of bone. To access the bone, external tissue was dissected away using scissors. Two sets of measurements were taken using an instant readout precision digital caliper, in mm (Grobet USA, Item # 35.180): 1) for snout length, from the cranial aspect to the tip of the snout, 2) for intra-orbital distance, between the medial-

edge of each orbit. In addition, X-rays were taken of fixed skulls using a CR 7 Digital Dental X-Ray imaging for vet dentistry, with a DPI Resolution of 505/505. A two-tailed t-test was used to assess if difference in measurements between the two groups were statistically significant.

Mouse Footprint Analysis

Littermate male *Gpc4* KO and WT mice from 18 to 26 weeks were used for analysis. Mice were placed on a treadmill (Columbus) and videos were captured of their paws as they walked. GaitScan software (Clever Systems) was used to analyze both fore and hind paws. Measurement output is in mm. A minimum of five images were measured per paw per mouse, then averaged. The tip of each phalanx (1-5) and the base of the paw (B) were assigned a marker. Because the forepaws have only four phalanges, the #3 marker is an arbitrary mark between phalange #2 and 4. Print Length measures were taken from marker B to #3 while print width measures were taken from marker #1 to #5. A two-tailed t-test was used to assess if differences in measurements between the two groups were statistically significant.

Mouse hearing analysis

Hearing experiments were approved by Stanford University Institutional Animal Care and Use Committee. Auditory brainstem responses (ABR) recordings were conducted in a sound-attenuating room at the Auditory Core for the Department of Otolaryngology, Stanford University. Mice were anesthetized with ketamine (100 mg/kg) and xylazine (10 mg/kg), injected intraperitoneally. Animal body temperature was maintained at 37 °C for all recordings. Stimulus presentation, ABR acquisition, equipment control and data management were coordinated using the computerized intelligent hearing system (Intelligent Hearing Systems, Miami, FL, USA). Acoustic stimuli were delivered to the ear canals via plastic tubes channeled to the speaker at 8, 16, and 32 kHz. Sound levels were incremented in five dB steps from 10-20 dB below threshold to 80 dB (for 8 and 16 kHz) or 100 dB (for 32 kHz). Threshold for ABR was defined as the lowest stimulus level at which repetitive waves I and V could be identified in the response waveform. The amplitude analysis was done by peak-to-peak measurement of the ABR waveform and latency was calculated as time delay from the onset of the stimulus (0 ms) until the occurrence of the ABR response peak. ABR waveform amplitude and latency analysis was performed at the 16 kHz frequency at 35 dB. To test the function of outer hair cells we measured

Distortion-Product Otoacoustic Emission (DPOAEs) as previously described,[11] using the computerized National Hearing Instruments conducted in a sound-attenuating room at the Auditory Core for Department of Otolaryngology, Stanford University. DPOAE thresholds were calculated by interpolating the data and identifying when the signal was >−5 dB sound pressure level (SPL) and greater than two standard deviations above the noise floor. If no DPOAE response was detected even at our equipment limits of 80 dB SPL, we arbitrarily defined the threshold to be 80 dB. A two-tailed t-test was used to assess if differences in measurements between the two groups were statistically significant.

Molecular evolution of the Glypican family

Human glypican NCBI Reference Sequences were used to identify homologues. The organisms interrogated and sequence accession numbers analysed are indicated in Figure S1. Amino acid sequences were aligned using MUSCLE and the evolution of the glypican family was inferred using the Maximum Likelihood method based on a the Jones-Taylor-Thornton model using Molecular Evolutionary Genetics Analysis (MEGA) software version 6.[12] The amino acid identity and similarity of full length sequences was determined by pairwise global alignment using Matrix Global Alignment Tool (MatGat) version 2.02 with the PAM250 alignment substitution matrix.[12]

REFERENCES

1. Amor, D.J., Dahl, H.H., Bahlo, M., and Bankier, A. (2007). Keipert syndrome (Nasodigitoacoustic syndrome) is X-linked and maps to Xq22.2-Xq28. Am J Med Genet A 143A, 2236-2241.
2. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36, 928-930.
3. Spath, M.A., Nillesen, W.N., Smits, A.P., Feuth, T.B., Braat, D.D., van Kessel, A.G., and Yntema, H.G. (2010). X chromosome inactivation does not define the development of premature ovarian failure in fragile X premutation carriers. Am J Med Genet A 152A, 387-393.
4. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164.

5. Hsu, F., Kent, W.J., Clawson, H., Kuhn, R.M., Diekhans, M., and Haussler, D. (2006). The UCSC Known Genes. Bioinformatics 22, 1036-1046.
6. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-249.
7. Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4, 1073-1081.
8. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46, 310-315.
9. Allen, N.J., Bennett, M.L., Foo, L.C., Wang, G.X., Chakraborty, C., Smith, S.J., and Barres, B.A. (2012). Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors. Nature 486, 410-414.
10. Tang, T., Li, L., Tang, J., Li, Y., Lin, W.Y., Martin, F., Grant, D., Solloway, M., Parker, L., Ye, W., et al. (2010). A mouse knockout library for secreted and transmembrane proteins. Nat Biotechnol 28, 749-755.
11. Xia, A., Visosky, A.M., Cho, J.H., Tsai, M.J., Pereira, F.A., and Oghalai, J.S. (2007). Altered traveling wave propagation and reduced endocochlear potential associated with cochlear dysplasia in the BETA2/NeuroD1 null mouse. J Assoc Res Otolaryngol 8, 447-463.
12. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30, 2725-2729.
13. Campanella, J.J., Bitincka, L., and Smalley, J. (2003). MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4, 29.