Published in final edited form as:

J Med Genet. 2021 July ; 58(7): 442–452. doi:10.1136/jmedgenet-2020-106844.

OTITIS MEDIA SUSCEPTIBILITY AND SHIFTS IN THE HEAD AND NECK MICROBIOME DUE TO *SPINK5* VARIANTS

Daniel N. Frank[1], **Arnaud P.J. Giese**[2], **Lena Hafrén**[3], **Tori C. Bootpetch**[4], **Talitha Karisse L. Yarza**[5,6], **Matthew J. Steritz**[4], **Melquiadesa Pedro**[5], **Patrick John Labra**[5], **Kathleen A. Daly**[8], **Ma. Leah C. Tantoco**[5,6,9], **Wasyl Szeremeta**[10], **Maria Rina T. Reyes-Quintos**[5,6,7,9], **Niaz**

[*]**Corresponding Author:** Regie Lyn P. Santos-Cortez, Department of Otolaryngology-Head and Neck Surgery, School of Medicine, CU-AMC, 12700 E. 19th Ave., Research Complex 2, MS:8606, Aurora, Colorado 80045, USA; regie.santos-cortez@cuanschutz.edu; Tel: +1 (303) 724-0289.

Author Contributions: R.L.P.S.C. conceptualized the study. T.K.L.Y., M.P., P.J.L., M.L.C.T., M.R.T.R.Q., E.G.d.V.L., R.A.R.d.C., R.M.A.N., K.M.C.O., J.P.M.M., A.N.E.C., M.C.E.C., M.L.S.A., T.L.G.C., G.T.A., E.M.C.d.P., A.L.C., C.M.C. and R.L.P.S.C. collected clinical data, saliva samples and microbial swabs from the indigenous Filipino population and PGH patients. W.S., H.S.P. and T.C. collected clinical data and saliva samples from Texan families. L.H., E.E., J.K. and P.S.M. provided clinical data and DNA samples from Finnish families. K.A.D. and M.M.S. provided clinical data and DNA samples from Minnesota families. N.R.L. and K.L.M. provided DNA samples from the CLHNS cohort. T.C.B., M.J.S. and N.A. performed human and microbial DNA isolation, PCR and Sanger sequencing. A.P.J.G., S.Y., S.R. and Z.M.A. performed immunolocalization experiments and sequencing for Pakistani families. A.F.R. provided microarray expression data from mouse ME. The UW-CMG, M.J.B., D.A.N. and S.M.L. performed exome sequencing of DNA samples from indigenous Filipinos and 12 Minnesota families. D.N.F., D.I. and C.E.R. performed 16S rRNA sequencing and analyses on microbial samples. R.L.P.S.C. performed exome, linkage and statistical analyses for variant data. D.N.F. and R.L.P.S.C. wrote the manuscript. All authors read, provided critical input and approved the final version of the manuscript.

Competing Interests: Dr. Allen Ryan is a co-founder of, shareholder in and uncompensated consultant to Otonomy, Inc., a relationship that was approved by the University of California San Diego. The authors declare no competing interests.

Consortium: Members of the University of Washington Center for Mendelian Genomics are listed in http://uwcmg.org/docs/Crediting_UW-CMG/UW_CMG_Banner.pdf

Data Availability: Novel variants are being deposited in ClinVar. Demultiplexed paired-end 16S rRNA sequence data were deposited in the NCBI Short Read Archive under accession number PRJNA439435. RNA-sequence data from cholesteatoma and mucosal samples from OM patients [40] are available in dbGaP:phs001941.v1.p1.

Web Resources:
ANNOVAR, annovar.openbioinformatics.org/en/latest/
Burrows-Wheeler Aligner, bio-bwa.sourceforge.net/
Combined Annotation Dependent Depletion (CADD), cadd.gs.washington.edu
dbNSFP, sites.google.com/site/jpopgen/dbNSFP
Explicet, www.explicet.org
FastX-ToolKit v0.0.13, hannonlab.cshl.edu/fastx_toolkit
fathmm, fathmm.biocompute.org.uk/
featureCounts v1.5.2, bioinf.wehi.edu.au/featureCounts
Genetic Variant Interpretation Tool, www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/
Genome Analysis Toolkit, software.broadinstitute.org/gatk/
Genotype-Tissue Expression Portal, gtexportal.org
gnomAD, gnomad.broadinstitute.org/
InterPro, www.ebi.ac.uk/interpro
Likelihood Ratio Test, www.genetics.wustl.edu/jflab/lrt_query.html
MERLIN, csg.sph.umich.edu/abecasis/merlin/
MutationAssessor, mutationassessor.org/r3/
MutationTaster, www.mutationtaster.org
NCBI Short Read Archive, www.ncbi.nlm.nih.gov/sra
Online Mendelian Inheritance of Man, www.omim.org
PolyPhen2, genetics.bwh.harvard.edu/pph2/
PROVEAN, provean.jcvi.org
R, www.r-project.org
silva, SINA Alignment Service, www.arb-silva.de/aligner/
STAR v2.5.3a, github.com/alexdobin/STAR
Superlink, cbl-hap.cs.technion.ac.il/superlink-snp/
UCSC Genome Browser, genome.ucsc.edu

Ahankoob[4], **Erasmo Gonzalo d.V. Llanes**[5,6,9], **Harold S. Pine**[10], **Sairah Yousaf**[2], **Diana Ir**[1], **Elisabet Einarsdottir**[11,12], **Rhodieleen Anne R. de la Cruz**[9], **Nanette R. Lee**[13], **Rachelle Marie A. Nonato**[9], **Charles E. Robertson**[1], **Kimberly Mae C. Ong**[9], **Jose Pedrito M. Magno**[9], **Alessandra Nadine E. Chiong**[5,9], **Ma. Carmina Espiritu-Chiong**[14], **Ma. Luz San Agustin**[5,9], **University of Washington Center for Mendelian Genomics**, **Teresa Luisa G. Cruz**[5,9], **Generoso T. Abes**[5,9], **Michael J. Bamshad**[15], **Eva Maria Cutiongco-de la Paz**[7,16], **Juha Kere**[11,12,17], **Deborah A. Nickerson**[15], **Karen L. Mohlke**[18], **Saima Riazuddin**[2], **Abner L. Chan**[5,9], **Petri S. Mattila**[3], **Suzanne M. Leal**[19], **Allen F. Ryan**[20], **Zubair M. Ahmed**[2], **Tasnee Chonmaitree**[21], **Michele M. Sale**[22], **Charlotte M. Chiong**[5,6,9], **Regie Lyn P. Santos-Cortez**[4,5,23,*]

[1]Division of Infectious Diseases, Department of Medicine, School of Medicine, University of Colorado Anschutz Medical Campus (CU-AMC), Aurora, Colorado 80045, USA [2]Department of Otorhinolaryngology, Head and Neck Surgery, School of Medicine, University of Maryland, Baltimore, Maryland 21201, USA [3]Department of Otorhinolaryngology, Head & Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki FI-00029 HUS, Finland [4]Department of Otolaryngology-Head and Neck Surgery, School of Medicine, CU-AMC, Aurora, Colorado 80045, USA [5]Philippine National Ear Institute, University of the Philippines (UP) Manila, Manila 1000, Philippines [6]National Hearing Screening Reference Center, University of the Philippines (UP) Manila, Manila 1000, Philippines [7]National Institutes of Health – University of the Philippines (UP) Manila, Manila 1000, Philippines [8]Department of Otolaryngology, Head and Neck Surgery, University of Minnesota, Minneapolis, Minnesota 55455, USA [9]Department of Otorhinolaryngology, UP College of Medicine – Philippine General Hospital (PGH), Manila 1000, Philippines [10]Department of Otolaryngology, University of Texas Medical Branch (UTMB), Galveston, Texas 77555, USA [11]Folkhälsan Institute of Genetics and Molecular Neurology Research Program, University of Helsinki, Helsinki FI-00014, Finland [12]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge SE-141 83, Sweden [13]USC-Office of Population Studies Foundation, Inc. and Department of Anthropology, Sociology and History, University of San Carlos, Cebu City 6000, Philippines [14]Renergy Dermatology & Laser Center, Taguig City 1630, Philippines [15]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA [16]Philippine Genome Center, University of the Philippines, Quezon City 1101, Philippines [17]Department of Medical and Molecular Genetics, King's College London, London SE1 9RT, United Kingdom [18]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA [19]Center for Statistical Genetics, Gertrude H. Sergievsky Center, Taub Institute for Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, New York 10027, USA [20]Division of Otolaryngology, Department of Surgery, University of California San Diego School of Medicine and Veterans Affairs Medical Center, La Jolla, California 92093, USA [21]Division of Infectious Diseases, Department of Pediatrics, UTMB, Galveston, Texas 77555, USA [22]Center for Public Health Genomics, School of Medicine, Department of Biochemistry and Molecular Genetics, and Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia 22908, USA [23]Center for Children's Surgery, Children's Hospital Colorado, Aurora, Colorado 80045, USA

Abstract

Background: Otitis media (OM) susceptibility has significant heritability, however the role of rare variants in OM is mostly unknown. Our goal is to identify novel rare variants that confer OM susceptibility.

Methods: We performed exome and Sanger sequencing of >1,000 DNA samples from 551 multi-ethnic families with OM and unrelated individuals, RNA-sequencing, and microbiome sequencing and analyses of swabs from the outer ear, middle ear, nasopharynx and oral cavity. We also examined protein localization and gene expression in infected and healthy middle ear tissues.

Results: A large, intermarried pedigree that includes 81 OM-affected and 53 unaffected individuals co-segregates two known rare *A2ML1* variants, a common *FUT2* variant and a rare, novel pathogenic variant c.1682A>G (p.Glu561Gly) within *SPINK5* (LOD=4.09). Carriage of the *SPINK5* missense variant resulted in increased relative abundance of Microbacteriaceae in the middle ear, along with occurrence of Microbacteriaceae in the outer ear and oral cavity but not the nasopharynx. Eight additional novel *SPINK5* variants were identified in twelve families and individuals with OM. A role for *SPINK5* in OM susceptibility is further supported by lower RNA counts in variant carriers, strong SPINK5 localization in outer ear skin, faint localization to middle ear mucosa and eardrum, and increased *SPINK5* expression in human cholesteatoma.

Conclusion: *SPINK5* variants confer susceptibility to non-syndromic OM. These variants potentially contribute to middle ear pathology through breakdown of mucosal and epithelial barriers, immunodeficiency such as poor vaccination response, alteration of head and neck microbiota, and facilitation of entry of opportunistic pathogens into the middle ear.

Keywords

16S rRNA sequencing; exome sequencing; LEKTI; linkage analysis; Microbacteriaceae; microbiome; microbiota; otitis media; *SPINK5*

INTRODUCTION

Otitis media (OM) is a common childhood disease characterized by middle ear (ME) inflammation. OM affects almost half of children <1 year old.[1] Known risk factors for OM include lack of breastfeeding, allergies, upper respiratory infection, exposure to second-hand smoke exposure, low social status, daycare attendance, multiple siblings and family history.[2] Around 17% of patients require surgery for recurrent acute (RAOM) or chronic OM (COM), whether suppurative (CSOM) or with effusion (COME), with a 5-fold increased risk for OM surgery among first-degree relatives.[3] As a complex genetic trait, the heritability of OM ranges from 22–74% depending on OM type and cohort.[4–5] Previous genome-wide association studies identified common variant loci for OM, however these loci explain only a small proportion of OM heritability.[6–7]

Previously we identified two rare *A2ML1* (MIM 610627) variants (one duplication, one splice) and a common *FUT2* (MIM 182100) c.604C>T (p.Arg202*) variant in an intermarried, indigenous Filipino population (*N*~250) with a ~50% prevalence of non-syndromic OM (supplementary table S1).[8–11] Carriage of *A2ML1* and/or *FUT2* variants is associated with different OM types (supplementary table S1) and shifts in the ME microbiome,[8–12] e.g. carriers of the *A2ML1* c.2478_2485dupGGCTAAAT

(p.Ser829Trpfs*9) duplication variant had increased relative abundance of Fusobacteria in the ME.[12] Furthermore multiple *A2ML1* and *FUT2* variants were identified in US, Finnish and Pakistani families and patients with OM.[8–10] In particular, a *FUT2* stop variant c.461G>A (p.Trp154*) was associated with ME microbiome shifts among US patients with OM, and four *FUT2* variants decreased A antigen levels in epithelial cells.[9]

However in the indigenous Filipino population, we identified individuals with chronic, effusive, acute or healed OM who were wildtype for *A2ML1* or *FUT2* variants, including one individual who had available exome data (supplementary table S1). Rather than dismiss these individuals as phenocopies, we analyzed the exome data to determine additional rare variant(s) that may contribute to OM within the population. We then followed up our exome findings with expression and microbiome studies, as well as potential replication in additional families.

METHODS

Ethics approval

Human studies were approved by these institutional review boards (IRB): Bahauddin Zakariya University; Colorado Multiple IRB; Helsinki University Hospital; National Commission on Indigenous Peoples; University of Maryland Baltimore; University of Minnesota; University of the Philippines Manila; University of Texas Medical Branch; University of Virginia; University of Washington. Informed consent was obtained from all study participants. Mouse studies were approved by the Institutional Animal Care and Use Committees of the University of Maryland Baltimore and the Veterans Affairs Medical Center, San Diego.

Subject ascertainment

A total of 145 indigenous Filipinos provided clinical data and saliva samples for DNA isolation using Oragene kits (DNAgenotek). All individuals with various OM types were considered affected (figure 1; supplementary table S1). Also using DNA Oragene kits, saliva samples were collected from eleven patients from PGH and 166 Texan probands who underwent surgery for RAOM, COME or CSOM.

DNA samples were obtained from two established OM family cohorts, including 140 Minnesota and 217 Helsinki families.[5,7] All available relatives from Minnesota families were examined by an otolaryngologist, tested by tympanometry, and considered affected if OM-positive for ≥2 of the following: personal history, medical records, otoscopy, tympanometry. Data on OM history, risk factors and surgery were obtained from Finnish family members who were considered OM-affected if they had ventilation tubes, COME for >2 months, or RAOM (>3 episodes in 6 months or >4 episodes in 12 months).

For 16 Pakistani families, information on pedigree structure, comorbidities, OM onset and symptoms were obtained. Presence of ear discharge and air/bone conduction audiometry determined OM affection status. DNA was extracted from blood samples from participating family members. The different OM cohorts are further described in supplementary table S2.

DNA sequencing

Six DNA samples from indigenous Filipinos with OM were submitted for exome sequencing at the University of Washington Center for Mendelian Genomics (UW-CMG), as previously described.[8–9] Of these six individuals, IPOM-2 is wildtype for *A2ML1* or *FUT2* variants (figure 1).[8–10] Sequence capture was performed using either the Roche NimbleGen SeqCap EZ Human Exome v.2.0 or the Big Exome 2011 Library. Exome sequencing was performed (average depth ~60×) using an Illumina HiSeq. Burrows-Wheeler Aligner/BWA [13] and Genome Analysis Toolkit/GATK [14] were used to generate BAM and vcf files, respectively. Selected exome variants (table 1) were Sanger-sequenced in all available indigenous Filipino DNA samples.

DNA samples from selected OM-affected members of 12 Minnesota families were also exome-sequenced at UW-CMG. Additionally DNA samples of affected individuals from 16 Minnesota and 11 Finnish families plus 206 Finnish trio probands were exome-sequenced at the Northwest Genomics Center. From each of the 39 Minnesotan or Finnish families with multiple affected relatives, only two affected relatives were selected for exome sequencing, except for two families with exome data from 3–4 siblings each. For all available DNA samples from two Minnesota and four Finnish families, Sanger sequencing was performed for identified *SPINK5* variants (supplementary figure S1).

DNA samples from probands of 14 Pakistani families were exome-sequenced at the University of Maryland, as previously described.[15] Libraries were constructed using the Agilent SureSelect Human Expanded All Exon V5 kit, then sequenced on an Illumina HiSeq 2500 (~100× coverage). BWA and GATK were likewise used to generate BAM and vcf files.

Additional DNA samples from 112 Minnesota probands, 166 Texas probands, two Pakistani probands, and 11 PGH patients were Sanger-sequenced for the 33 coding exons of *SPINK5* (NM_006846). To estimate the Filipino minor allele frequency (MAF) of selected variants from the IPOM-2 exome and PGH patient data, DNA samples from unrelated Filipinos ($n \geq 88$) were obtained from the Cebu Longitudinal Health and Nutrition Survey (CLHNS) and Sanger-sequenced for these variants. The CLHNS DNA samples are from a community-based birth cohort recruited in 1983–1984 in order to study health and nutrition outcomes, but not OM.[16] Population-specific MAF for all other variants identified in non-Filipino cohorts were derived from gnomAD.

Bioinformatic, linkage and statistical analyses

Exome variants were annotated using ANNOVAR software [17] and the gnomAD, avsnp150 and dbNSFP33a [18] databases. Exome variants in the IPOM-2 exome were selected if: (a) passing all GATK quality control filters; (b) absent in gnomAD; (c) for indels, deemed disease-causing by MutationTaster;[19] (d) for single nucleotide variants, have a scaled CADD [20] score≥20 plus predicted damaging by ≥2 bioinformatics tools from dbNSFP.

Among the indigenous Filipinos, 134 ascertained individuals are connected by a single pedigree (figure 1). Two-point affecteds-only analysis was performed using Superlink software,[21] genotypes for *A2ML1, FUT2* and selected exome variants (table 1), and CLHNS MAF per variant. For each analysis by variant (table 1), if a wildtype affected

individual carries a pathogenic *A2ML1* or *FUT2* variant, the specific individual is considered a phenocopy. Parameters included autosomal dominant inheritance with 90% penetrance, 5% phenocopy rate and 1% disease MAF. Analyses were also performed with disease MAF equals variant MAF. In addition, logistic mixed models regression analysis was used to determine the relation between carriage of any pathogenic variant and OM status. Using the R function glmmPQL from the MASS package, age, sex and the number of variant copies carried per individual were analyzed as fixed effects, while membership by household or pedigree branch was included as random variable (supplementary table S1).

Because the indigenous Filipino population is unique [8] and not represented in gnomAD, a gnomAD MAF of zero was used for selecting variants from the IPOM-2 exome. However due to the finding of OM-pathogenic variants with higher MAF and non-deleterious bioinformatics prediction in other cohorts (e.g. *FUT2* c.461G>A has non-Finnish European MAF=0.47 and MutationTaster prediction as polymorphism [9]), *SPINK5* variants in the exome data from non-Filipino cohorts were selected with less stringent criteria i.e. scaled CADD>3, considered deleterious by ≥1 bioinformatics tool and MAF<0.01 in an ethnically matched gnomAD population [10].

RNA-sequencing of saliva samples

Because ME tissue samples were not available from indigenous Filipinos, we collected saliva samples for RNA-sequencing to determine potential variant effects.[10] Nine indigenous Filipinos (three wildtype and six heterozygous for the *SPINK5* variant; supplementary table S3) provided saliva samples using Oragene·RNA RE-100 kits. RNA samples (RIN=5.1–7.4) were processed with the NuGen Trio RNA-Seq Kit at the UCD-AMC Genomics and Microarray Core. Libraries were sequenced on an Illumina HiSeq 4000 generating 50-bp single-end reads, with samples pooled at equimolar concentrations and sequenced in a single lane. Reads were trimmed and adaptor sequences removed using the FASTX-Toolkit then aligned to the hg38 human genome using STAR.[22] Aligned reads were summarized at gene level using featureCounts.[23] Counts were normalized using log2-counts per million and analyzed by gene according to exome variant genotypes using limma [24].

16S rRNA gene sequencing and analysis

Outer ear (OE), ME, nasopharyngeal (NP), and oral cavity swabs were obtained from 90 indigenous Filipinos with and without OM (supplementary table S4) using Oragene P117 prototype kits. Microbial DNA isolation was performed using the Epicentre MasterPure™ Kit. Bacterial profiles were determined by broad-range PCR amplification and sequence analysis of the 16S rRNA gene V1-V2 regions, as previously described.[9] Illumina paired-end sequencing was performed on MiSeq using the 600 cycle version 3 kit. Assembled and quality-filtered sequences were aligned and classified with SINA (1.3.0-r23838) using the 418,497 bacterial sequences in Silva 115NR99.[25] Operational taxonomic units (OTUs) were produced by clustering sequences with identical taxonomic assignments (median 34,508 sequences/sample; interquartile range: 25,832–54,136). Goods coverage scores were ≥99% for all samples at the rarefaction point of 10,000 sequences, indicating adequate depth of sequence coverage for all samples.

Overall differences in microbial community composition between *SPINK5* genotypes were assessed by PERMANOVA tests using Morisita-Horn dissimilarity scores with 10^5–10^6 permutations. Differences in relative abundances of individual OTUs between groups were assessed by linear regression of center-log-ratio-transformed relative abundance data as outcome variables and *SPINK5* genotype as predictor variable. Standard measures of α-diversity, including richness (number of OTUs per sample estimated by S_{Chao1}), community evenness (uniformity of OTU distributions estimated by Shannon H/H_{max}), and complexity (Shannon diversity, H), were estimated at the rarefaction point of 10,000 sequences through 1,000 re-samplings. Results were then assessed by analysis of variance tests with *SPINK5* genotype as predictor. For PERMANOVA analyses, relative abundance tests, and α-diversity analyses, these were evaluated as covariates: age (by quartiles), sex, OM status, *FUT2* genotype, *A2ML1* genotype. Results were adjusted as follows, based on results of PERMANOVA analyses: OE and ME were adjusted by *A2ML1* genotype, while NP and oral cavity were adjusted by age (figure 2; figure 3; supplementary figure S2). R and Explicet v2.10.8 [26] were used for data display, analysis, and figure generation. Regression tests used the "car" R package and report *p*-values based on type II sums of squares. Principal coordinates analysis (PCoA) was utilized to explore and visualize data similarities in 2-dimensional space; distance matrices were calculated using the Morisita-Horn index (vegdist function of vegan R package) [27] and applied to the cmdscale function in R.

ME expression and immunolocalization

Transbullar ME inoculation was performed on wildtype 320 WBxB6 F1 hybrid mice with PBS (sham control) or non-typeable *H. influenzae* (NTHi), as previously described.[28] Two independent samples were generated for each time point after inoculation: 0 hours (0h, uninfected controls), 3h, 6h, 1 day (1d), 2d, 3d, 5d, and 7d, from OM initiation to resolution. ME mucosal mRNA was profiled on Affymetrix mouse 430 2.0 whole-genome microarrays. Out of 3,605 genes (~14.4% of the mouse genome) that defined the signature of acute OM across time, *SPINK5* expression in mouse ME was derived by comparing NTHi-infected versus healthy mice.

For immunolocalization in adult non-infected wildtype mouse ME, these primary antibodies were used: anti-SPINK5 (1:200, ab0138511, abcam); anti-b-catenin (1:1000, sc-7199, Santa Cruz Biotechnology). Mouse ears were harvested and fixed in 4% paraformaldehyde for 24h at 4°C, then decalcified using 0.25M EDTA (Promega) for 24h and incubated overnight with a 30% sucrose solution. Samples were processed for OCT embedding and kept at −80°C. OCT blocks were sectioned at 14μm thickness. Sections were permeabilized in 0.10% Triton X-100 for 1h, and blocked with 10% NGS in PBS for 1h. The tissue samples were probed with primary antibody overnight and, after three washes, were incubated with the secondary antibody for 1h at room temperature. Rhodamine Phalloidin (Invitrogen) was used at a 1:300 dilution for F-actin labeling. Nuclei were stained with DAPI (Molecular Probes). Sections were imaged on a Zeiss 710 confocal microscope using a 4× objective for low magnification or a 100×1.46 NA oil immersion objective. Stacks of confocal images were acquired with a Z step of 0.5μm and processed using ImageJ software.

SPINK5 localization was also determined in human cholesteatoma tissue. Paraffin sections were obtained from the Pathology Biorepository Shared Services, University of Maryland. Sections were deparaffinized using xylene and rehydrated gradually through graded alcohols. Sections were heat-treated at 95°C for 5min in a 10mM sodium citrate buffer, then processed for immunostaining.

RESULTS

Exome variants in IPOM-2 and salivary RNA levels

Of six indigenous individuals with current or previous CSOM and exome data, one child who is *A2ML1-/FUT2*-wildtype carried multiple, rare variants that were absent in gnomAD (table 1). Of the 16 rare damaging missense or stop variants (CADD≥22) that were present in other indigenous exomes but rare in the general Filipino population, a genome-wide significant LOD score of >3.3 was obtained for seven variants (table 1). When LOD scores were re-calculated with disease MAF equal to variant MAF, six variants with LOD>3.3 remained, namely: *SPINK5* c.1682A>G (p.(Glu561Gly)), 3.91; *TMEM270* c.670C>T (p.(Gln224*)), 5.62; *SIDT2* c.1249G>T (p.(Asp417Tyr)), 4.32; *JMJD7* c.553C>G (p.(Leu185Val)), 3.83; *CDH19* c.794T>C (p.(Val265Ala)), 6.75; and *ZNRF3* c.2522G>A (p.(Gly841Asp)), 4.88.

Salivary RNA-sequence data from nine indigenous individuals were analyzed based on variant carriage (table 1). Of the variants with LOD>3.3, only the *SPINK5, SAT2* and *ZNRF3* variants have some effect on RNA expression (table 1). When corrected for *A2ML1* genotypes, nominal *p*-values for the gene expression for each gene were as follows: *SPINK5*=0.08, *SAT2*=0.08, and *ZNRF3*=0.02 (table 1).

Changes in microbiota due to the *SPINK5* c.1682A>G variant

SPINK5 was previously shown as expressed in skin, tonsils, and in sinonasal and oral mucosae,[29–31] consistent with findings in the Genotype-Tissue Expression database. Because *SPINK5* variants are associated with infection risk,[32–33] we hypothesized that the *SPINK5* c.1682A>G (p.(Glu561Gly)) variant may influence the commensal microbiota of various head and neck sites. We collected swabs of the ME, OE, NP and oral cavity of indigenous Filipinos (supplementary table S4) and submitted isolated bacterial DNA for 16S rRNA gene sequencing. As expected, significant differences in overall microbiota composition (i.e., β-diversity) were evident across sampling sites ($p<1\times10^{-6}$; supplementary figure S2), as well as between all pairs of sites ($p<1\times10^{-6}$) other than OE versus ME ($p=0.78$). Bivariate analyses (i.e., including sample site as a covariate) also indicated that age ($p=0.001$), but not sex (p=0.26) or OM (p=0.44), was associated with differences in microbiota across sites.

Analyzing β-diversity for individual sampling sites, age was highly significant in univariable analyses (supplementary table S5) for NP ($p=0.0009$) and oral cavity ($p=0.0004$), trended towards significance for OE ($p=0.08$), and was not significant for ME ($p=0.82$). Age remained significant only for NP and oral cavity in multivariable models (supplementary table S5; supplementary table S6). Occurrence of OM was associated with altered

microbiota in the NP (p=0.09 and p=0.047 for univariable and multivariable models, respectively (supplementary table S5; supplementary table S6), but not in the oral cavity (p=0.64). OM-microbiota associations could not be assessed in OE or ME because samples were available only from those with eardrum perforations (i.e., no non-OM control samples). Among those with OM, the oral cavity and NP microbiotas differed significantly from OE and ME microbiotas (pairwise $p < 1 \times 10^{-6}$; supplementary figure S3), suggesting there was no overall convergence in microbiota between anatomical sites with OM occurrence.

SPINK5 genotype was associated with altered microbiota composition only within the oral cavity (p=0.02 in multivariable models; figure 2; supplementary table S6). On the other hand, *A2ML1* genotype was significantly associated with altered microbiota composition in the OE (p=0.02) and the ME (p=0.02) in multivariable models, but not NP or oral cavity (supplementary table S6). No significant differences were noted for *FUT2*, *SAT2*, or *ZNFR3* genotypes in either univariable analyses or multivariable analyses adjusting for age, sex, OM, or other genotypes (supplementary table S5; supplementary table S6). Measures of microbial diversity (e.g., S_{Chao1}, Shannon diversity, Shannon evenness) did not differ significantly with *SPINK5* genotype or OM (data not shown).

Numerous bacterial taxa differed in relative abundance in association with *SPINK5* genotype (figure 3A–3D) and OM occurrence (figure 3E–3F), with results varying by anatomical site. Interestingly most taxa that differed by *SPINK5* genotype were of higher abundance in carriers of the *SPINK5* variant (figure 3). In contrast, similar numbers of taxa were either elevated or reduced in abundance in OM versus non-OM. The genera that varied with *SPINK5* genotype or OM status represented a variety of phyla, including Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, Spirochetes and Tenericutes, suggesting that both *SPINK5* genotype and OM resulted in broad-level differences in microbiota. However there was little overlap in the sets of genera that were altered across sites (figure 3A–3D). Only Microbacteriaceae differed significantly with *SPINK5* genotype in more than one anatomical site (OE, ME, and oral cavity). This might suggest that OE and oral cavity, rather than NP, might serve as reservoirs for ME-colonizing bacteria within the context of *SPINK5* variant carriage, e.g. the ME may be seeded with Microbacteriaceae from the OE through an eardrum perforation from previous OM. Within the NP, the genera *Parvimonas* and *Haemophilus* were enriched in *SPINK5* variant carriers and those with OM (figure 3C, 3E). Only *Lactobacillus* was enriched in the oral cavity in association with *SPINK5* genotype and OM.

Previously we identified Fusobacteria in the OE and ME as being nominally associated with *A2ML1* genotype.[12] Re-analyses using the current data showed that in the OE the genus Fusobacterium was associated with *A2ML1* genotype (p=0.02), but was non-significant after correction for sex or age. On the other hand, in the ME the genus Leptotrichia from the phylum Fusobacteria was significantly associated with *A2ML1* genotype (FDR-adjusted p=0.03), even after adjusting for age, sex or *SPINK5* genotype. Interestingly, Leptotrichia is associated with *SPINK5* genotype in the NP (figure 3C). Although Leptotrichia is commonly abundant in the oropharynx,[34] it has also been reported in adenoids and ME fluid from children with COME.[35]

Phenotypic description of *SPINK5* variant carriers

Individual IPOM-2 is *A2ML1-/FUT2*-wildtype but homozygous for a novel rare, missense variant c.1682A>G (p.(Glu561Gly)) within *SPINK5* (table 1). Three other affected individuals with exome data were also heterozygous for the same *SPINK5* missense variant (figure 1; table 1). Sanger sequencing of DNA samples from the CLHNS cohort showed that the *SPINK5* c.1682A>G (p.(Glu561Gly)) variant is rare in the general Filipino population with MAF=0.004 (table 1).

Homozygous or compound heterozygous variants within *SPINK5* (MIM 605010), a large majority of which are loss-of-function variants, are known to cause Comèl-Netherton syndrome (MIM 526500), a rare autosomal recessive disorder characterized by distinct skin and hair abnormalities.[32–33] Dermatologic examination of IPOM-2 who is homozygous for the *SPINK5* missense variant c.1682A>G (p.(Glu561Gly)), her mother who is heterozygous (supplementary figure S4), and four additional heterozygous carriers ruled out the known clinical features of Comèl-Netherton syndrome. In particular, IPOM-2 is homozygous for the *SPINK5* variant and has a history of left-sided CSOM which healed after medical treatment with Ofloxacin otic drops. Although IPOM-2 had a previous history of asthma, eczema and skin dryness, dermatologic examination revealed normal skin and hair and none of the known clinical features of Comèl-Netherton syndrome.

Sanger sequencing of 139 additional DNA samples revealed that 84 of 87 affected indigenous individuals carried an *A2ML1, FUT2* and/or *SPINK5* variant, suggesting that carriage of any of these variants is associated with OM (p=0.004; supplementary table S1). In addition, two-point linkage analysis resulted in LOD scores (θ=0) of 6.30 for the *A2ML1* duplication variant, LOD=2.83 for the *FUT2* variant, and LOD=4.09 for the *SPINK5* variant (figure 1). Based on guidelines from the American College of Medical Genetics,[36] the *SPINK5* variant c.1682A>G (p.(Glu561Gly)) is classified as pathogenic and likely contributes to OM susceptibility.

Additional *SPINK5* variants in OM-affected individuals

To further identify *SPINK5* variants that play a role in OM susceptibility, additional DNA samples from US, Finnish and Pakistani families were submitted for exome and Sanger sequencing. A rare, damaging *SPINK5* variant c.802C>T (p.(Arg268Cys)) co-segregates with OM in a Minnesota family in an autosomal dominant pattern, and is also heterozygous in two Texan probands and three children from two Finnish families, all of whom have both RAOM and COME (table 2; supplementary figure S1). However in Finnish family 34, incomplete co-segregation of the *SPINK5* c.802C>T (p.(Arg268Cys)) variant with OM suggests intra-familial genetic heterogeneity (supplementary figure S1).[37] Furthermore affected individuals from four Finnish families (table 2) were heterozygous for rare, damaging missense variants c.551G>A (p.(Gly184Asp)), c.2243A>G (p.(Glu748Gly)), c.2493T>A (p.(Asn831Lys)) and c.2974A>T (p.(Met992Leu)). The stop variant c.652C>T (p.Arg218*) was previously identified as homozygous or compound heterozygous with a loss-of-function variant in individuals with Comèl-Netherton syndrome.[33] This variant was heterozygous in a Finnish mother-proband pair who did not exhibit syndromic features but had RAOM (with the child also having COME).

Additionally we identified a recurrent, rare damaging variant *SPINK5* c.1345C>A (p.(Leu449Ile)) in two out of 11 PGH patients (table 2). One Texas proband is heterozygous for a 3'UTR variant c.*74A>G that is absent in gnomAD and predicted by MutationTaster to induce splice defects (table 2), however we have no available data to test potential epigenetic effects of this variant. No rare or low-frequency, potentially damaging *SPINK5* variant was identified in the 16 Pakistani families. In summary, we found eight novel *SPINK5* variants that are mostly heterozygous in individuals with non-syndromic OM from different ethnic groups (table 2). However these variants remain of unknown significance until replicated further for OM susceptibility.[36]

SPINK5 is strongly localized to the OE more than ME mucosa

In wildtype mouse ears, SPINK5 was localized strongly to OE skin, but was also present though faintly staining in ME mucosa and eardrum (figure 4). There was no increased SPINK5 staining in paraffinized human cholesteatoma (supplementary figure S5). There was also no significant *SPINK5* regulation in mouse ME mucosa following acute NTHi infection based on previous microarray data (data not shown).[28] However our previous RNA-sequencing study showed that *SPINK5* was upregulated in fresh human cholesteatoma tissue compared to ME mucosa (DESeq2 nominal p=0.009; adj-p=0.07).[40] Unfortunately we have no RNA data from eardrum tissue.

DISCUSSION

In this report we investigated (A) the effect of carriage of a novel rare, pathogenic missense *SPINK5* variant on gene expression and bacterial profiles at various sites of the head and neck including the ME, and (B) the occurrence of *SPINK5* variants in families and individuals with OM. We also examined SPINK5 localization and gene expression in mouse ME. Our analyses of multi-ethnic families and individuals with OM identified rare single nucleotide variants within *SPINK5*, including a novel pathogenic missense variant c.1682A>G (p.(Glu561Gly)) that conferred susceptibility to non-syndromic OM and increased relative abundance of Microbacteriaceae in the ME (figure 1; figure 3). We also observed lower *SPINK5* expression in saliva of variant carriers (table 1; supplementary table S3) and higher expression in cholesteatoma tissue [40]. In our previous RNA-sequencing study, *SPINK5* was co-upregulated with *A2ML1* in saliva of OM patients.[10] Both *A2ML1* and *SPINK5* encode serine protease inhibitors. We hypothesized that defects in serine protease inhibitors lead to ME mucosal damage due to proteases secreted by bacteria or inflammatory cells.[8] Previously a *SPINK5* missense variant c.1258A>G (p.Glu420Lys; overall gnomAD MAF=0.49) was reported to increase protease activity and breakdown of proteins that are essential for formation of epithelial barriers.[41] The rare p.Glu561Gly variant occurs between LEKTI Kazal-type domains D8-D9, which are cleaved from SPINK5 protein in order to potently inhibit kallikreins and maintain the balance required for epidermal lipid barrier integrity.[33] In OM, kallikrein activity was shown to be higher in serous than in mucoid effusions in the ME.[42] Interestingly *SPINK5* variant carriers had RAOM and/or COME, while indigenous Filipino variant carriers with eardrum perforations were noted to have mucoid or mucopurulent discharge. Taken together, these findings suggest that SPINK5 variants affect inhibition of kallikrein activity and increase

predisposition to mucoid or mucopurulent effusions in the ME as part of RAOM, COME or CSOM.

Moreover deficiency in lympho-epithelial Kazal-type 5 related inhibitor (LEKTI) protein due to autosomal recessive *SPINK5* variants leads to Comèl-Netherton syndrome that includes skin barrier defects and increased frequency of infections such as recurrent OM.[32–33] Among families with Comèl-Netherton syndrome, 92% had skin and hair phenotypes (e.g. trichorrhexis invaginata, scaly erythroderma, ichthyosis linearis circumflexa) due to diallelic, rare, loss-of-function *SPINK5* variants.[33] In patients with OM as part of Comèl-Netherton syndrome, primary immunodeficiency with increased proinflammatory cytokine levels, decreased natural killer cell cytotoxicity, reduced memory B cells and defective response to pneumococcal vaccination was documented but the immune condition improved after immunoglobulin therapy.[32] In contrast, the families and individuals reported here do not have syndromic features and carry mostly missense variants that are heterozygous or with autosomal dominant inheritance, suggesting phenotypic heterogeneity due to rare variant allelic heterogeneity. Even though we present non-syndromic OM due to rare *SPINK5* variants as a distinct phenotype, treatment modalities designed for Comèl-Netherton syndrome might be modified for genotype-based OM treatment. Therapies suggested for SPINK5 defects are calcium, kallikrein inhibition, phototherapy and gene therapy.[43–46] *SPINK5* variant carriers generally respond to standard OM therapies such as antibiotic otic drops and surgery (supplementary table S2), however we expect some OM cases to be recurrent or antibiotic-resistant.

Microbacteriaceae had elevated relative abundance in the ME, and also occurred in the OE and oral cavities, but not in the NP of *SPINK5* variant carriers (figure 3). This might suggest that bacterial taxa gain entry into the ME through the OE and an eardrum perforation, usually from previous OM e.g. previous NTHi-induced OM. Within the indigenous Filipino community, ME pathogens are potentially acquired via swimming in contaminated seawater e.g. by wastewater effluent containing Microbacteriaceae/Actinobacterium.[47] Because SPINK5 is localized to skin, the OE acting as a reservoir for ME pathogens might be expected in *SPINK5* variant carriers. OM due to *Pseudoclavibacter* which belongs to the Microbacteriaceae family is rare but was previously reported in an immunocompromised child.[48] Microbacteriaceae species were also identified as antibiotic-resistant (e.g. quinolones) or as opportunistic pathogens that flourish in the ME, skin wounds, lungs and gastrointestinal tracts of immunocompromised hosts.[49–51] Species-level determination will help guide antibiotic treatment, however our 16S rRNA sequence data does not allow species-level identification.

SPINK5 genotype was associated with differences in microbiota composition of various anatomical sites including the ME (figures 2 and 3). Altered microbiota may contribute to OM by diminishing the ability of commensal microbiota to compete with infectious agents or support immune homeostasis.[52–53] However our cross-sectional study design precluded determination of whether differences in microbiota arose before or after OM. Additionally false-discovery rate correction by the Benjamini-Hochberg method resulted in $p>0.05$ for all tests of individual taxa. Indeed, the substantial inter-subject variability in microbiota observed within each anatomic site (supplemental figure S2), coupled with

sample size limitation, and occurrence of multiple variants, likely decreased the power of statistical tests to discern between-group differences in α-diversity, β-diversity, or individual taxon abundance. Follow-up of these exploratory findings will require expanded longitudinal, taxa-specific or animal model studies.

To conclude, *SPINK5* variants confer susceptibility to non-syndromic OM in multi-ethnic families and individuals. These variants potentially contribute to ME pathology through breakdown of mucosal and epithelial barriers, immunodeficiency such as poor vaccination response, alteration of head and neck microbiota, and facilitation of entry into the ME of opportunistic pathogens through a perforated eardrum from a previous OM episode.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank all study participants. We also thank: E Valdivia of UTMB and SM Lagrana-Villagracia of PNEI for data collection efforts; E Leinonen for collection and handling of Finnish samples; the Office of Population Studies Foundation research and data collection teams; I Abbe for additional Sanger sequencing; and J Cardwell for pre-processing of RNA-seq data. This work is dedicated to Dr. Michele M. Sale, outstanding scientist, mentor and friend, who passed away during the conduct of this study.

Funding: The Cebu Longitudinal Health and Nutrition Survey was supported by NIH grants DK078150, TW005596, HL085144 and TW008288 and pilot funds from RR020649, ES010126 and DK056350. Exome sequencing of DNA samples from the indigenous Filipino population and 12 Minnesota families was provided by the UW-CMG and was supported by the US NIH-National Human Genome Research Institute and the National Heart, Lung and Blood Institute grants UM1 HG006493 and U24 HG008956 (to D.A.N., M.J.B. and S.M.L.). This study was supported by the Philippine Council for Health Research and Development - Department of Science and Technology Balik Scientist Program and the US NIH – National Institute on Deafness and Other Communication Disorders grant R01 DC015004 (both to R.L.P.S.C.).

References:

1. Chonmaitree T, Trujillo R, Jennings K, Alvarez-Fernandez P, Patel JA, Loeffelholz MJ, Nokso-Koivisto J, Matalon R, Pyles RB, Miller AL, McCormick DP. Acute otitis media and other complications of viral respiratory infection. Pediatrics 2016;137:e20153555. [PubMed: 27020793]
2. Zhang Y, Xu M, Zhang J, Zeng L, Wang Y, Zheng QY. Risk factors for chronic and recurrent otitis media – a meta-analysis. PLoS One 2014;9:e86397. [PubMed: 24466073]
3. Padia R, Alt JA, Curtin K, Muntz HR, Orlandi RR, Berger J, Meier JD. Familial link of otitis media requiring tympanostomy tubes. Laryngoscope 2017;127:962–6. [PubMed: 27861935]
4. Casselbrant ML, Mandel EM, Fall PA, Rockette HE, Kurs-Lasky M, Bluestone CD, Ferrell RE. The heritability of otitis media: a twin and triplet study. JAMA 1999;282;2125–30. [PubMed: 10591333]
5. Hafrén L, Kentala E, Järvinen TM, Leinonen E, Onkamo P, Kere J, Mattila PS. Genetic background and the risk of otitis media. Int J Pediatr Otorhinolaryngol 2012;76;41–4. [PubMed: 22018929]
6. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, Hinds DA. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun 2017:8;599. [PubMed: 28928442]
7. Allen EK, Chen WM, Weeks DE, Chen F, Hou X, Mattos JL, Mychaleckyj JC, Segade F, Casselbrant ML, Mandel EM, Ferrell RE, Rich SS, Daly KA, Sale MM. A genome-wide association study of chronic otitis media with effusion and recurrent otitis media identified a novel susceptibility locus on chromosome 2. J Assoc Res Otolaryngol 2013;14:791–800. [PubMed: 23974705]

8. Santos-Cortez RLP, Chiong CM, Reyes-Quintos MRT, Tantoco MLC, Wang X, Acharya A, Abbe I, Giese AP, Smith JD, Allen EK, Li B, Cutiongco-de la Paz EM, Garcia MC, Llanes EGdV, Labra PJ, Gloria-Cruz TLI, Chan AL, Wang GT, Daly KA, Shendure J, Bamshad MJ, Nickerson DA, Patel JA, Riazuddin S, Sale MM, University of Washington Center for Mendelian Genomics, Chonmaitree T, Ahmed ZM, Abes GT, Leal SM. Rare *A2ML1* variants confer susceptibility to otitis media. Nat Genet 2015;47;917–20. [PubMed: 26121085]

9. Santos-Cortez RLP, Chiong CM, Frank DN, Ryan AF, Giese APJ, Bootpetch Roberts T, Daly KA, Steritz MJ, Szeremeta W, Pedro M, Pine HS, Yarza TKL, Scholes MA, Llanes EGdV, Yousaf S, Friedman NR, Tantoco MLC, Wine TM, Labra PJ, Benoit J, Ruiz AG, de la Cruz RAR, Greenlee C, Yousaf A, Cardwell J, Nonato RMA, Ray D, Ong KMC, So E, Robertson CE, Dinwiddie J, Lagrana-Villagracia SM, University of Washington Center for Mendelian Genomics, Gubbels SP, Shaikh RS, Cass SP, Einarsdottir E, Lee NR, Schwartz DA, Gloria-Cruz TLI, Bamshad MJ, Yang IV, Kere J, Abes GT, Prager JD, Riazuddin S, Chan AL, Yoon PJ, Nickerson DA, Cutiongco-de la Paz EM, Streubel SO, Reyes-Quintos MRT, Jenkins HA, Mattila PS, Chan KH, Mohlke KL, Leal SM, Hafrén L, Chonmaitree T, Sale MM, Ahmed ZM. *FUT2* variants confer susceptibility to familial otitis media. Am J Hum Genet 2018;103;679–90. [PubMed: 30401457]

10. Larson ED, Magno JPM, Steritz MJ, Llanes EGdV, Cardwell J, Pedro M, Bootpetch Roberts T, Einarsdottir E, Rosanes RAQ, Greenlee C, Santos RAP, Yousaf A, Streubel SO, Santos ATR, Ruiz AG, Lagrana-Villagracia SM, Ray D, Yarza TKL, Scholes MA, Anderson CB, Acharya A, University of Washington Center for Mendelian Genomics, Gubbels SP, Bamshad MJ, Cass SP, Lee NR, Shaikh RS, Nickerson DA, Mohlke KL, Prager JD, Cruz TLG, Yoon PJ, Abes GT, Schwartz DA, Chan AL, Wine TM, Cutiongco-de la Paz EM, Friedman NR, Kechris K, Kere J, Leal SM, Yang IV, Patel JA, Tantoco MLC, Riazuddin S, Chan KH, Mattila PS, Reyes-Quintos MRT, Ahmed ZM, Jenkins HA, Chonmaitree T, Hafrén L, Chiong CM, Santos-Cortez RLP. *A2ML1* and otitis media: novel variants, differential expression, and relevant pathways. Hum Mutat 2019;40:1156–71. [PubMed: 31009165]

11. Santos-Cortez RLP, Reyes-Quintos MRT, Tantoco MLC, Abbe I, Llanes EGdV, Ajami NJ, Hutchinson DS, Petrosino JF, Padilla CD, Villarta RL Jr, Gloria-Cruz TL, Chan AL, Cutiongco-de la Paz EM, Chiong CM, Leal SM, Abes GT. Genetic and environmental determinants of otitis media in an indigenous Filipino population. Otolaryngol Head Neck Surg 2016;155:856–62. [PubMed: 27484237]

12. Santos-Cortez RLP, Hutchinson DS, Ajami NJ, Reyes-Quintos MRT, Tantoco MLC, Labra PJ, Lagrana SM, Pedro M, Llanes EGdV, Gloria-Cruz TLI, Chan AL, Cutiongco-de la Paz EM, Belmont JW, Chonmaitree T, Abes GT, Petrosino JF, Leal SM, Chiong CM. Middle ear microbiome differences in indigenous Filipinos with chronic otitis media due to a duplication in the *A2ML1* gene. Infect Dis Poverty 2016;5;97. [PubMed: 27799062]

13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60. [PubMed: 19451168]

14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303. [PubMed: 20644199]

15. Riazuddin S, Hussain M, Razzaq A, Iqbal Z, Shahzad M, Polla DL, Song Y, van Beusekom E, Khan AA, Tomas-Roca L, Rashid M, Zahoor MY, Wissink-Lindhout WM, Basra MAR, Ansar M, Agha Z, van Heeswijk K, Rasheed F, Van de Vorst M, Veltman JA, Gilissen C, Akram J, Kleefstra T, Assir MZ, UK10K, Grozeva D, Carss K, Raymond FL, O'Connor TD, Riazuddin SA, Khan SN, Ahmed ZM, de Brouwer APM, van Bokhoven H, Riazuddin S. Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. Mol Psychiatry 2017;22:1604–14. [PubMed: 27457812]

16. Adair LS, Popkin BM, Akin JS, Guilkey DK, Gultiano S, Borja J, Perez L, Kuzawa CW, McDade T, Hindin MJ. Cohort profile: the Cebu longitudinal health and nutrition survey. Int J Epidemiol 2011;40:619–25. [PubMed: 20507864]

17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;28:e164.

18. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum Mutat 2016;37:235–41. [PubMed: 26555599]
19. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575–6. [PubMed: 20676075]
20. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5. [PubMed: 24487276]
21. Fishelson M, Geiger D. Exact genetic linkage computations for general pedigrees. Bioinformatics 2002;18 Suppl 1:S189–98. [PubMed: 12169547]
22. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21. [PubMed: 23104886]
23. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30. [PubMed: 24227677]
24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47. [PubMed: 25605792]
25. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41:D590–6. [PubMed: 23193283]
26. Robertson CE, Harris JK, Wagner BD, Granger D, Browne K, Tatem B, Feazel LM, Park K, Pace NR, Frank DN. Explicet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. Bioinformatics 2013;29:3100–1. [PubMed: 24021386]
27. Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MH, Szoecs E, Wagner H. Vegan: Community Ecology Package. R package version 2.5–4. http://vegan.r-forge.r-project.org (accessed March 2020).
28. Hernandez M, Leichtle A, Pak K, Webster NJ, Wasserman SI, Ryan AF. The transcriptome of a complete episode of acute otitis media. BMC Genomics 2015;16:259. [PubMed: 25888408]
29. Mägert HJ, Ständker L, Kreutzmann P, Zucht HD, Reinecke M, Sommerhoff CP, Fritz H, Forssmann WG. LEKTI, a novel 15-domain type of human serine proteinase inhibitor. J Biol Chem 1999;274:21499–502. [PubMed: 10419450]
30. Ong C, O'Toole EA, Ghali L, Malone M, Smith VV, Callard R, Harper JI. LEKTI demonstrable by immunohistochemistry of the skin: a potential diagnostic skin test for Netherton syndrome. Br J Dermatol 2004;151:1253–7. [PubMed: 15606522]
31. Richer SL, Truong-Tran AQ, Conley DB, Carter R, Vermylen D, Grammer LC, Peters AT, Chandra RK, Harris KE, Kern RC, Schleimer RP. Epithelial genes in chronic rhinosinusitis with and without nasal polyps. Am J Rhinol 2008:22;228–34. [PubMed: 18588753]
32. Renner ED, Hartl D, Rylaarsdam S, Young ML, Monaco-Shawver L, Kleiner G, Markert ML, Stiehm ER, Belohradsky BH, Upton MP, Torgerson TR, Orange JS, Ochs HD. Comel-Netherton syndrome – defined as primary immunodeficiency. J Allergy Clin Immunol 2009;124:536–43. [PubMed: 19683336]
33. Sarri CA, Roussaki-Schulze A, Vasilopoulos Y, Zafiriou E, Patsatsi A, Stamatis C, Gidarokosta P, Sotiriadis D, Sarafidou T, Mamuris Z. Netherton syndrome: a genotype-phenotype review. Mol Diagn Ther 2017;21:137–52. [PubMed: 27905021]
34. Wang H, Dai W, Feng X, Zhou Q, Wang H, Yang Y, Li S, Zheng Y. Microbiota composition in upper respiratory tracts of healthy children in Shenzhen, China, differed with respiratory sites and ages. Biomed Res Int 2018;2018:6515670. [PubMed: 30013985]
35. Chan CL, Wabnitz D, Bardy JJ, Bassiouni A, Wormald PJ, Vreugde S, Psaltis AJ. The microbiome of otitis media with effusion. Laryngoscope 2016;126:2844–51. [PubMed: 27335217]
36. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus

recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24. [PubMed: 25741868]

37. Rehman AU, Santos-Cortez RLP, Drummond MC, Shahzad M, Lee K, Morell RJ, Ansar M, Jan A, Wang X, Aziz A, Riazuddin S, Smith JD, Wang GT, Ahmed ZM, Gul K, Shearer AE, Smith RJ, Shendure J, Bamshad MJ, Nickerson DA, University of Washington Center for Mendelian Genomics, Hinnant J, Khan SN, Fisher RA, Ahmad W, Friderici KH, Riazuddin S, Friedman TB, Wilch ES, Leal SM. Challenges and solutions for gene identification in the presence of familial locus heterogeneity. Eur J Hum Genet 2015;23:1207–15. [PubMed: 25491636]
38. Maniu A, Harabagiu O, Perde Schrepler M, Cătană A, Fănuţă B, Mogoantă CA. Molecular biology of cholesteatoma. Rom J Morphol Embryol 2014;55:7–13. [PubMed: 24715159]
39. Lim DJ, Saunders WH. Acquired cholesteatoma: light and electron microscopic observations. Ann Otol Rhinol Laryngol 1972;81:1–11. [PubMed: 5009814]
40. Baschal EE, Larson ED, Bootpetch Roberts T, Pathak S, Frank G, Handley E, Dinwiddie J, Moloney M, Yoon PJ, Gubbels SP, Scholes MA, Cass SP, Jenkins HA, Frank DN, Yang IV, Schwartz DA, Ramakrishnan V, Santos-Cortez RLP. Identification of novel genes and biological pathways that overlap in infectious and non-allergic diseases of the upper and lower airways using network analyses. *Front Genet* 2020;10:1352.
41. Fortugno P, Furio L, Teson M, Berretti M, El Hachem M, Zambruno G, Hovnanian A, D'Alessio M. The 420K LEKTI variant alters LEKTI proteolytic activation and results in protease deregulation: implications for atopic dermatitis. Hum Mol Genet 2012;21:4187–200. [PubMed: 22730493]
42. Hamaguchi Y, Majima Y, Ukai K, Sakakura Y, Miyoshi Y. Significance of kallikrein-kinin system in otitis media with effusion. Ann Otol Rhinol Laryngol Suppl 1986;124:5–8. [PubMed: 3087261]
43. Kobashi M, Morizane S, Sugimoto S, Sugihara S, Iwatsuki K. Expression of serine protease inhibitors in epidermal keratinocytes is increased by calcium but not 1,25-dihydroxyvitamin D3 or retinoic acid. Br J Dermatol 2017;176:1525–32. [PubMed: 27797397]
44. Tiryakioglu NO, Onal Z, Saygili SK, Onal H, Tunali NE. Treatment of ichthyosis and hypernatremia in a patient with Netherton syndrome with a SPINK5 c.153delT mutation using kallikrein inhibiting ointment. Int J Dermatol 2017;56:106–8. [PubMed: 26889743]
45. Maatouk I, Moutran R, Tomb R. Narrowband ultraviolet B phototherapy associated with improvement in Netherton syndrome. Clin Exp Dermatol 2012;37:364–6. [PubMed: 22582910]
46. Di WL, Mellerio JE, Bernadis C, Harper J, Abdul-Wahab A, Ghani S, Chan L, Martinez-Queipo M, Hara H, McNicol AM, Farzaneh F, McGrath J, Thrasher A, Qasim W Phase I study protocol for ex vivo lentiviral gene therapy for the inherited skin disease, Netherton syndrome. Hum Gene Ther Clin Dev 2013;24:182–90. [PubMed: 24329107]
47. Liu SG, Luo YR, Huang LF. Dynamics of size-fractionated bacterial communities during the coastal dispersal of treated municipal effluents. App Microbiol Biotechnol 2016:100;5839–48.
48. Ayuthaya SI, Leelaporn A, Kiratisin P, Oberdofer P. *Pseudoclavibacter* otitis media in a 3-year-old boy with pulmonary and spinal tuberculosis. Medicine (Baltimore) 2015;94:e709. [PubMed: 25929901]
49. Oyaert M, De Baere T, Breyne J, De Laere E, Mariën S, Waets P, Laffut W. First case of *Pseudoclavibacter bifida* bacteremia in an immunocompromised host with chronic obstructive pulmonary disease (COPD). J Clin Microbiol 2013;51:1973–6. [PubMed: 23536403]
50. Schieffer KM, Sabey K, Wright JR, Toole DR, Drucker R, Tokarev V, Harris LR, Deiling S, Eshelman MA, Hegarty JP, Yochum GS, Koltun WA, Lamendella R, Stewart DB. The microbial ecosystem distinguished chronically diseases tissue from adjacent tissue in the sigmoid colon of chronic, recurrent diverticulitis patients. Sci Rep 2017;7:8467. [PubMed: 28814777]
51. Yamamoto K, Hayakawa K, Nagamatsu M, Fujiya Y, Mawatari M, Kutsuna S, Takeshita N, Tamura S, Mezaki K, Ohmagari N. Bacteremia due to *Arthrobacter creatinolyticus* in an elderly diabetic man with acute cholangitis. Jpn J Infect Dis 2017;70:201–2. [PubMed: 27357988]
52. Ghoul M, Mitri S. The ecology and evolution of microbial competition. Trends Microbiol 2016;24:833–45. [PubMed: 27546832]
53. Honda K, Littman DR. The microbiome in infectious disease and inflammation. Annu Rev Immunol 2012;30:759–95. [PubMed: 22224764]

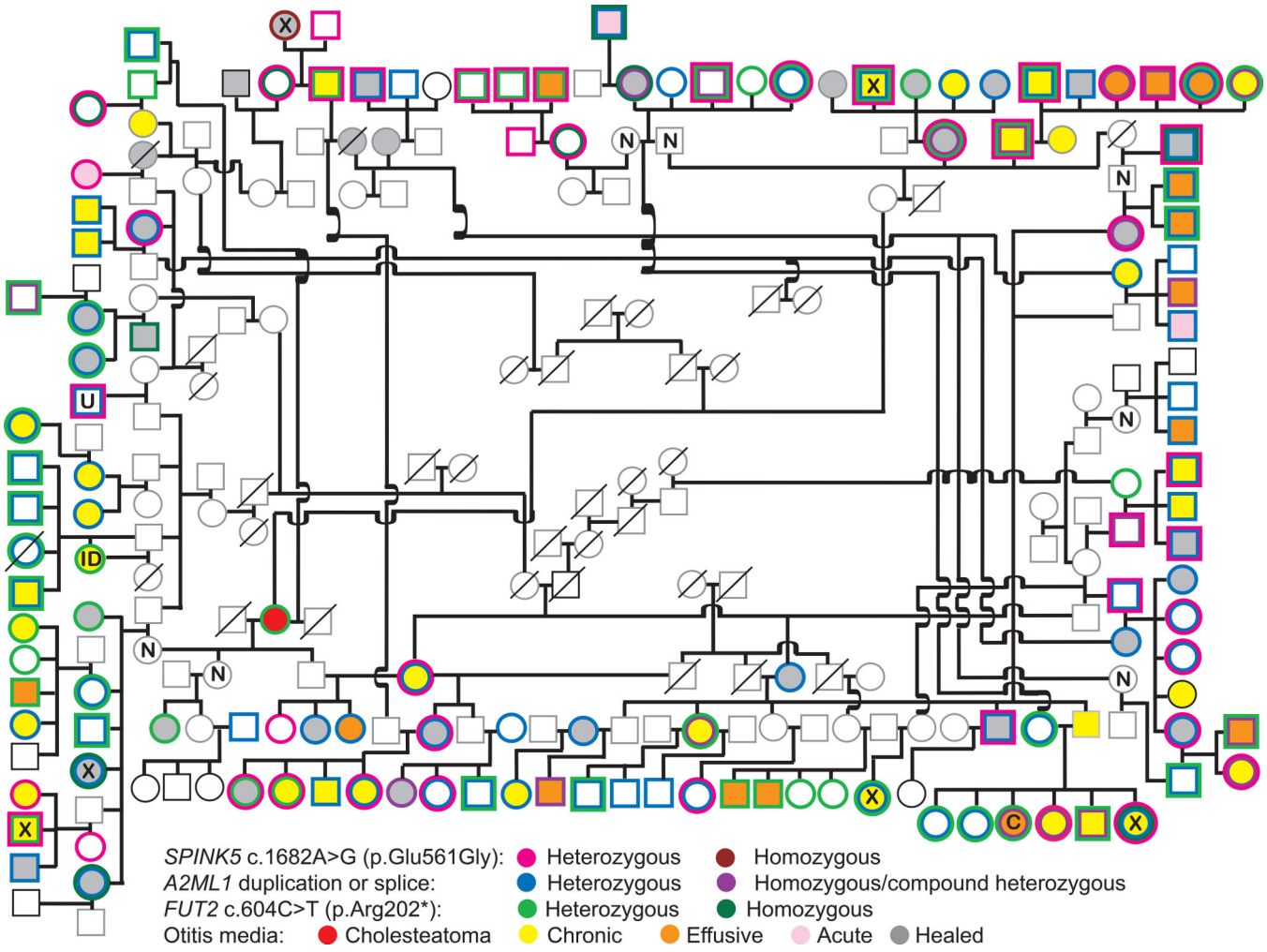

Figure 1. Indigenous Filipino pedigree with otitis media.

All individuals with chronic (*yellow fill*), acute (*pink fill*), effusive (*orange fill*) and healed otitis media (*gray fill*) were considered affected. One individual with chronic otitis media has cholesteatoma (*red fill*). X=DNA sample sent for exome sequencing; N=normal ears, no DNA sample; U=unknown phenotype; ID=intellectual disability; C=cleft lip and palate. Wildtype individuals with normal ears are indicated by *clear symbols with black lines,* unascertained individuals by *clear symbols with gray lines.* Variants carried are indicated by *line color* as in the legend below the pedigree. Individual IPOM-2 is indicated with an *X at the upper left corner* of the pedigree drawing. This figure represents our most complete record of pedigree relations to date and only about half of the ascertained individuals in this pedigree were noted in previous publications for *A2ML1* and *FUT2*.[8–10] Two-point affecteds-only analyses using 90% penetrance, 5% phenocopy and disease allele frequency of 0.01 revealed the following LOD scores (θ=0): *A2ML1* duplication, 6.30; *FUT2* nonsense variant, 2.83; *SPINK5* missense variant, 4.09. When the linkage analyses were performed conditional on the proband who is wildtype for *A2ML1* and *FUT2* variants and with disease MAF equal to variant MAF, the LOD scores for the *A2ML1* duplication and *FUT2* stop

variants dropped to 1.22 and 1.43 respectively, while the LOD score for the *SPINK5* variant remained significant at 3.91.

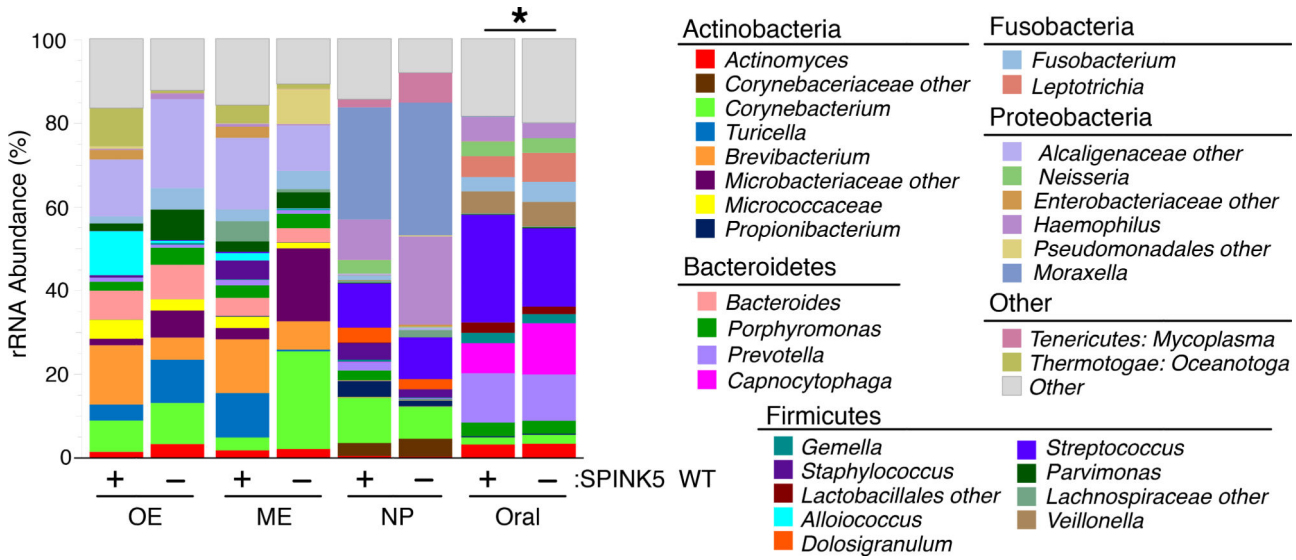

Figure 2. Relative abundance profiles of bacterial genera comparing *SPINK5*-wildtype (+) and *SPINK5*-variant (−) individuals.

16S rRNA gene profiling was performed to profile bacterial communities in the outer ear (OE), middle ear (ME), nasopharynx (NP), and oral cavity. Each profile averages the relative abundances of genera across all members of a comparison group. Results of permutation-based multiple analysis of variance (PERMANOVA) tests are indicated above the barcharts. Results for OE and ME were adjusted for *A2ML1* genotype, while those for NP and oral cavity were adjusted for age. *: $p<0.05$. *SPINK5* genotype was associated with altered microbiota composition within the oral cavity ($p=0.02$).

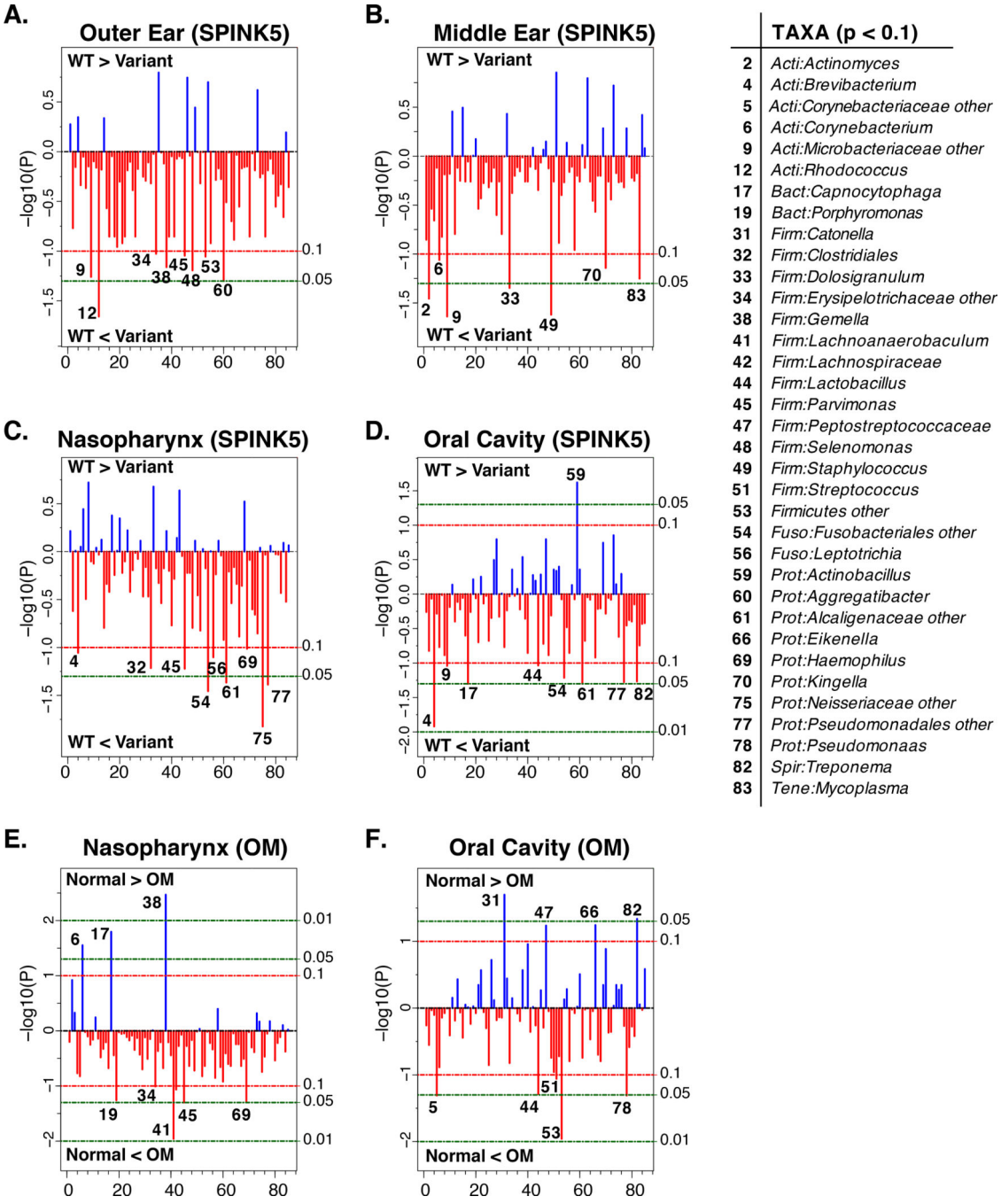

Figure 3. Association of *SPINK5* genotype and otitis media status with resident bacterial communities.

16S rRNA gene profiling was performed to profile bacterial communities in the outer ear (OE), middle ear (ME), nasopharynx (NP), and oral cavity. Manhattan plots display the -$\log_{10}$(*p*-values) for linear model regression tests comparing the relative abundances (RA) of each genus (center-log transformed RA) between (A-D) *SPINK5*-wildtype and *SPINK5*-variant individuals, and (E-F) otitis media status. *Blue lines* denote taxa with WT RA > variant RA (A-D) or Normal RA > OM RA (E-F). *Red lines* denote genera with WT RA < variant RA (A-D) or Normal RA < OM RA (E-F); -$\log_{10}$(*p*-values) for these genera are

plotted below the 0 line for graphing purposes to distinguish them from blue lines. Based on results of PERMANOVA tests, linear models for OE and ME included *A2ML1* genotype as a covariate, while those for NP and oral cavity were adjusted for age. Genera that are significant in one or more statistical tests ($p<0.1$) are listed. *Horizontal lines* denote p-value thresholds of 0.1, 0.05, and 0.01.

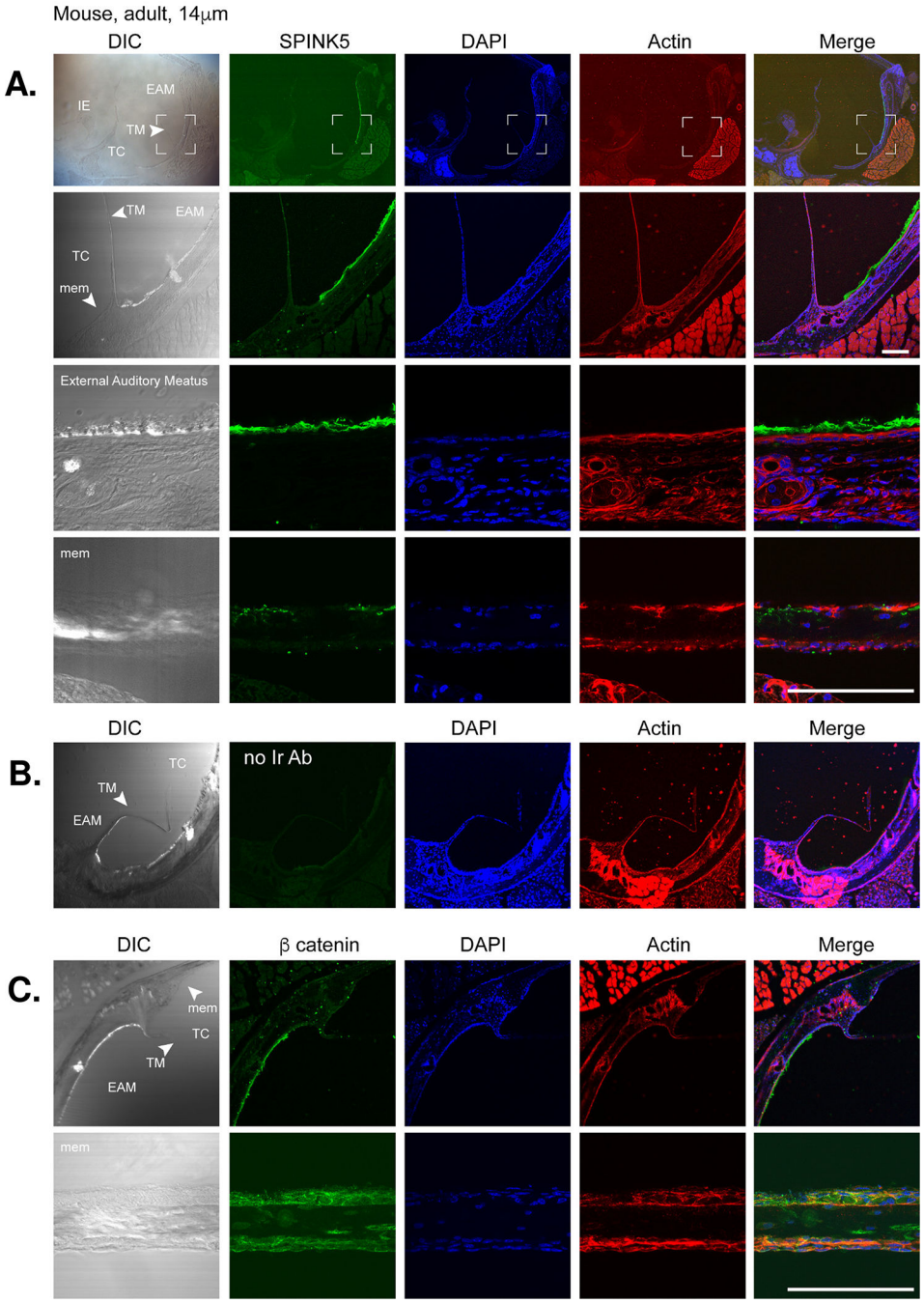

Figure 4. SPINK5 is faintly expressed in the mucosal epithelium of the middle ear and the tympanic membrane and strongly expressed in the external auditory canal skin of adult wildtype mice.

(A, B, C) Sagittal cryosections of an adult wildtype mouse ear showing the tympanic cavity (TC), the middle ear mucosa (MEM), the external auditory meatus (EAM), the tympanic membrane (TM) and the inner ear (IE). Confocal images of cryosections immunostained with anti-SPINK5 antibody (A, *green*), secondary antibody only (B, *green*) or anti-β-catenin antibody (C, *green*) and counterstained with rhodamine phalloidin (*Actin, red*) and DAPI (*Nuclei, blue*) are shown. (A) SPINK5 is mostly localized in the external auditory meatus

epithelium and faintly expressed in the middle ear mucosa and the tympanic membrane as compared to the negative control (B). (C) β-catenin immunostaining is used as a positive control. DIC: Differential interference contrast. Scale bars: 100 μm.

Table 1.

Rare damaging exome variants in indigenous Filipino individual IPOM-2 who is *A2ML1*- and *FUT2*-negative

Gene	Chr	Variant[a]	Damaging Results	Scaled CADD	IPOM-2 Genotype	Exome ID (+)[b]	CLHNS MAF	LOD score[c]	Saliva RNA-seq limma nominal-*p*[d]
GAK	4p16.3	NM_005255:c.3032T>C(p.(Phe1011Ser))	MA,MT,PP2,PR	25.0	Het	14	0	−0.30	ND
SPINK5	**5q32**	**NM_006846:c.1682A>G(p.(Glu561Gly))**	**MA,PR**	**25.3**	**Hom**	**14/41/76**	**0.004**	**4.09**	**0.16**
TMEM270/ WBSCR28	7q11.23	NM_182504:c.670C>T(p.(Gln224*))	MT	39.0	Het	14/41/45	0	4.95	very low exp.
MRPL15	8q11.23	NM_014175:c.191G>A(p.(Arg64His))	LRT,MA,mLR,mSVM,, MT,PP2,PR,SI	34.0	Het	14/76	0	2.45	very low exp.
GALNT18	11p15.3	NM_198516:c.1132T>C(p.(Ser378Pro))	LRT,MA,mSVM,MT,PR,SI	28.6	Het	14	0	0.50	ND
SIDT2	11p15.3	NM_001040455:c.1249G>T(p.(Asp417Tyr))	LRT,MT,PP2,PR,SI	25.3	Het	18/76	0.006	4.30	0.67
CHD8	14q11.2	NM_001170629:c.2105A>G(p.(Lys702Arg))	LRT,MT	23.9	Het	14/76	0	1.40	ND
JMJD7	15q15.1	NM_005090:c.553C>G(p.(Leu185Val))	LRT,MA,MT,PP2	26.7	Het	14/41	0	3.00	0.81
PDP2	16q22.1	NM_020786:c.1279G>T(p.(Val427Leu))	LRT,MA,MT,SI	24.4	Het	14/76	0	2.39	0.37
SAT2	**17p13.1**	**NM_001320845:c.385C>G(p.(Pro129Ala))**	**MA,MT,PP2,PR,SI**	**22.4**	**Hom**	**41**	**0**	**5.02**	**0.03**
CARD14	17q25.3	NM_024110:c.2517G>C(p.(Lys839Asn))	MA,SI	28.2	Het	14/76	0	1.21	ND
CDH19	18q22.1	NM_021153:c.794T>C(p.(Val265Ala))	LRT,MA,PP2,PR,SI	23.9	Het	14/18/41	0	5.72	0.54
FBXO15	18q22.3	NM_001142958:c.802T>G(p.(Leu268Val))	LRT,MA,MT,PP2,SI	22.6	Het	14/45	0	1.47	ND
ZNF790	19q13.12	NM_206894:c.1478T>G(p.(Leu493Arg))	MA,PP2,PR,SI	25.5	Het	18/76	0	4.38	0.64
SLC13A3	20q13.12	NM_022829:c.401C>T(p.(Thr134Ile))	LRT,MA,MT,PP2	24.9	Het	41	0	1.20	ND
ZNRF3	**22q12.1**	**NM_001206998:c.2522G>A(p.(Gly841Asp))**	**LRT,MT,PP2,SI**	**25.2**	**Het**	**14/18/ 41/76**	**0**	**4.77**	**0.03**

Abbreviations: LRT, likelihood ratio test; MA, MutationAssessor; mLR, meta-logistic regression; mSVM, meta-support vector machine; MT, MutationTaster; PR, PROVEAN; SI, SIFT; CADD, Combined Annotation Dependent Depletion scaled scores; het, heterozygous; hom, homozygous; CLHNS, Cebu Longitudinal Health and Nutrition Survey; MAF, minor allele frequency; ND, not determined due to LOD<2.

[a] Variants are listed according to chromosomal position. Variants in *bold* were further selected for microbiome analyses.

[b] IDs of indigenous Filipinos who, in addition to IPOM-2, have exome data (n=5) and carry the variant.

[c] LOD scores at θ=0. Specified parameters include autosomal dominant inheritance with 90% penetrance, 1% disease allele frequency and variant MAF from CLHNS.

[d] Correction for *A2ML1* genotype was done for the variants in *bold* by summing counts for variant allele copies (e.g. *A2ML1+SPINK5*). Corrected *p*-values: *SPINK5*=0.08, *SAT2*=0.08, *ZNRF3*=0.02. In general RNA counts are lower for carriers of the three variants.

Table 2.

SPINK5 variants identified in multiple cohorts with otitis media

Population	Variant (NM_006846.3)	CADD score	Damaging prediction	Ancestry	Control population	Control MAF	gnomAD population	gnomAD MAF
IPOM	c.1682A>G (p.(Glu561Gly))	25.4	MA,PP2,PR	Filipino	CLHNS	0.004	Overall	0
PGH (*n*=2)	c.1345C>A (p.(Leu449Ile))	21.1	PP2,SI	Filipino	CLHNS	0.008	OTH	0.0002
UMN8009[a]	c.802C>T (p.(Arg268Cys))	23.1	MA,PP2,SI	Eur-Am	NFE	0.008	ASJ	0.02
UTMB (*n*=2)	c.802C>T (p.(Arg268Cys))	23.1	MA,PP2,SI	Eur-Am	NFE	0.008	ASJ	0.02
UTMB (*n*=1)	c.*74A>G	9.98	MT (splice)	Eur-Am	NFE	0	Overall	0
UHF (*n*=1)	c.551G>A (p.(Gly184Asp))	27.5	MA,MT,PP2,Pr	Finnish	FIN	0.001	FIN	0.001
UHF (*n*=1)	c.652C>T (p.(Arg218*))	33.0	MT (NMD)	Finnish	FIN	0.0003	FIN	0.0003
UHF (*n*=2)[a]	c.802C>T (p.(Arg268Cys))	23.1	MA,PP2,SI	Finnish	FIN	0.003	ASJ	0.02
UHF (*n*=1)	c.2243A>G (p.(Glu748Gly))	16.3	PP2,SI	Finnish	FIN	0.003	NFE	0.004
UHF (*n*=1)	c.2493T>A (p.(Asn831Lys))	4.7	PP2	Finnish	FIN	0.0001	FIN	0.0001
UHF (*n*=1)	c.2974A>T (p.(Met992Leu))	23.1	PP2	Finnish	FIN	0.00005	FIN	0.00005

Abbreviations as in table 1. Additional abbreviations: IPOM, indigenous Filipino population; PGH, Philippine General Hospital; UMN, Minnesota family with otitis media, of European-American ancestry; UTMB, University of Texas Medical Branch; UHF, Finnish probands/families with otitis media. For PGH, UTMB and UHF, number of carriers per variant is indicated. CLHNS represents the general Filipino population. The rest are matched populations from gnomAD: NFE, non-Finnish European; FIN, Finnish; ASJ, Ashkenazi Jewish; OTH, other.

[a]The two families with multiple affected relatives carrying this variant are shown in supplementary figure S2. Due to lack of power, non-parametric linkage analysis for c.802C>T (p.(Arg268Cys)) and c.2478T>G (p.(Asp826Glu)) using MERLIN was non-significant for 41 Minnesotan or Finnish families who were mostly wildtype.