Pathogenic Variants in *GPC4* Cause Keipert Syndrome

David J. Amor,[1,2] Sarah E.M. Stephenson,[1,2] Mirna Mustapha,[3] Martin A. Mensah,[4] Charlotte W. Ockeloen,[5] Wei Shern Lee,[1,2] Rick M. Tankard,[6,7] Dean G. Phelan,[1,2] Marwan Shinawi,[8] Arjan P.M. de Brouwer,[9] Rolph Pfundt,[5] Cari Dowling,[10] Tomi L. Toler,[8] V. Reid Sutton,[11,12] Emanuele Agolini,[13] Martina Rinelli,[13] Rossella Capolino,[14] Diego Martinelli,[15] Giuseppe Zampino,[16] Miroslav Dumić,[17] William Reardon,[18] Charles Shaw-Smith,[19] Richard J. Leventer,[1,2,20] Martin B. Delatycki,[1,2] Tjitske Kleefstra,[9] Stefan Mundlos,[4] Geert Mortier,[21] Melanie Bahlo,[6,7] Nicola J. Allen,[10] and Paul J. Lockhart[1,2,*]

Glypicans are a family of cell-surface heparan sulfate proteoglycans that regulate growth-factor signaling during development and are thought to play a role in the regulation of morphogenesis. Whole-exome sequencing of the Australian family that defined Keipert syndrome (nasodigitoacoustic syndrome) identified a hemizygous truncating variant in the gene encoding glypican 4 (*GPC4*). This variant, located in the final exon of *GPC4,* results in premature termination of the protein 51 amino acid residues prior to the stop codon, and in concomitant loss of functionally important N-linked glycosylation (Asn514) and glycosylphosphatidylinositol (GPI) anchor (Ser529) sites. We subsequently identified seven affected males from five additional kindreds with novel and predicted pathogenic variants in *GPC4*. Segregation analysis and X-inactivation studies in carrier females provided supportive evidence that the *GPC4* variants caused the condition. Furthermore, functional studies of recombinant protein suggested that the truncated proteins p.Gln506* and p.Glu496* were less stable than the wild type. Clinical features of Keipert syndrome included a prominent forehead, a flat midface, hypertelorism, a broad nose, downturned corners of mouth, and digital abnormalities, whereas cognitive impairment and deafness were variable features. Studies of *Gpc4* knockout mice showed evidence of the two primary features of Keipert syndrome: craniofacial abnormalities and digital abnormalities. Phylogenetic analysis demonstrated that *GPC4* is most closely related to *GPC6*, which is associated with a bone dysplasia that has a phenotypic overlap with Keipert syndrome. Overall, we have shown that pathogenic variants in *GPC4* cause a loss of function that results in Keipert syndrome, making *GPC4* the third human glypican to be linked to a genetic syndrome.

Keipert syndrome, also known as nasodigitoacoustic syndrome (MIM: 255980), is a rare, X-linked disorder characterized by craniofacial and digital abnormalities and variable learning difficulties and sensorineural deafness.[1] Keipert et al.[2] first described the syndrome in two brothers, and it has subsequently been reported in one other male sibling pair,[3] five isolated male individuals,[4–7] and a girl and her less severely affected father.[8] The facial appearance is distinctive and comprises a broad forehead, hypertelorism, a prominent nose, a wide mouth, and a prominent upper lip with a cupid's bow configuration. Changes of the digits are also distinctive: there is widening of all distal phalanges, particularly those of the thumbs and great toes.

Previously, we described the diagnosis of Keipert syndrome in the maternal nephew of the brothers originally reported by Keipert et al.[2] and mapped Keipert syndrome to Xq22.2-Xq28 in this family (KS1).[1] To identify the underlying genetic cause of Keipert syndrome, we performed whole-exome capture and massively parallel sequencing of gDNA isolated from the maternal nephew (Figure 1A, KS1 II:5). Institutional ethics approval was provided by the Royal Children's Hospital (Melbourne), and written informed consent was received from all participants prior to study. Three variants were identified within the linkage region after exclusion of synonymous variants and filtering against population databases (Table S1). The

[1]Murdoch Children's Research Institute, Flemington Road, Parkville, Victoria 3052, Australia; [2]Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia; [3]Department of Otolaryngology—Head and Neck Surgery, Stanford University, Stanford, CA 94305, USA; [4]Institut für Medizinische Genetik, Charité - Universitätsmedizin Berlin, Augustenburger Platz 1, Berlin 13353, Germany; [5]Department of Human Genetics, Radboud University Medical Center, Nijmegen 6500 HB, the Netherlands; [6]The Walter and Eliza Hall Institute of Medical Research, Royal Parade, Parkville, Victoria 3052, Australia; [7]Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010, Australia; [8]Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA; [9]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, 6500HB Nijmegen, The Netherlands; [10]Molecular Neurobiology Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA; [11]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [12]Texas Children's Hospital, Houston, TX 77030, USA; [13]Laboratory of Medical Genetics, Department of Laboratories, IRCCS Bambino Gesù Children's Hospital, Rome 00165, Italy; [14]Medical Genetics, Department of Pediatrics, IRCCS Bambino Gesù Children's Hospital, Rome 00165, Italy; [15]Division of Metabolism, Department of Pediatric Subspecialties, IRCCS Bambino Gesù Children's Hospital Rome 00165, Italy; [16]Center for Rare Disease and Congenital Defects, Fondazione Policlinico Universitario A. Gemelli, Catholic University, Rome 00168, Italy; [17]Department of Pediatrics Rebro, University of Zagreb Medical School, Kispaticeva 12, 10 000 Zagreb, Croatia; [18]Department of Clinical Genetics, Our Lady's Children Hospital Crumlin, Dublin, D12, V004, Ireland; [19]Department of Clinical Genetics, Royal Devon and Exeter National Health Service Foundation Trust, Exeter EX1 2ED, UK; [20]Department of Neurology, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia; [21]Department of Medical Genetics, Antwerp University Hospital and University of Antwerp, B-2650 Antwerp (Edegem), Belgium
*Correspondence: paul.lockhart@mcri.edu.au
https://doi.org/10.1016/j.ajhg.2019.02.026.

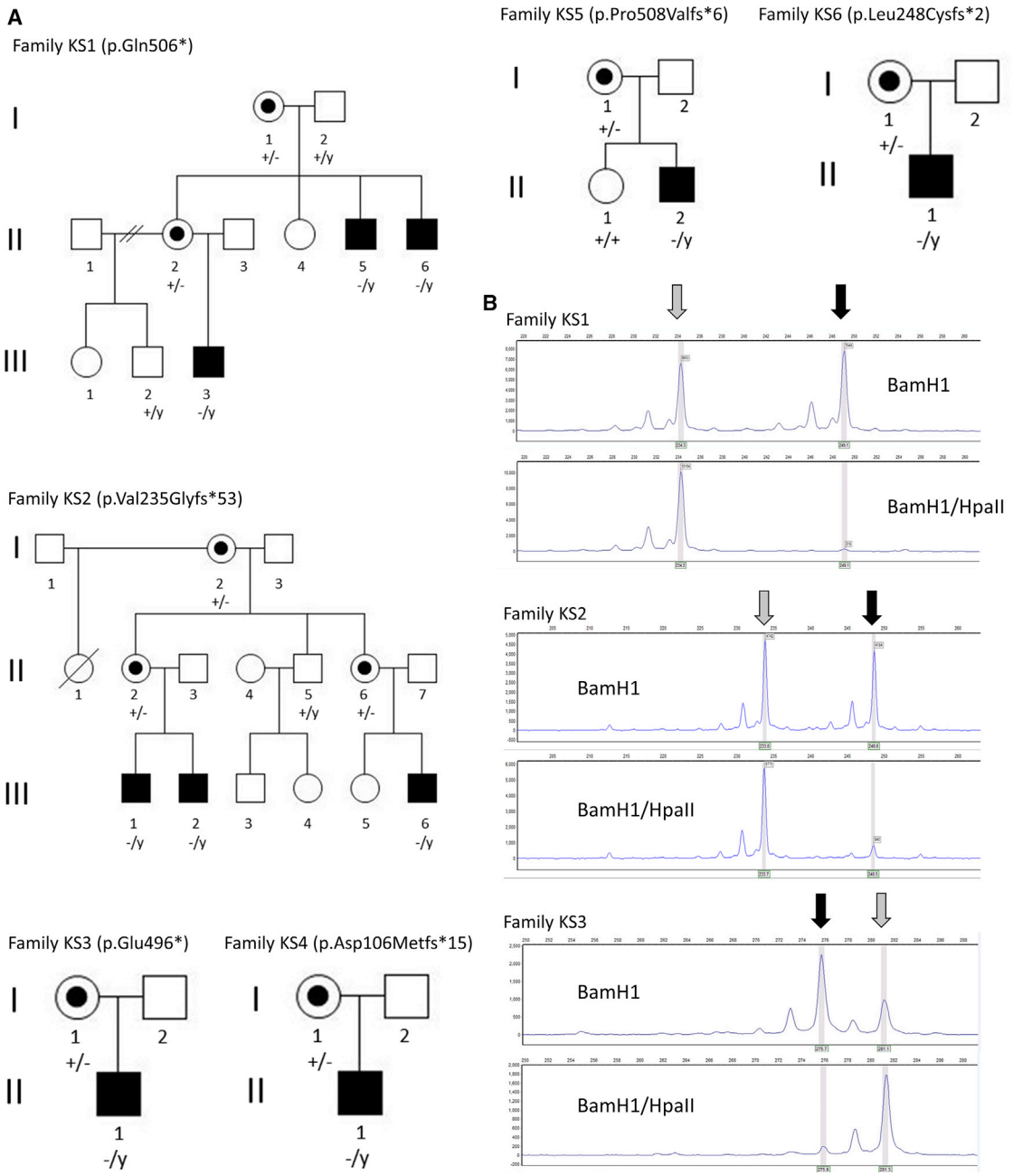


Figure 1. Pedigree Structure and Segregation of Pathogenic Variants in *GPC4*

(A) Pedigrees showing the clinical phenotypes and segregation of the WT (+) and mutant (−) *GPC4* allele. Affected individuals are shaded, Y indicates the Y chromosome, and carrier females are indicated by a dot.

(B) Representative images of X-inactivation analysis in one carrier female from families KS1 (I:1), KS2 (I:2), and KS3 (I:1) demonstrate extreme skewing of X inactivation, as evidenced by the inability of the methylation-sensitive restriction endonuclease HpaII to digest one allele of the androgen receptor (gray arrow). The second allele, with minimal methylation (black arrow), is almost fully digested by HpaII.

nonsynonymous variant in the gene encoding the arginine vasopressin receptor-2 (*AVPR2*, c.743G>A [p. Arg248His], GenBank: NM_000054.5) was excluded because it was classified as a variant of uncertain significance (VUS) according to ACMG guidelines, and pathogenic variants in this gene cause diabetes insipidus (MIM: 300538). A truncating variant in the gene encoding melanoma-associated antigen (mutated) 1-like 1 was identified (*MUM1L1*, c.1404G>A [p.Trp468*], GenBank: NM_001171020.2). This variant was classified as a VUS and was not considered to be a strong candidate for Keipert syndrome. Moreover, analysis of the Genotype-Tissue Expression (GTEx) database demonstrated negligible expression of *MUM1L1* in all tissues except ovary, and dysregulation is associated with ovarian cancers.[9] Analysis of the gene in gnomAD revealed moderate evidence of gene intolerance to loss of function[10] (pLI = 0.49; observed/expected = 0.2, CI 0.09–0.53). The third variant, which was identified as most likely to be

disease causing, was a hemizygous C-to-T transition variant in the ninth and final exon of glypican 4; this variant resulted in premature protein termination (*GPC4*, c.1516C>T [p.Gln506*], GenBank: NM_001448.2). *GPC4* is classified as extremely intolerant of loss-of-function variants (pLI = 0.94; observed/expected = 0.11, CI 0.04–0.36) and is expressed during development in a variety of tissue types relevant to Keipert syndrome. Notably, GPC4 is a member of the glypican protein family, whose members are cell-surface heparan sulfate proteoglycans that regulate growth-factor signaling during development.[11,12] Sanger sequencing confirmed the variant and segregation within the KS1 pedigree (Figure 1A).

Further interrogation of the gnomAD database identified, from within the ~182,000 sequenced alleles, two hemizygous alleles predicted to be loss of function with high-confidence (LoF allele frequency = 1.1×10^{-5}; c.283C>T, [p.Gln95*], and c.1150C>T, [p.Arg384*]), even though the database is considered to be relatively free of sequences from individuals with pediatric disorders. We did not believe this would exclude *GPC4* as the gene underlying the disorder in family KS1 because the phenotype within the family is quite variable. For example, KS1 II-5 has subtle phenotypic features and minimal impairment from his condition; he has a job and has fathered a child. We consider it likely that the phenotype of Keipert syndrome in some cases might escape clinical recognition and therefore be included in gnomAD. This situation has been observed for other craniofacial syndromes with phenotypic heterogeneity; for example, gnomAD contains two individuals who have the pathogenic p.Pro250Arg *FGFR3* variant that causes Muenke syndrome but presents with variable phenotypes.[13]

We next searched for additional individuals with pathogenic variants in *GPC4* to provide genetic validation of family KS1. Five affected males from three additional kindreds with novel variants in *GPC4* were recruited to this study through GeneMatcher (Figure 1A).[14] In family KS2, a frameshift variant was identified in exon 3 in two brothers and a male cousin (c.701dup [p.Val235Glyfs*53]). In family KS3, a truncating variant was identified in exon 9 in one affected male (c.1486G>T [p.Glu496*]), and in family KS4, a frameshift variant was identified in exon 2 (c.316delG [p.Asp106Metfs*15]) in one affected male. Notably, no potentially pathogenic variants were identified in either *AVPR2* or *MUM1L1* in these cases. In addition, two additional families were identified during the course of the study. In family KS5, a frameshift variant (c.1518_1521dupGTGC [p.Pro508Valfs*6]) was identified in exon 9 in one affected male, and in family KS6, a frameshift variant (c.742delC [p.Leu248Cysfs*2]) was identified in exon 4 in one affected male. Genotype analysis in all available samples from the six families confirmed segregation of the variants with disease and demonstrated inheritance from a carrier mother (Figure 1A). In the two multigenerational families (KS1 and KS2), this analysis also confirmed that the variant was present in a carrier grandmother. Transmission of the variant was observed through nine segregating meioses in total, providing additional genetic support for the linkage and gene-identification findings. Carrier females were clinically unaffected, although the mother of KS4 I:1 was noted to have possible subtle dysmorphic features such as hypertelorism and a flat nasal bridge. 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, essentially as described previously.[15] In this assay, differential methylation of the X chromosomes is quantitated by methylation-sensitive digestion of genomic DNA. Allelic ratios from 50%–79% were considered to reflect a normal X-inactivation pattern, from 80%–90% were considered to reflect moderate skewing, and >90% were considered to reflect strong skewing. In total, we analyzed six individuals from four families (KS1 I:1 and II:2; KS2 I:2 and II:6; KS3 I:1; and KS4 I:1). Although analysis of the (CAG)n repeat was not informative in family KS4, X-inactivation studies in all the carrier females from KS1, KS2, and KS3 showed extreme skewing of >90% methylation of one allele (representative results are shown in Figure 1B). Collectively, these observations provide significant support for X linkage in the families and strongly suggest that variants in *GPC4* underlie Keipert syndrome.

Interestingly, with the exception of the first family (KS1) described with Keipert syndrome, all *GPC4* variants were detected via a "genotype-first" approach, yet these individuals share very similar phenotypes with the original kindred (Figures 2A–2O and Table 1). Clinical features present in seven or more of the ten affected individuals were macrocephaly, a prominent forehead, a flat midface, hypertelorism, a broad nose, downturned corners of mouth, and digital abnormalities. The most prominent peripheral skeletal features were brachydactyly, clinodactyly and/or camptodactyly, broad terminal phalanges, and broad thumbs and great toes. Cognitive impairment was variable; intellectual ability ranged from average to moderate intellectual disability, and autism or autistic features were observed in three affected individuals. Identification of these newly reported cases enabled us to extend and refine the phenotype of Keipert syndrome. In particular, the spectrum of digital abnormalities was broadened: absent toenails were observed in individuals KS2 III:2 and KS3 II:1. The latter individual also had synostosis between metatarsals III and IV in the right foot and a missing metatarsal III with hypoplastic phalanx in the left foot. Although sensorineural deafness was described initially as a cardinal feature of Keipert syndrome,[1,2] we did not observe deafness in any of the other families with confirmed *GPC4* pathogenic variants, suggesting that deafness might be relatively infrequent in Keipert syndrome. To further test this, we obtained genomic DNA from one female and three male individuals previously described as having a clinical diagnosis of Keipert syndrome[5,7,8] Three of the four affected individuals were reported to have moderate sensorineural hearing loss. However, *GPC4* variants were not detected in any, suggesting that these individuals have an overlapping but distinct disorder.

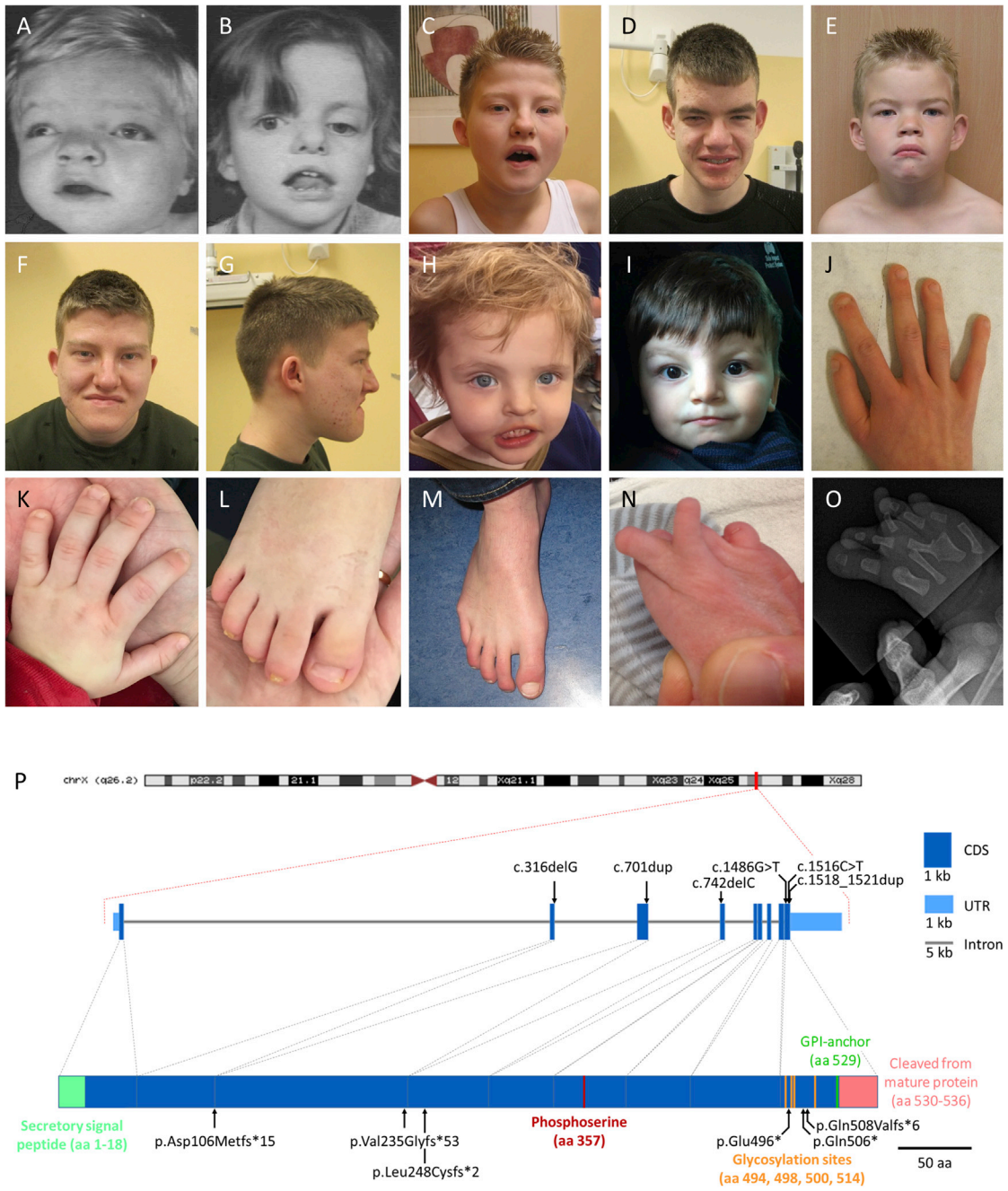


Figure 2. Clinical Presentation of Subjects with Keipert Syndrome and Mutations in *GPC4*

(A–I) Facial features of individuals, demonstrating hypertelorism, a broad forehead, a broad nose, a flat midface, prominent lips, and downturned corners of mouth. Illustrated are (A) KS1 II:5 at age 15 months, (B) KS1 II:6 at age 5 years, (C and D) KS2 III:1 at age 12 years and 19 years, (E, F, and G) KS2 III:2 at ages 7 years and 17 years, (H) KS4 II:1 at age 3 years, and (I) KS5 II:2 at age 2 years.
(J) Right hand of KS2 III:1, showing brachydactyly and broad terminal phalanges.
(K–L) Left hand and right foot of KS5 II:2, showing brachydactyly and a broad great toe.
(M) Right foot of KS2 III:2, showing clinodactyly and broad terminal phalanges.
(N–O) Right foot of KS3 II:1, showing synostosis between metatarsals III and IV.
(P) Schematic representation of the location of pathogenic variants identified in *GPC4* and the exon and protein domain structure. The signal peptide sequence (green), phosphoserine (red), glycosylation (orange), lipidation (dark blue), and cleaved GPI anchor (pink) are indicated. The gene structure was derived from the UCSC Genome Bioinformatics database, and protein coordinates were obtained from Uniprot. Informed consent for publication of photos was obtained from all individuals or their parents.

All identified variants were truncating or resulted in a frameshift, suggesting loss of function as the likely disease mechanism. The variants in exon 2 (c.316delG [p.Asp106Metfs*15]), exon 3 (c.701dup [p.Val235Glyfs*53]), and exon 4 (c.742delC [p.Leu248Cysfs*2]) result in loss of >50% of the 556 amino acid protein and

Table 1. Clinical Features and *GPC4* Variants of Affected Individuals

Family	KS1	KS1	KS1	KS2	KS2	KS2	KS3	KS4	KS5	KS6	Summary
Pedigree ID	II:5	II:6	III:3	III:1	III:2	III:6	II:1	II:1	II:2	II:1	
Gene variant	c.1516C>T (p.Gln506*)	c.1516C>T (p.Gln506*)	c.1516C>T (p.Gln506*)	c.701dup (p.Val235Glyfs*53)	c.701dup (p.Val235Glyfs*53)	c.701dup (p.Val235Glyfs*53)	c.1486G>T (p.Glu496*)	c.316delG, p.(Asp106Metfs*15)	c.1518_1521dupGTGC (p.Pro508Valfs*6)	c.742delC (p.Leu248Cysfs*2)	
Age at last examination	40 years	38 years	7 months	19 years	12 years	6 years	3 years	6 years	2 years	3 years	
Gender	male	male	male	male	male	male	male	male	male	male	
Facial Features											
Macrocephaly	no	yes	yes	no	no	yes	borderline	yes	yes	yes	7/10
Prominent forehead	yes	yes	yes	no	yes	unknown	yes	yes	no	yes	7/9
Flat midface	yes	yes	yes	yes	yes	unknown	yes	yes	no	no	7/9
Hypertelorism	yes	yes	yes	yes	yes	unknown	yes	yes	yes	yes	9/9
Broad nose	yes	yes	yes	yes	yes	unknown	yes	yes	yes	yes	9/9
Downturned corners of mouth	yes	yes	yes	yes	yes	unknown	yes	no	yes	no	7/9
Prominent lip	yes	yes	yes	yes	yes	unknown	yes	no	no	no	6/9
Ears simple or low set	yes	no	no	yes	no	unknown	no	yes	yes	yes	5/9
Skeletal											
Brachydactyly	yes	yes	no	yes	no	unknown	yes	no	yes	no	5/9
Clinodactyly	yes	yes	yes	no	yes	unknown	yes	no	no	no	5/9
Camptodactyly	yes	yes	yes	no	no	unknown	no	no	no	no	3/9
Broad thumb	yes	yes	yes	no	no	unknown	no	no	yes	yes	5/9
Broad first toe	yes	yes	yes	yes	no	unknown	yes	no	yes	yes	7/9
Broad terminal phalanges	yes	yes	yes	yes	yes	yes	yes	no	no	no	7/10
Other Features											
Sensorineural hearing loss	moderate unilateral	severe bilateral	moderate bilateral	no	no	unknown	no	no	no	no	3/9
Cognitive impairment	no	mild intellectual disability	learning difficulties	moderate intellectual disability (IQ 52)	borderline intellectual disability (IQ 76)	no (IQ 92)	borderline intellectual disability (IQ 70)	intellectual disability	probable	yes	8/10

(*Continued on next page*)

Table 1. *Continued*

Family	KS1	KS1	KS1	KS2	KS2	KS2	KS3	KS4	KS5	KS6	Summary
Additional features	–	unilateral ptosis	unilateral absent kidney	hyperlaxity of joints	hyperlaxity of joints	autism	synostosis of metatarsal bones with hypoplastic phalanx	behavioral difficulties and autistic traits	neonatal cholestasis (self-resolving)	autistic features	–
	–	strabismus	–	amblyopia	delayed eruption of permanent dentition	–	–	–	neonatal hypotonia	stereotypic movements	–
	–	double alveolar margin	–	patent ductus arteriosus	strabismus	–	–	–	speech delay	–	–
	–	–	–	hypodontia	absent fifth toenails	–	–	–	gastresophageal reflux	–	–
	–	–	–	cryptorchidism (bilateral)	finger-like thumbs	–	–	–	–	–	–
	–	–	–	finger-like thumbs	–	–	–	–	–	–	–

were classified as pathogenic. However, the truncating variants in families KS1, KS3, and KS6 (c.1516C>T [p.Gln506*], c.1486G>T [p.Glu496*]), and c.1518_1521dupGTGC [p.Pro508Valfs*6]) are within the last exon and therefore might not be subject to nonsense-mediated decay. Although the resultant proteins lack glycosylation sites and the glycosylphosphatidylinositol (GPI) anchor (Ser529) that are critical for sorting and localization of GPC proteins (Figure 2P), it is possible that the variants do not result in a loss of function. To further test the pathogenic mechanism underlying p.Glu496* and p.Gln506*, we analyzed recombinant N-terminal V5-tagged GPC4 in cultured cells. We have previously utilized similar methodology to investigate the function of glypicans in the formation of active synapses.[16] Expression constructs encoding V5-tagged wild-type (WT) or truncated (p.Glu496* and p.Gln506*) recombinant human GPC4 were transfected into HEK293 cells, and stably transfected populations were isolated after 3 weeks of culturing in media supplemented with 400 μg/mL geneticin (G418 sulfate, ThermoFisher Scientific). Qualitative visualization of the cell populations by immunocytochemical analysis with an antibody directed against V5 suggested reduced amounts of the mutant GPC4 proteins compared to WT (data not shown). Therefore, we performed immunoblot analysis to determine steady-state amounts of GPC4 under basal conditions and after inhibition of the ubiquitin proteasome system (UPS), the major intracellular protein degradation pathway for short-lived or damaged proteins. Cells were cultured overnight in basal media or media supplemented with 10 μM MG-132, which we have previously shown is sufficient to inhibit the UPS without causing significant cellular toxicity in HEK293 cells.[17] Protein extracts were prepared, and immunoblot analysis, performed essentially as previously described,[18] identified an apparent increase in the steady-state amounts of mutant but not WT recombinant V5 GPC4 after proteasome inhibition (Figure 3A). Quantitation and comparison of the recombinant GPC4/β-actin ratio in the untreated cells suggested the steady-state amount of p.Gln506* was reduced in comparison to that of the WT (WT 0.026 ± 0.002 vs p.Gln506* 0.014 ± 0.002, mean ± SEM, $p = 0.042$), whereas there was no significant difference in the amounts of p.Glu496* compared to WT (WT 0.026 ± 0.002 vs p.Glu496* 0.029 ± 0.004, mean ± SEM, $p = 0.565$). In contrast, after exposure to 10 μM MG-132 for 14 h, both p.Gln506* and p.Glu496* steady-state amounts were significantly elevated in comparison to steady-state amounts in the treated WT cells (WT 0.030 ± 0.002 vs p.Gln506* 0.041 ± 0.003, mean ± SEM, $p = 0.024$; and WT 0.030 ± 0.002 vs p.Glu496* 0.199 ± 0.009, mean ± SEM, $p = 0.002$) (Figure 3B). Therefore, having an active UPS resulted in a relative decrease in steady-state protein amounts of ~2.5-fold for p.Gln506* and ~6-fold for p.Glu496* compared to those in UPS-inhibited cells.

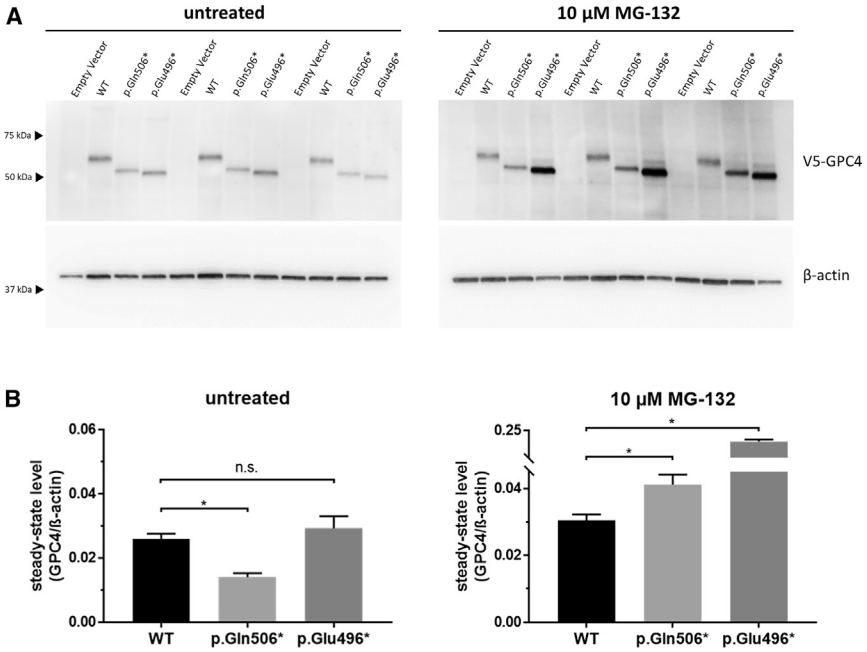


Figure 3. Truncated GPC4 Proteins Are Unstable and Degraded by the UPS

(A) HEK293 cells transfected with constructs encoding N-terminally tagged WT and truncated GPC4 were grown in the absence or presence of 10 μM MG-132 for 14 hours. Total protein was isolated and analyzed by SDS-PAGE and immunoblotting with antibodies directed against V5 (46-0705, ThermoFisher Scientific, 1:5000) and β-actin (A5441, Sigma; 1:20,000). A representative image is shown, and approximate sizes in kDa are indicated (MultiMark Standard, Invitrogen).

(B) Quantification of three independent experiments performed in triplicate. The immunoreactive signals were quantified with a LAS4000 imager. Steady-state protein amounts are expressed as the ratio of GPC4/β-actin, which is the loading control for normalization.

An asterisk denotes $p \leq 0.05$; statistical comparisons were made with a two-tailed t test; error bars represent mean ± SEM.

Collectively, the reduced stability and removal of critical protein domains suggest a loss of GPC4 function as the pathogenic mechanism underlying disease in all six families described in this study.

To date, there has been very limited information about *GPC4*-associated phenotypes in humans. A hemizygous missense variant (c.1235G>A [p.Arg412Lys]) was reported as likely pathogenic in an individual with a Robinow-syndrome-like phenotype that included some findings seen in Keipert syndrome; such findings included macrocephaly; hypertelorism; broad thumbs and great toes; and camptodactyly. However, this individual had more prominent skeletal features, including mesomelia and cranial sclerosis,[19] which were not seen in the Keipert-syndrome-affected individuals with truncating *GPC4* mutations in our study. Duplication of *GPC4* has also been implicated in one family that is affected by the overgrowth syndrome Simpson-Golabi-Behmel syndrome (SGBS, MIM: 312870), an X-linked disorder caused by pathogenic variation in the gene encoding glypican 3 (*GPC3*). SGBS is characterized by pre- and postnatal overgrowth, skeletal abnormalities, variable intellectual disability, and congenital anomalies such as congenital heart defects and diaphragmatic hernia.[20–22] This SGBS-affected family, initially reported by Golabi and Rosen,[21] was followed up on and was reported in 2010 as not having a *GPC3* mutation but as having a duplication of exons 1–9 of *GPC4*.[23] This duplication segregated with a phenotype that included macrosomia, coarse facies, a short nose, a broad nasal bridge, macroglossia, and accessory nipples. The duplication mapped close to the 3′ end of *GPC3*, and the authors speculated that the duplication might result in decreased expression of *GPC3*, or alternatively that *GPC4* might be a second gene for SGBS.[23] However, since that publication, sequencing and deletion analysis of *GPC4* in other individuals with SGBS and other overgrowth disorders has not detected any pathogenic variants.[20,24]

There have been two reports of SGBS associated with deletions that encompass both *GPC3* and *GPC4*. In a family that included some members with a clinical diagnosis of SGBS and that was originally described by Pilia et al.,[22] a deletion that removed the entire *GPC4* gene and the last two exons of *GPC3* was detected.[25] The SGBS phenotype in males in this family was noted to include renal-tract abnormalities and hydrocephalus. The second report was of a 1 Mb deletion encompassing three genes, *GPC4*, *GPC3*, and *CCDC160*, in a male fetus that was terminated at 24 weeks; his phenotype comprised macrosomia, visceromegaly, macroglossia, polyhydramnios, and mild ventriculomegaly.[26] In both families, there was insufficient clinical information provided to allow determination of whether an overlapping phenotype, comprising features of both SGBS and Keipert syndrome, might have been present.

Glypicans are a family of cell-surface heparan sulfate proteoglycans characterized by a GPI anchor that localizes them to the cell surface, where they regulate growth-factor signaling during development and disease.[27] They are expressed predominantly during development and are thought to play a role in the regulation of morphogenesis. In mammals, there are six glypicans, GPC1 to GPC6. The genes encoding GPC3 and GPC4 and those encoding GPC5 and GPC6 are clustered on chromosome Xq26 and chromosome 13q32, respectively, suggesting that the glypican family arose as a result of gene duplication. Filmus et al.[28] reported evidence of homologs of glypican throughout Eumetazoa but were unable to identify glypican homologs outside the Metazoa linage. To analyze the evolution of the glypican family, we used the maximum likelihood method with Molecular Evolutionary Genetics Analysis (MEGA) software version 6,[29] and in doing this,

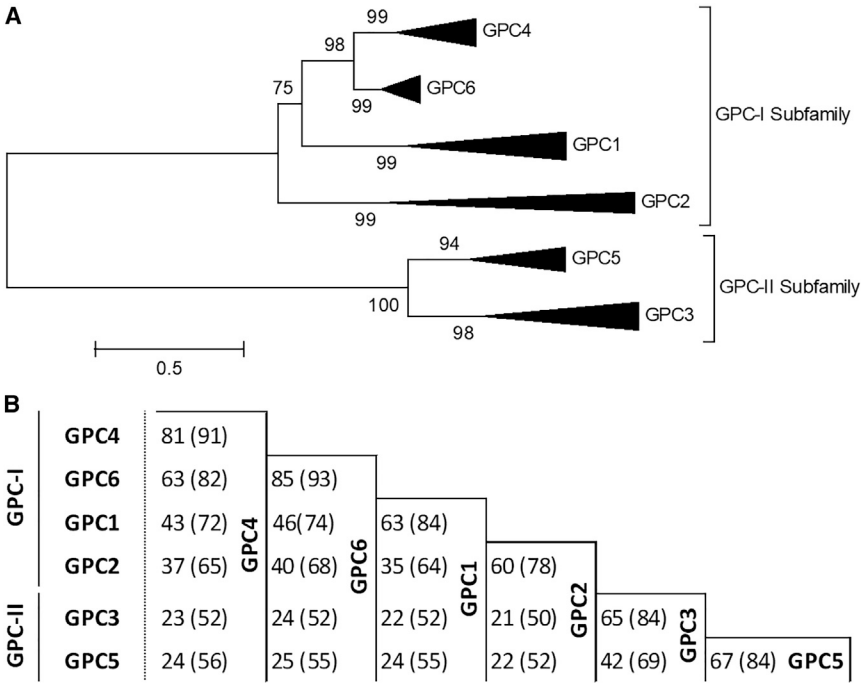


Figure 4. GPC4 and GPC6 Are the Most Closely Related of All Glypican Proteins
(A) Phylogenetic analysis of vertebrate (2R) glypicans was inferred via the maximum-likelihood method. GPC1, GPC2, GPC4, and GPC6 share a common ancestor; GPC3 and GPC5 share a common ancestor. GPC4 and GPC6 have undergone considerably less diversification than have other glypican proteins. The tree is drawn to scale, and branch lengths are in the same units as those of the evolutionary distances used to infer the phylogenetic tree.
(B) The average amino acid identity and similarity (brackets) of vertebrate (2R) glypican proteins were determined by global pairwise alignment. Sequence accession numbers are indicated in Figure S1.

we independently replicated the Eumetazoa result. In addition, evidence of glypicans was identified in the genomes of both the Placozoa *Trichoplax adhaerens* and the sponge *Amphimedon queenslandica*. These genomes represent the basal group of multicellular organisms (Figure S1). To determine the orthology of the glypican family, we performed phylogenetic analysis by using the maximum likelihood method. There are two major glypican subfamilies, the *GPC-I* subfamily, which comprises *GPC1*, *GPC2*, *GPC4,* and *GPC6*, and the *GPC-II* subfamily, which comprises *GPC3* and *GPC5*. Robust branch partitioning suggested the likelihood that a single gene ancestor of both subfamilies originated prior to the divergence of the Eumetazoan genome (Figure S2). Notably, in the chordate genome of the lancelet (*Branchiostoma floridae*, Bflo) the single *GPC-I* ortholog and the single *GPC-II* ortholog were identified side-by-side on a single scaffold (BRAFLscaffold_196; GenBank: NW_003101409, data not shown) in an arrangement reflecting both the *GPC3-GPC4* and the *GPC5-GPC6* genomic clustering. *GPC3* and *GPC5* are inferred to be descended from a single chordate *GPC-I* ortholog and *GPC4* and *GPC6* from a single *GPC-II* ortholog. Therefore, both loci probably arose from a single locus. To refine the relationship among vertebrate glypicans, we performed phylogenetic analysis by using only glypican proteins from vertebrate genomes that had undergone two rounds of whole-genome duplication. Although glypicans do not appear to encode well-characterized functional protein domains[30] and although the amino acid homology of mammalian proteins is as low as 25%, the three-dimensional structure is conserved across the family.[11] Among all glypicans, the greatest similarity at the protein level is found between GPC3 and GPC5 and between GPC4 and GPC6 (Figure 4). In addition, this study has demonstrated that GPC4 and GPC6 are the most closely related of all glypicans.

Apart from *GPC4* (this report, Keipert syndrome) and *GPC3* (SGBS), only *GPC6* has a clear association with Mendelian disorders. Pathogenic variants in *GPC6* cause autosomal-recessive omodysplasia (MIM: 258315), a skeletal dysplasia characterized by short limbs and craniofacial abnormalities (frontal bossing, a depressed nasal bridge with a short nose, and a long and prominent philtrum).[31] It is notable that there is phenotypic overlap between the craniofacial abnormalities in Keipert syndrome and those in omodysplasia; however, in omodysplasia, limb abnormalities are far more severe than in Keipert syndrome. Pathogenic variants in *GPC1*, *GPC2,* and *GPC5* have not yet been shown to cause any Mendelian disorder, although common variants in *GPC5* have been implicated in acquired nephrotic syndrome[32] and in the risk of lung cancer in never-smokers.[33] Amplification and overexpression of *GPC5* has been observed in rhabdomyosarcoma.[34]

We have previously identified GPC4 and GPC6 as astrocyte-secreted proteins that induce synapse formation in rodent neurons.[16] GPC4 and GPC6 have redundant functions *in vitro*, but *Gpc4* knockout (KO) mice have defective synapse formation in the developing hippocampus, a part of the brain that is enriched for Gpc4.[16] In day 7–10.5 embryos, *Gpc4* mRNA localizes to a range of sites that include the anterior forebrain neuroepithelium, branchial arches, optic and otic vesicles, limb buds, and somites.[12] We performed additional studies in *Gpc4*-deficient mice to look for the two primary features of Keipert syndrome: craniofacial abnormalities and digital abnormalities. All bone measurement experiments were approved by the Salk Institute Institutional Animal Care and Use Committee (IACUC). In addition, we tested the mice for sensorineural deafness, which was observed in the originally described Keipert-syndrome-affected family but might not be a defining feature of the syndrome. Hearing experiments were approved by Stanford University IACUC.

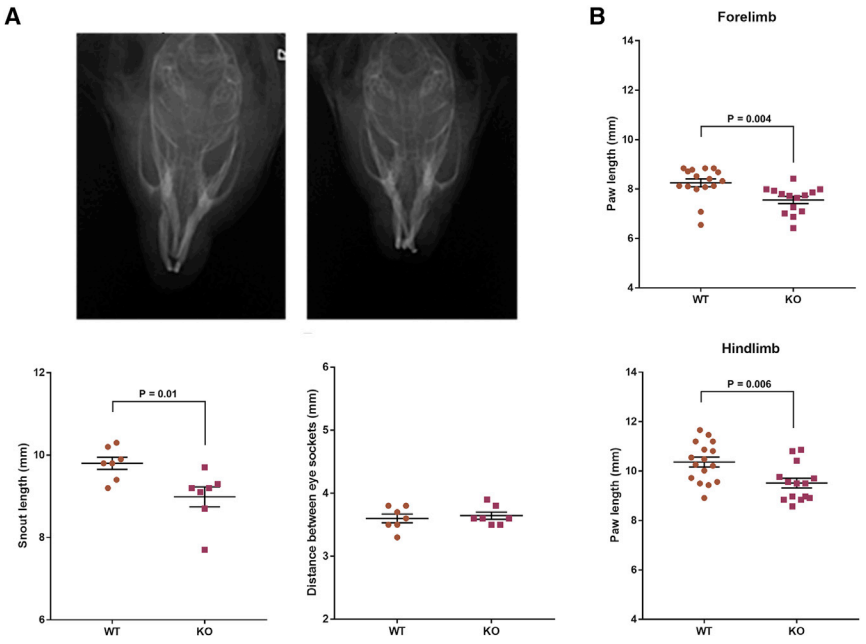


Figure 5. *Gpc4* KO Mice Exhibit Nasodigital Deficits

(A) Example radiographs showing typical skull morphology of adult WT (left) and *Gpc4* KO (right) mice. Note the shortened snout in the KO mouse. Quantification of radiograph data demonstrated that *Gpc4* KO mice have significantly shorter snouts; snouts were 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.

(B) *Gpc4* KO mice have significantly shorter front and rear paws.

n = 7 animals per group; data combine right and left paw to give 14 measurements; statistical comparisons were made with a two-tailed t test; error bars represent mean ± SEM.

Glypican-4-null (male) mice with targeted KO of exon 3 (strain MMRRC_032331-UCD) were maintained as we previously described,[16,35] and 16- to 20-week-old littermate male *Gpc4* KO and WT mice were used for analysis. To determine the effect of *Gpc4* loss of function on skull geometry, we took X-rays of the skull by using a CR 7 Digital Dental X-Ray imager and performed two sets of measurements: (1) from the cranial aspect to the tip of the snout and (2) for intra-orbital distance, between the medial-edge of each orbit. There was a significant decrease in the length of the snout in *Gpc4* KO mice compared to WT (WT—9.8 ± 0.1 mm, n = 7 versus KO—9.0 ± 0.2 mm, n = 7, mean ± SEM, p = 0.014), but there was no difference in intra-orbital separation (WT—3.6 ± 0.1 mm, n = 7 versus KO—3.6 ± 0.1 mm, n = 7, mean ± SEM, p = 0.64) (Figure 5A). Similar results were obtained when measurements for snout length were taken with an instant-readout precision digital caliper (Figure S3), suggesting that loss of GPC4 causes craniofacial abnormalities in the mouse.

To determine the effect of *Gpc4* loss of function on the digits, mice were placed on a treadmill, and videos of their paws were captured as they walked. We used GaitScan software to analyze both fore and hind paws and analyzed a minimum of five images per paw per mouse. We marked the position of the base of the paw and the tip of each phalanx. The distance between the tip of the longest phalanx (#3 of 5) and base was measured for hind paws. Because the fore paws have only four phalanges, the #3 marker was positioned equidistant between phalange 2 and 4. There was a significant decrease in the length of both the fore and hind paws in *Gpc4* KO mice compared to WT (fore paws—WT 8.3 ± 0.16 mm, n = 7 versus KO 7.6 ± 0.14 mm, n = 7, mean ± SEM, p = 0.004; hind paws—WT 10.4 ± 0.2 mm, n = 7 versus KO 9.5 ± 0.2 mm, n = 7, mean ± SEM, p = 0.006) (Figure 5B). We also measured the distance between the phalanges, but there was no difference in the total spread of digits for the forelimb and hindlimbs, as measured from digit 1 to digit 5 (Figure S4). Thus, *Gpc4* KO mice recapitulate the skeletal features seen in Keipert syndrome, albeit more subtly.

To test hearing function in the *Gpc4* KO mice, we performed auditory brainstem response (ABR) recordings. Mice were anesthetized, and acoustic stimuli were delivered to the ear canals 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). The 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 ABR thresholds of P18 *Gpc4* KO mice were compared to those of littermate WT control mice. There was no difference in hearing thresholds at the 8, 16, and 32 kHz frequencies measured between the two genotypes (Figure S5). We analyzed the ABR data at 16 kHz, within the most sensitive frequency range for mice for peak 1 amplitudes and latencies. However, we did not observe a shift in first peak amplitude or latency in *Gpc4* KO mice compared to controls. In addition, to test the function of outer hair cells, we measured distortion-product otoacoustic emissions (DPOAEs) as previously described.[36] No difference in the function of the cochlear amplifier was revealed by DPOAE measurements between the KO and control mice.

In conclusion, we have demonstrated that pathogenic variants in *GPC4* underlie Keipert syndrome. The primary and invariant features of the disorder include craniofacial and digital abnormalities. However, despite being described in KS1, the prototypical family, sensorineural deafness does not appear to be consistently associated with loss of *GPC4* function. We report that *Gpc4* KO mice display morphological abnormalities reminiscent of Keipert syndrome: they have significantly shorter paw and snout lengths than WT mice. In contrast to the relatively subtle physical phenotype

in *Gpc4*-deficient mice, the zebrafish *GPC4* mutant *knypek* does not survive beyond 5–7 days post-fertilization; this is due to severely reduced convergence and extension movements.[37] However, when these gastrulation defects are suppressed by *gpc4* mRNA injection, *knypek* embryos display more subtle abnormalities of the craniofacial cartilages; these abnormalities include shortening of the skull and jaw.[38] These features are consistent with those observed in individuals with Keipert syndrome and in the *Gpc4* KO mice. *GPC4* is the third human glypican to be linked to a genetic syndrome, and our data strengthen the evidence linking the family of glypican genes to disorders of cartilage and bone morphogenesis, information that may be relevant to phenotypes associated with dysregulation of glypicans 1, 2, and 5.

Accession Numbers

The ClinVar accession numbers for the *GPC4* variants reported in this paper are ClinVar: RCV000659264.1, RCV000659265.1, RCV000659266.1, and RCV000659267.1.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.02.026.

Acknowledgments

This work was supported by the Australian Government National Health and Medical Research Council (program grant 1054618 and fellowship 1002098). P.J.L. was supported by the National Health and Medical Research Council (NHMRC) Career Development Fellowship GNT1032364. M.B. was supported by an NHMRC senior research fellowship (1102971) and an NHMRC program grant (1054618). Additional funding was provided by the Independent Research Institute Infrastructure Support Scheme and the Victorian State Government Operational Infrastructure Program. M.A.M. is a participant in the BIH Charité Junior Clinician Scientist Program funded by the Charité—Universitätsmediz in Berlin and the Berlin Institute of Health. V.R.S. receives support from the National Human Genome Research Institute (NHGRI) UM1 HG006542-07. C.D. and N.J.A. were supported by the National Institutes of Health National Institute of Neurological Disorders and Stroke (NIH-NINDS) R01 NS089791. M.M. was supported by the National Institute on Deafness and other Communicative Disorders (NIDCD) R01 DC09590.

Declaration of Interests

The authors declare no competing interests.

Received: April 24, 2018
Accepted: February 25, 2019
Published: April 11, 2019

Web Resources

ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/
GeneMatcher, https://www.genematcher.org/
Genome Aggregation Database (gnomAD), https://gnomad.broadinstitute.org/
Genotype-Tissue Expression (GTEx) project, https://gtexportal.org/home/
Online Mendelian Inheritance in Man, http://www.omim.org/
UCSC Genome Bioinformatics database, https://genome.ucsc.edu/
Uniprot, https://www.uniprot.org/
Varsome, https://varsome.com

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. Keipert, J.A., Fitzgerald, M.G., and Danks, D.M. (1973). A new syndrome of broad terminal phalanges and facial abnormalities. Aust. Paediatr. J. *9*, 10–13.
3. Balci, S., and Dagli, S. (1996). Keipert syndrome in two brothers from Turkey. Clin. Genet. *50*, 223–228.
4. Cappon, S.M., and Khalifa, M.M. (2000). Additional case of Keipert syndrome and review of the literature. Med. Sci. Monit. *6*, 776–778.
5. Reardon, W., and Hall, C.M. (2003). Broad thumbs and halluces with deafness: a patient with Keipert syndrome. Am. J. Med. Genet. A. *118A*, 86–89.
6. Derbent, M., Bikmaz, Y.E., and Agildere, M. (2011). A patient with Keipert syndrome and isolated fibrous dysplasia of the sphenoid sinus. Am. J. Med. Genet. A. *155A*, 1496–1499.
7. Nik-Zainal, S., Holder, S.E., Cruwys, M., Hall, C.M., and Shaw-Smith, C. (2008). Keipert syndrome: two further cases and review of the literature. Clin. Dysmorphol. *17*, 169–175.
8. Dumic, M., Kokic, D.D., Matic, T., and Potocki, K. (2006). Daughter and her mildly affected father with Keipert syndrome. Am. J. Med. Genet. A. *140*, 2488–2492.
9. Emmanuel, C., Gava, N., Kennedy, C., Balleine, R.L., Sharma, R., Wain, G., Brand, A., Hogg, R., Etemadmoghadam, D., George, J., et al.; Australian Ovarian Cancer Study Group (2011). Comparison of expression profiles in ovarian epithelium in vivo and ovarian cancer identifies novel candidate genes involved in disease pathogenesis. PLoS ONE *6*, e17617.
10. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
11. Filmus, J., and Capurro, M. (2014). The role of glypicans in hedgehog signaling. Matrix Biol. *35*, 248–252.
12. Ybot-Gonzalez, P., Copp, A.J., and Greene, N.D. (2005). Expression pattern of glypican-4 suggests multiple roles during mouse development. Dev. Dyn. *233*, 1013–1017.
13. González-Del Angel, A., Estandía-Ortega, B., Alcántara-Ortigoza, M.A., Martínez-Cruz, V., Gutiérrez-Tinajero, D.J., Rasmussen, A., and Gómez-González, C.S. (2016). Expansion of the variable expression of Muenke syndrome: hydrocephalus without craniosynostosis. Am. J. Med. Genet. A. *170*, 3189–3196.
14. 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.

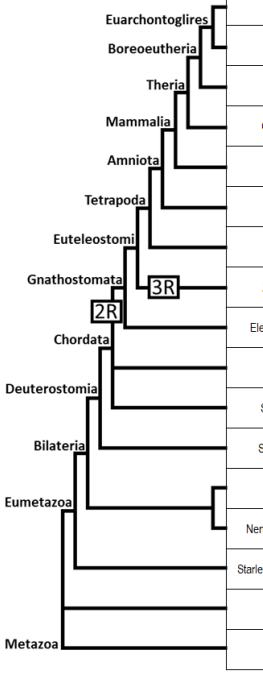
15. Spath, M.A., Nillesen, W.N., Smits, A.P., Feuth, T.B., Braat, D.D., van Kessel, A.G., and Yntema, H.G. (2010). X chromosome inactivation does not define the development of premature ovarian failure in fragile X premutation carriers. Am. J. Med. Genet. A. *152A*, 387–393.
16. Allen, N.J., Bennett, 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.
17. Lockhart, P.J., Lincoln, S., Hulihan, M., Kachergus, J., Wilkes, K., Bisceglio, G., Mash, D.C., and Farrer, M.J. (2004). DJ-1 mutations are a rare cause of recessively inherited early onset parkinsonism mediated by loss of protein function. J. Med. Genet. *41*, e22.
18. Stephenson, S.E.M., Aumann, T.D., Taylor, J.M., Riseley, J.R., Li, R., Mann, J.R., Tomas, D., and Lockhart, P.J. (2018). Generation and characterisation of a parkin-Pacrg knockout mouse line and a Pacrg knockout mouse line. Sci. Rep. *8*, 7528.
19. White, J.J., Mazzeu, J.F., Coban-Akdemir, Z., Bayram, Y., Bahrambeigi, V., Hoischen, A., van Bon, B.W.M., Gezdirici, A., Gulec, E.Y., Ramond, F., et al.; Baylor-Hopkins Center for Mendelian Genomics (2018). WNT signaling perturbations underlie the genetic heterogeneity of Robinow syndrome. Am. J. Hum. Genet. *102*, 27–43.
20. Cottereau, E., Mortemousque, I., Moizard, M.P., Bürglen, L., Lacombe, D., Gilbert-Dussardier, B., Sigaudy, S., Boute, O., David, A., Faivre, L., et al. (2013). Phenotypic spectrum of Simpson-Golabi-Behmel syndrome in a series of 42 cases with a mutation in GPC3 and review of the literature. Am. J. Med. Genet. C. Semin. Med. Genet. *163C*, 92–105.
21. Golabi, M., and Rosen, L. (1984). A new X-linked mental retardation-overgrowth syndrome. Am. J. Med. Genet. *17*, 345–358.
22. Pilia, G., Hughes-Benzie, R.M., MacKenzie, A., Baybayan, P., Chen, E.Y., Huber, R., Neri, G., Cao, A., Forabosco, A., and Schlessinger, D. (1996). Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. Nat. Genet. *12*, 241–247.
23. Waterson, J., Stockley, T.L., Segal, S., and Golabi, M. (2010). Novel duplication in glypican-4 as an apparent cause of Simpson-Golabi-Behmel syndrome. Am. J. Med. Genet. A. *152A*, 3179–3181.
24. Veugelers, M., Cat, B.D., Muyldermans, S.Y., Reekmans, G., Delande, N., Frints, S., Legius, E., Fryns, J.P., Schrander-Stumpel, C., Weidle, B., et al. (2000). Mutational analysis of the GPC3/GPC4 glypican gene cluster on Xq26 in patients with Simpson-Golabi-Behmel syndrome: identification of loss-of-function mutations in the GPC3 gene. Hum. Mol. Genet. *9*, 1321–1328.
25. Veugelers, M., Vermeesch, J., Watanabe, K., Yamaguchi, Y., Marynen, P., and David, G. (1998). GPC4, the gene for human K-glypican, flanks GPC3 on xq26: deletion of the GPC3-GPC4 gene cluster in one family with Simpson-Golabi-Behmel syndrome. Genomics *53*, 1–11.
26. Weichert, J., Schröer, A., Amari, F., Siebert, R., Caliebe, A., Nagel, I., Gillessen-Kaesbach, G., Mohrmann, I., and Hellenbroich, Y. (2011). A 1 Mb-sized microdeletion Xq26.2 encompassing the GPC3 gene in a fetus with Simpson-Golabi-Behmel syndrome report, antenatal findings and review. Eur. J. Med. Genet. *54*, 343–347.
27. Gupta, M., and Brand, M. (2013). Identification and expression analysis of zebrafish glypicans during embryonic development. PLoS ONE *8*, e80824.
28. Filmus, J., and Capurro, M. (2008). The role of glypican-3 in the regulation of body size and cancer. Cell Cycle *7*, 2787–2790.
29. 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.
30. Filmus, J., Capurro, M., and Rast, J. (2008). Glypicans. Genome Biol. *9*, 224.
31. Campos-Xavier, A.B., Martinet, D., Bateman, J., Belluoccio, D., Rowley, L., Tan, T.Y., Baxová, A., Gustavson, K.H., Borochowitz, Z.U., Innes, A.M., et al. (2009). Mutations in the heparan-sulfate proteoglycan glypican 6 (GPC6) impair endochondral ossification and cause recessive omodysplasia. Am. J. Hum. Genet. *84*, 760–770.
32. Okamoto, K., Tokunaga, K., Doi, K., Fujita, T., Suzuki, H., Katoh, T., Watanabe, T., Nishida, N., Mabuchi, A., Takahashi, A., et al. (2011). Common variation in GPC5 is associated with acquired nephrotic syndrome. Nat. Genet. *43*, 459–463.
33. Li, Y., Sheu, C.C., Ye, Y., de Andrade, M., Wang, L., Chang, S.C., Aubry, M.C., Aakre, J.A., Allen, M.S., Chen, F., et al. (2010). Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. Lancet Oncol. *11*, 321–330.
34. Williamson, D., Selfe, J., Gordon, T., Lu, Y.J., Pritchard-Jones, K., Murai, K., Jones, P., Workman, P., and Shipley, J. (2007). Role for amplification and expression of glypican-5 in rhabdomyosarcoma. Cancer Res. *67*, 57–65.
35. 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.
36. 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.
37. Topczewski, J., Sepich, D.S., Myers, D.C., Walker, C., Amores, A., Lele, Z., Hammerschmidt, M., Postlethwait, J., and Solnica-Krezel, L. (2001). The zebrafish glypican knypek controls cell polarity during gastrulation movements of convergent extension. Dev. Cell *1*, 251–264.
38. LeClair, E.E., Mui, S.R., Huang, A., Topczewska, J.M., and Topczewski, J. (2009). Craniofacial skeletal defects of adult zebrafish glypican 4 (knypek) mutants. Dev. Dyn. *238*, 2550–2563.

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.

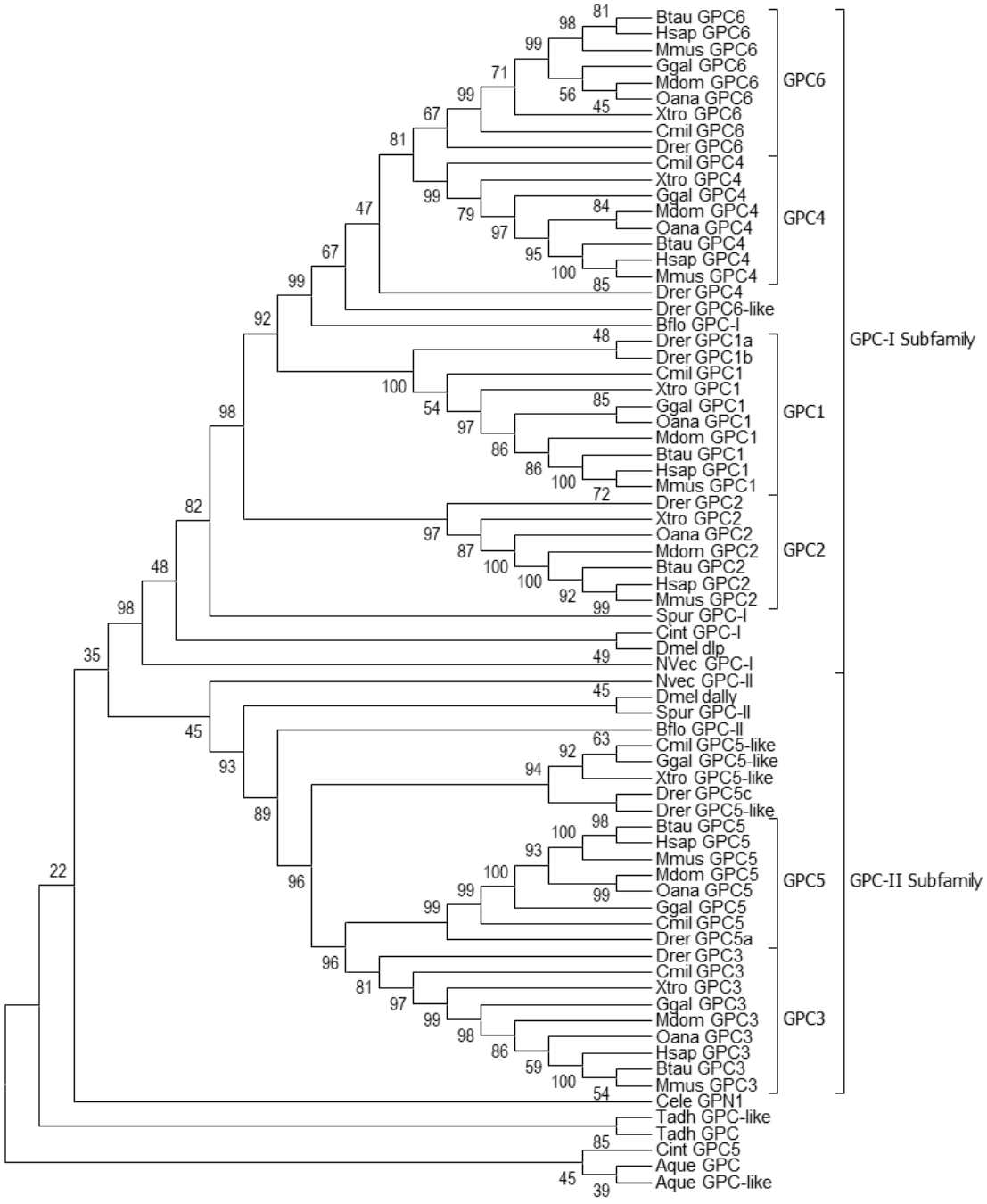


Figure S2: Molecular phylogenetic analysis of the glypican family.

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.

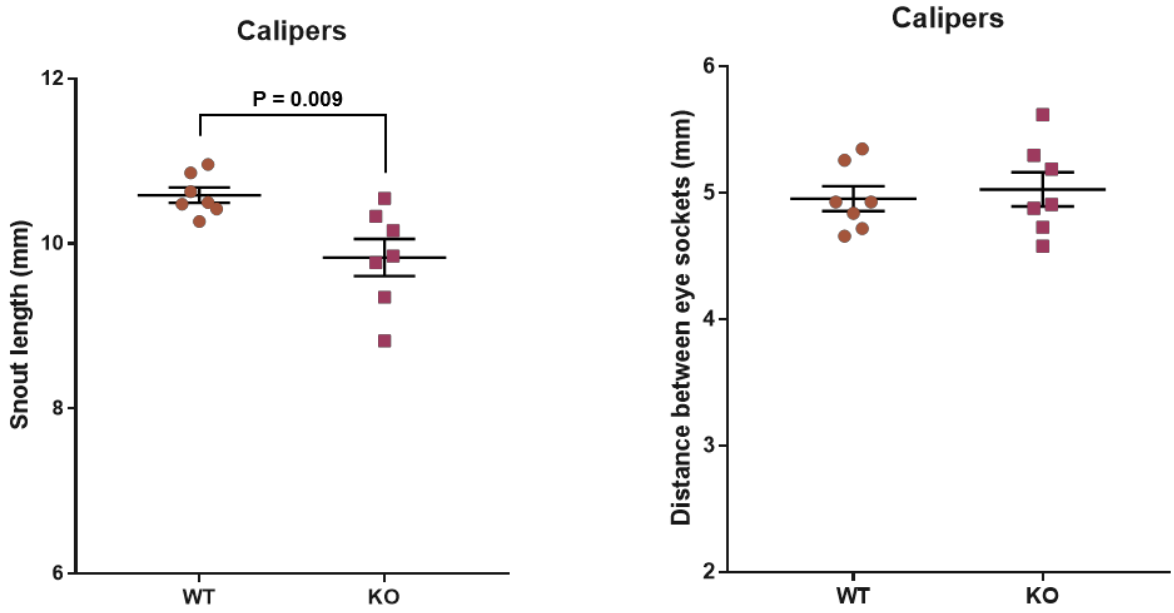


Figure S3: *Gpc*4 KO mice exhibit nasodigital deficits.

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.

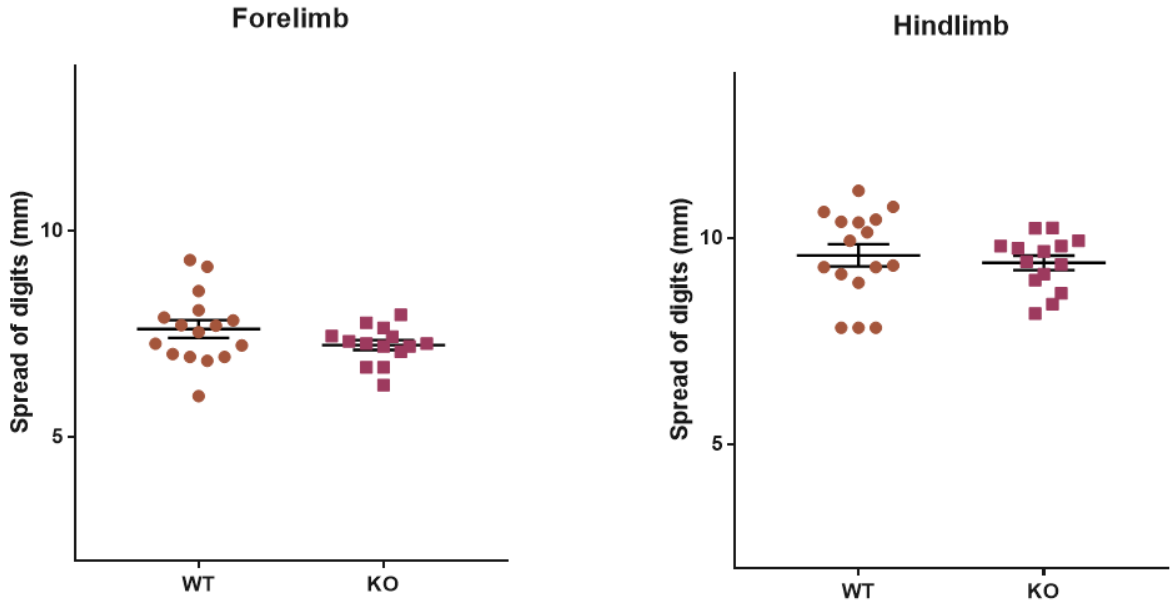


Figure S4: *Gpc4* KO mice do not display alterations in toe spread.

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.

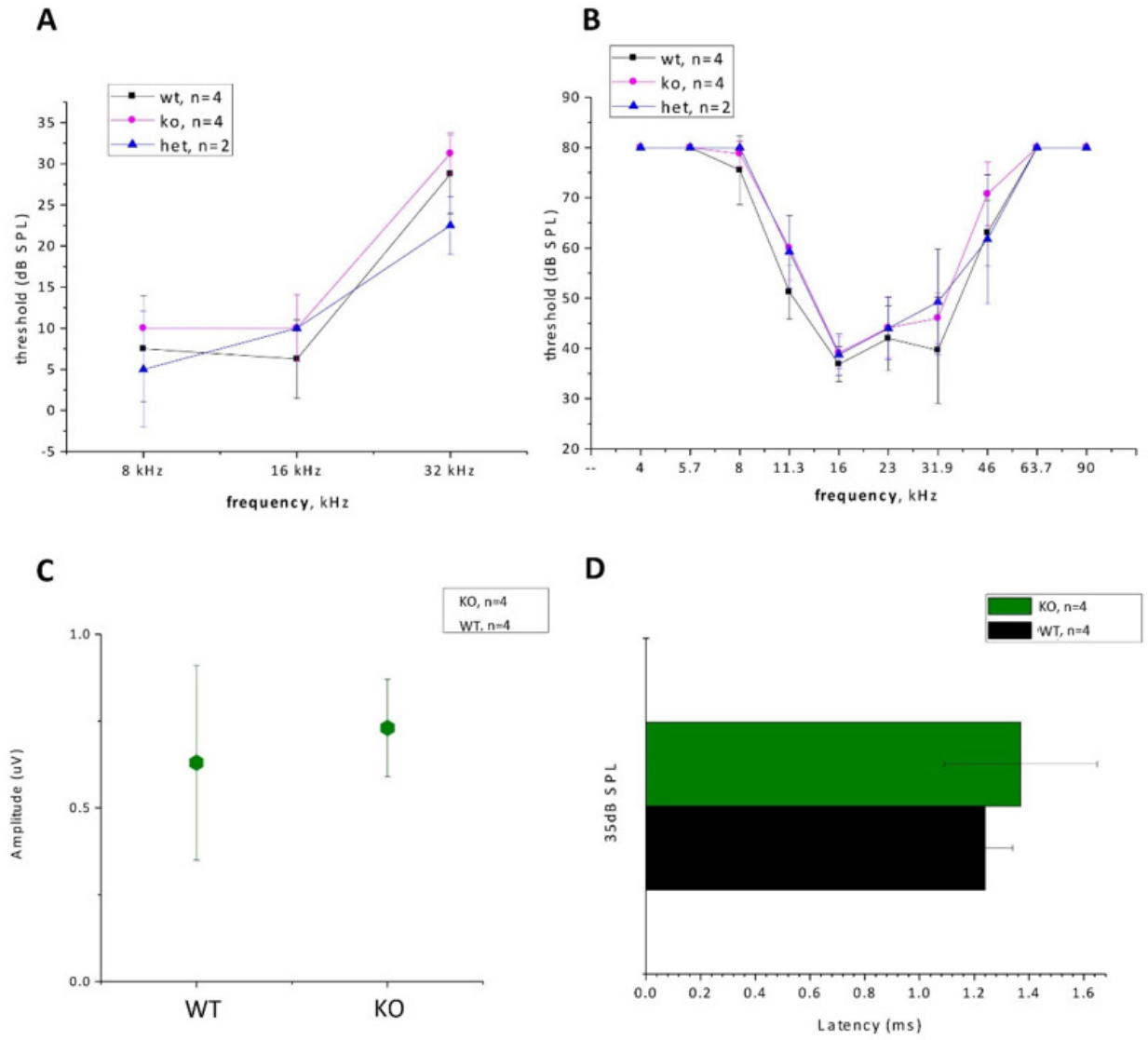


Figure S5: Assessment of auditory function in *Gpc4* KO mice by ABR and DPOAE. ABR (A) and DPOAE (B) thresholds are shown for mice at P18 of age. There were no significant differences in the thresholds at any frequency that were recorded in *Gpc4* KO mice compared to WT controls. ABR waveform amplitude 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.