FUT2 Variants Confer Susceptibility to Familial Otitis Media

Regie Lyn P. Santos-Cortez,[1,2,3,*] Charlotte M. Chiong,[3,4,5] Daniel N. Frank,[6] Allen F. Ryan,[7] Arnaud P.J. Giese,[8] Tori Bootpetch Roberts,[1] Kathleen A. Daly,[9] Matthew J. Steritz,[1] Wasyl Szeremeta,[10] Melquiadesa Pedro,[3] Harold Pine,[10] Talitha Karisse L. Yarza,[3,4] Melissa A. Scholes,[1,11] Erasmo Gonzalo d.V. Llanes,[3,5] Saira Yousaf,[8] Norman Friedman,[1,11] Ma. Leah C. Tantoco,[3,5] Todd M. Wine,[1,11] Patrick John Labra,[5] Jeanne Benoit,[6] Amanda G. Ruiz,[1,11] Rhodieleen Anne R. de la Cruz,[5] Christopher Greenlee,[1,11] Ayesha Yousaf,[12] Jonathan Cardwell,[13] Rachelle Marie A. Nonato,[5] Dylan Ray,[1] Kimberly Mae C. Ong,[5] Edward So,[8] Charles E. Robertson,[6] Jordyn Dinwiddie,[1,11] Sheryl Mae Lagrana-Villagracia,[3] University of Washington Center for Mendelian Genomics (UWCMG), Samuel P. Gubbels,[1] Rehan S. Shaikh,[12] Stephen P. Cass,[1] Elisabet Einarsdottir,[14,15] Nanette R. Lee,[16] David A. Schwartz,[13] Teresa Luisa I. Gloria-Cruz,[3,5] Michael J. Bamshad,[17] Ivana V. Yang,[13] Juha Kere,[14,15,18] Generoso T. Abes,[3,5] Jeremy D. Prager,[1,11] Saima Riazuddin,[8] Abner L. Chan,[3,5] Patricia J. Yoon,[1,11] Deborah A. Nickerson,[17] Eva Maria Cutiongco-de la Paz,[19,20] Sven-Olrik Streubel,[1,11] Maria Rina T. Reyes-Quintos,[3,4,5,19] Herman A. Jenkins,[1] Petri Mattila,[21] Kenny H. Chan,[1,11] Karen L. Mohlke,[22] Suzanne M. Leal,[23] Lena Hafrén,[21] Tasnee Chonmaitree,[24] Michele M. Sale,[25,26,27] and Zubair M. Ahmed[8]

Non-secretor status due to homozygosity for the common *FUT2* variant c.461G>A (p.Trp154*) is associated with either risk for autoimmune diseases or protection against viral diarrhea and HIV. We determined the role of *FUT2* in otitis media susceptibility by obtaining DNA samples from 609 multi-ethnic families and simplex case subjects with otitis media. Exome and Sanger sequencing, linkage analysis, and Fisher exact and transmission disequilibrium tests (TDT) were performed. The common *FUT2* c.604C>T (p.Arg202*) variant co-segregates with otitis media in a Filipino pedigree (LOD = 4.0). Additionally, a rare variant, c.412C>T (p.Arg138Cys), is associated with recurrent/chronic otitis media in European-American children ($p = 1.2 \times 10^{-5}$) and US trios (TDT p = 0.01). The c.461G>A (p.Trp154*) variant was also over-transmitted in US trios (TDT p = 0.01) and was associated with shifts in middle ear microbiota composition (PERMANOVA $p < 10^{-7}$) and increased biodiversity. When all missense and nonsense variants identified in multi-ethnic US trios with CADD > 20 were combined, *FUT2* variants were over-transmitted in trios (TDT p = 0.001). *Fut2* is transiently upregulated in mouse middle ear after inoculation with non-typeable *Haemophilus influenzae*. Four FUT2 variants—namely p.Ala104Val, p.Arg138Cys, p.Trp154*, and p.Arg202*—reduced A antigen in mutant-transfected COS-7 cells, while the nonsense variants also reduced FUT2 protein levels. Common and rare *FUT2* variants confer susceptibility to otitis media, likely by modifying the middle ear microbiome through regulation of A antigen levels in epithelial cells. Our families demonstrate marked intra-familial genetic heterogeneity, suggesting that multiple combinations of common and rare variants plus environmental factors influence the individual otitis media phenotype as a complex trait.

Introduction

FUT2 (MIM: 182100), which encodes alpha-(1,2)-fucosyltransferase, is the human secretor gene that controls expression of the Lewis and ABO(H) antigens on mucosal epithelia. In the Genotype-Tissue Expression (GTEx) database, it is highly expressed in salivary gland, gastrointestinal tract, vagina, and bladder. Non-secretor status for

[1]Department of Otolaryngology, University of Colorado School of Medicine (CUSOM), Aurora, CO 80045, USA; [2]Center for Children's Surgery, Children's Hospital Colorado (CHCO), Aurora, CO 80045, USA; [3]Philippine National Ear Institute, University of the Philippines (UP) Manila – National Institutes of Health (NIH), Manila 1000, Philippines; [4]National Hearing Screening Reference Center, UP Manila-NIH, Manila 1000, Philippines; [5]Department of Otorhinolaryngology, UP Manila College of Medicine – Philippine General Hospital, Manila 1000, Philippines; [6]Division of Infectious Diseases, Department of Medicine, CUSOM, Aurora, CO 80045, USA; [7]Division of Otolaryngology, Department of Surgery, University of California San Diego School of Medicine, La Jolla, CA 92093, USA; [8]Department of Otorhinolaryngology, Head and Neck Surgery, School of Medicine, University of Maryland, Baltimore, MD 21201, USA; [9]Department of Otolaryngology, Head and Neck Surgery, University of Minnesota, Minneapolis, MN 55455, USA; [10]Department of Otolaryngology, University of Texas Medical Branch (UTMB), Galveston, TX 77555, USA; [11]Department of Pediatric Otolaryngology, CHCO, Aurora, CO 80045, USA; [12]Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Punjab, Pakistan; [13]Department of Medicine, CUSOM, Aurora, CO 80045, USA; [14]Folkhälsan Institute of Genetics and Molecular Neurology Research Program, University of Helsinki, Helsinki 00014, Finland; [15]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge 141 83, Sweden; [16]USC-Office of Population Studies Foundation and Department of Anthropology, Sociology and History, University of San Carlos, Cebu City 6000, Philippines; [17]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [18]Department of Medical and Molecular Genetics, King's College London, London SE1 9RT, UK; [19]UP Manila – NIH, Manila 1000, Philippines; [20]Philippine Genome Center, UP, Quezon City 1101, Philippines; [21]Department of Otorhinolaryngology, Head & Neck Surgery, University of Helsinki and Helsinki University Hospital, 00029 HUS, Finland; [22]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; [23]Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [24]Department of Pediatrics, Division of Infectious Diseases, UTMB, Galveston, TX 77555, USA; [25]Center for Public Health Genomics, School of Medicine, University of Virginia, Charlottesville, VA 22908, USA; [26]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA; [27]Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908, USA
*Correspondence: regie.santos-cortez@ucdenver.edu
https://doi.org/10.1016/j.ajhg.2018.09.010.

FUT2 has been associated with multiple phenotypes, either (1) conferring risk for recurrent urinary tract infection, infection with *Neisseria meningitidis* or *Streptococcus pneumoniae,* meningitis/sepsis/cellulitis due to *Haemophilus influenzae*, mucosal candidiasis, and autoimmune diseases, or (2) protecting against specific viruses and *Pseudomonas sp*. infection in cystic fibrosis (Table S1). These associations were detected for a stop variant *FUT2* c.461G>A (p.Trp154*; rs601338) that has a minor allele frequency (MAF) of 0.26–0.50 in multiple gnomAD populations except East Asians (MAF = 0.002). It was hypothesized that positive selection for rs601338 occurred in non-East Asian populations because the variant confers protection against viral diarrhea[1] (Table S1). In the GTEx Portal, *cis*-eQTL analysis further supports a significant effect of rs601338 on *FUT2* expression in salivary gland, colon, and lung; however, GTEx did not include the middle ear. A genome-wide association study (GWAS) involving >120,000 European-descent individuals identified a synonymous *FUT2* variant (rs681343) in strong linkage disequilibrium (LD) with rs601338 as the most significant locus for childhood ear infections,[2,3] but the locus has not been replicated and the estimated effect was small and protective (odds ratio/OR = 0.9; $p = 3.5 \times 10^{-30}$). Additionally, case status was based on adult anamnesis rather than medical records and/or otologic examination.[2,3]

The goal of this study is to determine the role of *FUT2* in otitis media by sequencing DNA samples from families and individuals with otitis media from various ethnicities, by testing *Fut2* expression or FUT2 localization in mouse and human middle ear and also expression in human saliva from individuals with otitis media, and by determining the effect of *FUT2* variants on cellular FUT2 and A antigen levels. Previously using exome data, we identified two rare *A2ML1* variants (duplication c.2478_2485dupGGCTAAAT and splice variant c.4061+1G>A) conferring otitis media susceptibility in an intermarried, indigenous Filipino population.[4] However, 22 of 83 (26.5%) affected indigenous individuals with DNA samples are wild-type for *A2ML1* variants. Here we report a population-specific *FUT2* variant c.604C>T (p.Arg202*) as a risk factor for otitis media susceptibility in indigenous Filipinos who are *A2ML1*-negative. Moreover, we identified additional missense variants in European-American and Pakistani families with recurrent acute or chronic otitis media. We also show that the c.461G>A (p.Trp154*) variant likewise confers otitis media risk and induces shifts in the middle ear microbiome. Our findings are further supported by increased *Fut2* expression in infected wild-type mouse middle ear and lower FUT2 and A antigen levels in cells transfected with FUT2 variants.

Subjects and Methods

Ethics Approval

Prior to study onset, approval was obtained from the following institutional review boards (IRB): Colorado Multiple IRB; Helsinki University Hospital; the National Commission on Indigenous Peoples, Philippines; University of Maryland Baltimore; University of Minnesota; University of the Philippines Manila (UPM); University of Texas Medical Branch (UTMB) Galveston; University of Virginia; and University of Washington. Informed consent was obtained from all adult participants and parents of children enrolled in the study. In addition, approval was obtained from the Institutional Animal Care and Use Committee of the Veterans Affairs Medical Center, San Diego, California.

Subject Ascertainment

An intermarried, indigenous Filipino population was reported to have a high prevalence of otitis media due to *A2ML1* variants.[4] Expanding on the previous study, a total of 137 individuals or 70% of the population provided clinical data from multiple visits and saliva samples for DNA isolation. The indigenous Filipinos diagnosed with otitis media are further described in Tables S2 and S3. Middle ear swabs for microbiome study were also obtained from 17 individuals with perforated eardrums due to chronic otitis media then stored and shipped using Oragene P117 prototype microbial kits (DNAgenotek). Microbial DNA isolation was performed using the Epicenter MasterPure Kit. In addition, salivary DNA samples were collected from 11 simplex case subjects from the Philippine General Hospital (PGH) who underwent surgery for chronic suppurative otitis media.

Otitis media-affected families from different cohorts were first identified upon referral of a child or proband for otitis media surgery (Table S2), except for the indigenous Filipino population. For this study, 257 trios from Texas were enrolled. Clinical data including age, sex, otitis media status and surgical technique, and saliva samples were obtained from each individual (Table S2). For Texas trios, both biological parents also provided saliva samples. Additionally, 76 individuals from Colorado who underwent surgery provided clinical data and saliva samples (Table S2). All saliva samples were collected using Oragene DNA collection kits from which DNA was isolated using the manufacturer's protocol. For microbiome study, 13 Colorado individuals with otitis media provided middle ear samples from which microbial DNA was isolated using Epicentre kits.

DNA samples were also obtained from two previously established family cohorts for otitis media, including 140 families from the University of Minnesota[5] and 105 families from Helsinki University Hospital[6] (Table S2). In Minnesota, all family members were examined by an otolaryngologist and tested by tympanometry for middle ear function, and family members were considered affected if ≥2 data sources, whether otoscopy, tympanometry, medical records, or personal history, were positive for otitis media. For Finnish families, clinical data, including history of otitis media, risk factors for otitis media, and otolaryngologic surgery, were obtained. Finnish individuals were considered positive for otitis media if they had insertion of tympanostomy tubes, effusive otitis media for >2 months, or recurrent otitis media (i.e., >3 episodes in 6 months or >4 episodes in 12 months).

Nineteen families with otitis media were enrolled from the southern Punjab province of Pakistan (Table S2). Age at onset and recurrence of otitis media episodes were determined from medical history and otitis media status by otoscopy. DNA was extracted from blood samples provided by family members who participated in the study. For all cohorts, individuals with known genetic, craniofacial, and immunodeficiency syndromes were excluded.

Because Filipinos are not represented in any genomic database, DNA samples were obtained from the Cebu Longitudinal Health and Nutrition Survey (CLHNS) cohort[7] in order to estimate variant MAF in the general Filipino population. The CLHNS DNA samples are from a prospective community-based birth cohort recruited in 1983–1985 from Cebu, Philippines in order to study infant feeding, socio-economic, and environmental factors that affect health and nutrition outcomes, but not otitis media.[7]

Human DNA Sequencing

Using DNA samples obtained from families and individuals with otitis media, exome and Sanger sequencing were performed to determine carriage of *FUT2* variants. DNA samples from six indigenous Filipinos were submitted for exome sequencing at the UWCMG, as previously described.[4] Using an Illumina HiSeq instrument, exome sequencing was performed to an average depth of ~60×. Sequence capture was performed using either the Roche NimbleGen SeqCap EZ Human Exome Library v.2.0 or the Big Exome 2011 Library. Burrows-Wheeler Aligner,[8] Genome Analysis Toolkit,[9] and SeattleSeq were used to generate BAM, vcf, and annotation files, respectively.

Sanger sequencing for *FUT2* (GenBank: NM_000511.5) was performed using DNA samples from the following: (1) the coding region in 257 Texan trios, 140 Minnesota probands and selected Minnesota families, 105 Finnish probands, 76 Colorado probands, and 19 Pakistani families; and (2) the c.604C>T (p.Arg202*; rs1800028) variant in 137 indigenous Filipinos, 167 unrelated CLHNS samples, and 11 individuals from the PGH who underwent surgery for chronic otitis media.

Bioinformatic, Linkage, and Statistical Analyses

In order to determine the predicted functional effect of identified *FUT2* variants, bioinformatic analyses were performed using web tools available for the following software: fathmm, MutationAssessor, MutationTaster, PolyPhen-2, and PROVEAN/SIFT. Two-point affecteds-only analysis using Superlink[10] was performed using: (1) for Minnesota families, previously available genotype data[5] for the *FUT2* c.772G>A (p.Gly258Ser; rs602662) variant, while noting both LOD and HLOD scores; and (2) for the indigenous Filipino pedigree, genotypes from Sanger sequencing of rs1800028. For both analyses, otitis media was assumed to be autosomal dominant with 90% penetrance and 1% phenocopy rate. All individuals with various otitis media types were considered affected. For the rs602662 variant in Minnesota families, nonparametric linkage analysis was also performed using MERLIN.[11]

For variants identified in US and Finnish cohorts for which gnomAD data are available, Fisher exact tests were performed when comparing variant MAF by cohort versus population-matched gnomAD MAF. In Texas and Colorado trios, the transmission disequilibrium test (TDT) as described by Spielman et al.[12] was performed for *FUT2* variants. The TDT method is applicable to the multi-ethnic Texas and Colorado cohorts because it is robust against population stratification, allowing the inclusion of all 333 trios in TDT analysis.

16S rRNA Sequencing and Microbiome Analysis

A total of 30 middle ear samples, 17 from indigenous Filipino and 13 from Colorado individuals with otitis media, were submitted for 16S rRNA sequencing and microbiome analysis. For 16S amplicon library construction, bacterial profiles were determined by broad-range amplification and sequence analysis of 16S rRNA genes following previously described methods.[13,14] In brief, amplicons were generated using primers that target approximately 300 base pairs of the V1-V2 variable region of the 16S rRNA gene. PCR products were normalized using a SequalPrep kit (Invitrogen/Thermo Fisher Scientific), pooled, lyophilized, purified, and concentrated using a DNA Clean and Concentrator Kit (Zymo). Pooled amplicons DNA was diluted to 15 pM and spiked with 25% of the Illumina PhiX control DNA prior to loading the sequencer. Illumina paired-end sequencing was performed on the Miseq platform with versions 2.4 of the Miseq Control Software and of MiSeq Reporter, using a 600 cycle version 3 reagent kit.

Illumina Miseq paired-end reads were aligned to human reference genome hg19 with Bowtie 2[15] and matching sequences discarded. The remaining non-human paired-end sequences were sorted by sample via barcodes in the paired reads with a Python script.[14] Paired reads were assembled using phrap.[16,17] Pairs that did not assemble were discarded. Assembled sequence ends were trimmed over a moving window of 5 nucleotides until average quality met or exceeded 20. Trimmed sequences with more than 1 ambiguity or shorter than 250 nt were discarded. Potential chimeras identified with UCHIME (usearch6.0.203_i86linux32)[18] using the Schloss[19] Silva reference sequences were removed from subsequent analyses. Assembled sequences were aligned and classified with SINA (1.3.0-r23838)[20] using the 418,497 bacterial sequences in Silva 115NR99[21] as reference configured to yield the Silva taxonomy. Operational taxonomic units (OTUs) were produced by clustering sequences with identical taxonomic assignments. This process generated 1,193,018 sequences for 30 samples (median of 28,066 sequences/sample; interquartile range: 6,785–35,245 sequences/sample). Goods coverage scores were>96% for all samples at the rarefaction point of 500 sequences, indicating adequate depth of sequence coverage for all samples. Overall differences in microbial community composition between groups were assessed by PERMANOVA tests using Bray-Curtis dissimilarity scores with 10^6–10^7 permutations. Differences in relative abundances of individual OTUs were assessed by non-parametric Kruskal-Wallis tests. Standard measures of alpha biodiversity, including richness (the number of OTUs per sample estimated by Chao1), community evenness (the uniformity of OTU distributions estimated by Shannon H/H_{max}), and complexity (Shannon diversity), were estimated at the rarefaction point of 500 sequences through 1,000 re-samplings; results were assessed by analysis of variance tests. The software packages R and Explicet v2.10.8[22] were used for data display, analysis, and figure generation.

Expression of Fut2 mRNA in Mouse Middle Ear during Acute Otitis Media

Information on *Fut2* expression in the middle ear was derived from microarray analysis of mRNA from wild-type mice that had acute otitis media after inoculation with *H. influenzae* compared to healthy control mice. Animals were 60- to 90-day-old wild-type C57/WB F1 hybrid mice (Jackson). Non-typeable *H. influenzae* (ntHI) strain 3655 (biotype II), originally isolated from the middle ear of an individual with otitis media from St. Louis and provided by Dr. Asa Melhus, Lund University, Sweden, was used at a concentration of 10^5–10^6 bacteria/mL. Mice were anesthetized with rodent cocktail (2.0 mg/kg xylazine and 40.0 mg/kg ketamine i.m.). Both middle ear bullae were exposed in an aseptic surgery by cervical midline incision. A small hole was made in each bulla so as not to perforate the tympanic membrane. A 30 g needle was

then used to carefully inject 5 μL of inoculum into the middle ear. The incision was closed and the tympanic membranes visually confirmed to be intact. Animals were sacrificed at various intervals after inoculation for tissue harvest. Uninoculated animals served as controls.

Two groups of 20 C57BL/6:CB F1 hybrid mice per time point were inoculated in the middle ear with ntHI strain 3655. Middle ear tissues were harvested at 0 hr (control), 3 hr, 6 hr, 1 day, 2 days, 3 days, 5 days, and 7 days after inoculation to obtain two independent biological replicates of the otitis media time course. The tissue samples were homogenized in TRIzol (Invitrogen) and total RNA extracted. After confirmation of RNA quality, mRNA was reverse transcribed using a T7-oligodT primer then *in vitro* transcribed using T7 RNA polymerase to generate biotinylated cRNA probes that were hybridized to Affymetrix MU430 2.0 whole-genome microarrays. Raw hybridization intensity data were median-normalized and statistical differences in gene transcript expression levels were evaluated using variance-modeled posterior inference (VAMPIRE).[23] Bonferroni multiple testing correction was applied. The expression of gene transcripts encoding FUT2 was extracted from this dataset and normalized to that in control (untreated) middle ears.

Immunolocalization Studies

In order to test changes in cellular localization of FUT2 due to variants, human FUT2 was amplified using human genomic DNA and the PCR product was cloned into pEGFP-C2 vector. GFP-FUT2^{Ala104Val}, GFP-FUT2^{Arg138Cys}, GFP-FUT2^{Trp154*}, and GFP-FUT2^{Arg202*} constructs were generated using site-directed mutagenesis protocol (Agilent). The following primary antibodies were used: anti-FUT2 (Figure S1; 1:50e, NBP1-80775, Novus Biologicals) and anti-Mannosidase II (1:200e, AB-12277, Abcam).

COS-7 and HEK293T cells were maintained in DMEM supplemented with 10% FBS, glutamine, and penicillin-streptomycin (Invitrogen) at 37°C, 5% CO_2. Twenty-four hours before transfection, cells were plated on a coverslip. Cells were then transfected with 2 μg of cDNA according to the Lipofectamine 2000 protocol (Fisher Scientific). After 24 hr of transfection, cells were fixed with 4% PFA for 20 min, washed three times with PBS, permeabilized, and blocked with 10% normal goat serum and 0.10% Triton X-100 in PBS for 1 hr at room temperature. Cells were incubated with primary antibodies overnight at 4°C, washed three times with PBS, and incubated with secondary antibodies for 1 hr at room temperature. Cells were imaged using a Zeiss 710 confocal microscope. Stacks of confocal images were acquired with a Z step of 0.5 μm and processed using ImageJ software (NIH).

FUT2 localization was also determined in cholesteatoma tissue from human middle ear. Paraffin sections of human cholesteatoma were obtained from the Pathology Biorepository Shared Services, University of Maryland School of Medicine. Sections were deparaffinized using xylene and rehydrated gradually through graded alcohols. Sections were heat-treated at 95°C for 5 min in a 10 mM sodium citrate buffer. Sections were then processed for immunostaining.

Flow Cytometry Studies

To examine the effect of FUT2 variants on cellular FUT2 and A antigen levels, COS-7 cells were plated in a 10 cm dish and transiently transfected with 10 μg of GFP-FUT2 variant constructs. After 72 hr, cells were trypsinized and counted and 250K cells were collected for each condition. Cells were immunostained overnight with IgM monoclonal blood Group A antigen antibody (Santa Cruz; sc-69951) or IgM control antibody diluted in FACS buffer (3% FBS PBS 1X). Cells were washed 3 times with FACS buffer and incubated for 45 min with an APC conjugated anti-IgM secondary antibody. Cells were washed 3 times and processed for flow cytometry analysis (BD LSR II).

Western Blot

In order to further determine FUT2 levels due to identified variants, COS-7 cells were plated in 100 mm culture dish for 24 hr at 37°C in 5% CO_2. On the day of transfection, 10 μg of each GFP-FUT2 variants were transfected into cells using Lipofectamine 2000. After 48 hr, cells were washed with cold 1× PBS, then homogenized with a sonicator (Fisher Scientific) at intensity setting 2 for 10 s in RIPA buffer containing a protease inhibitor mixture (Roche/Sigma-Aldrich). Samples were diluted in 2× SDS sample buffer. Samples were processed for western blot using 4%–20% Tris Glycine gel (Novex, Thermo Fisher Scientific). Membranes were probed with anti-GFP (Invitrogen, A-11122) and anti-GAPDH (Santa Cruz, sc-32233) antibodies.

Human RNA-Sequencing and Network Analysis

RNA-sequencing and network analysis was also performed in order to determine how *FUT2* is related to other genes previously identified for otitis media. Using Oragene RNA kits, saliva was collected from 23 individuals seen at the Children's Hospital Colorado and the University of Colorado Hospital who underwent surgery for otitis media. Saliva was isolated according to manufacturer's protocol, achieving a median RIN of 7. RNA samples were submitted to the University of Colorado Denver Genomics and Microarray Core for RNA-sequencing using the NuGen Trio RNA kit. An Illumina HiSeq instrument was used for sequencing. The RNA-sequence data were aligned to the human reference sequence (GRCh38) using STAR.[24] Summary counts were then generated using featureCounts.[25] Differential expression analysis using edgeR[26] was implemented in NetworkAnalyst.[27] The InnateDB database[28] was used to identify subnetworks of otitis-media-related genes according to carriage of the *FUT2* c.461G>A (p.Trp154*) variant.

Results

From the exome data of six indigenous individuals with chronic otitis media, three were heterozygous and two homozygous for *FUT2* c.604C>T (p.Arg202*; rs1800028), which ClinVar classifies as pathogenic for plasma vitamin B12 levels. Based on these exomes and Sanger sequencing of an additional 131 samples, 54 indigenous individuals were heterozygous and 6 homozygous for rs1800028 (MAF = 0.24, Table S3). Of these 60 carriers of rs1800028, a majority (58.3%) have current or previous diagnosis of otitis media, including 15 out of 60 (25%) who carry rs1800028 but not an *A2ML1* variant (Table S3). Sanger sequencing of 167 CLHNS DNA samples demonstrated that in the general Filipino population the MAF of rs1800028 is 0.13, consistent with previous reports in Southeast Asians.[1,29] In gnomAD the rs1800028 variant is most frequent in East Asians (MAF = 0.006) but is rarer in other populations (MAF ≤ 0.0005) and was not detected in our non-Filipino cohorts.

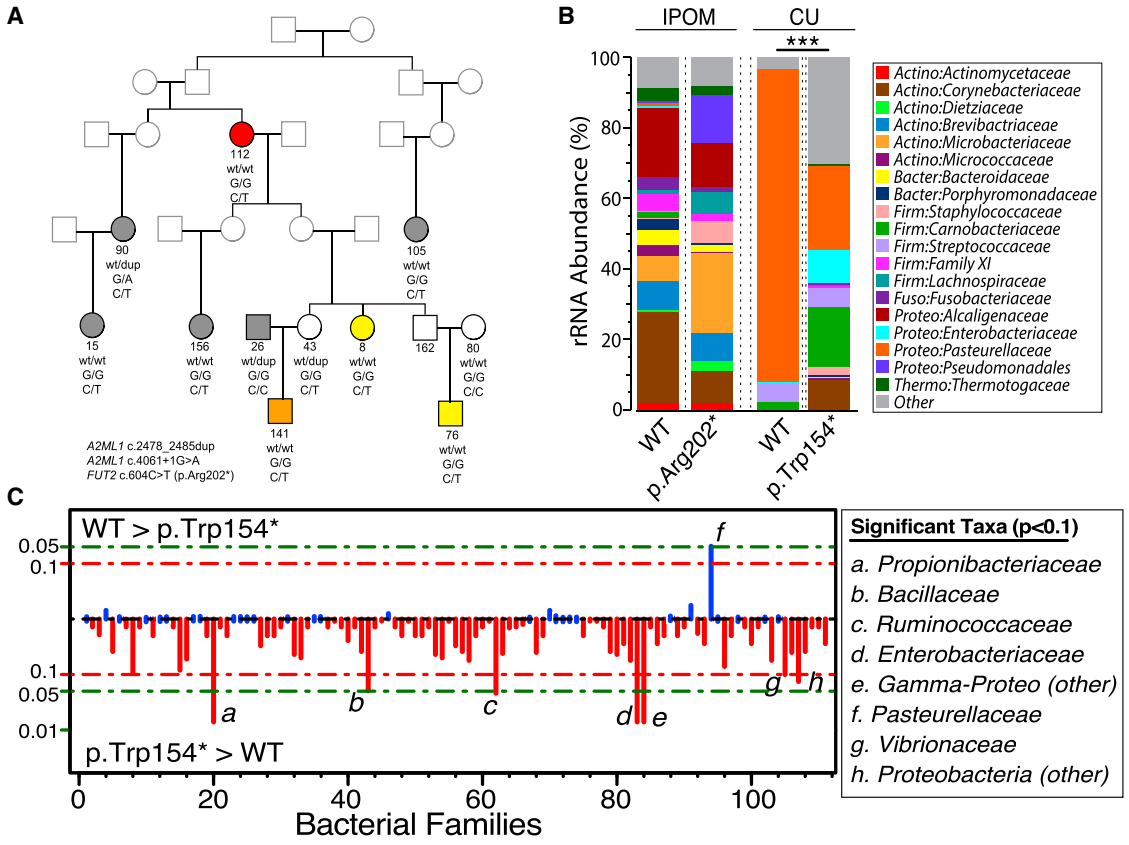

Figure 1. Indigenous Filipino (IPOM) Subpedigree and Microbiome Results for Middle Ear Samples from the IPOM and Colorado Cohorts

(A) Indigenous Filipino subpedigree with the *FUT2* c.604C>T (p.Arg202*; rs1800028) variant co-segregating with otitis media status, as follows: IPOM112 (87 y.o.), cholesteatoma (red); IPOM8 (24 y.o.) and IPOM76 (9 y.o.), chronic otitis media (yellow); IPOM141 (5 y.o.), effusive otitis media (orange); IPOM15 (12 y.o.), IPOM26 (50 y.o.), IPOM90 (47 y.o.), IPOM105 (36 y.o.), and IPOM156 (17 y.o.), healed otitis media with healed perforations, sclerosis, and/or previously documented otitis media (gray); IPOM43 (25 y.o.), IPOM80 (32 y.o.), and IPOM162 (40 y.o.), normal (clear with black border); unknown phenotype, clear with gray border. dup, *A2ML1* duplication; wt, wild-type.

(B) Average family-level relative abundances for individuals with otitis media from the indigenous Filipino community (IPOM) and Colorado (CU) defined by genotype for *FUT2* stop variants. Nominally significant differences were identified in the Colorado cohort according to genotype by permutation-based multiple analysis of variance (PERMANOVA) test ($p < 10^{-7}$) using Bray-Curtis dissimilarities.

(C) Manhattan plot showing $\log_{10}$-transformed p values per bacterial family, ascertained by Kruskal-Wallis tests of wild-type versus rs601338 (p.Trp154*) carriers within the Colorado cohort. Families with $p < 0.1$ are labeled on the plot and named to the right of the plot.

Linkage analysis for rs1800028 using the entire pedigree connecting majority of ascertained individuals resulted in LOD score deflation, even when *A2ML1*-variant-carriers were specified as having unknown status. To resolve the intra-familial genetic heterogeneity,[30] a subpedigree connecting indigenous individuals who are wild-type for *A2ML1* variants but carried the *FUT2* variant was reconstructed (Figure 1A). Two-point linkage analysis resulted in a statistically significant LOD score of 4.0 ($\theta = 0$). However, among 17 indigenous individuals with middle ear swabs, we did not detect differences in the middle ear microbiome due to carriage of the *FUT2* c.604C>T (p.Arg202*) variant (Figures 1B and 2).

When we Sanger-sequenced the *FUT2* coding region in 257 trios from Texas, a rare missense variant *FUT2* c.412C>T (p.Arg138Cys; rs1800022) that is not in LD with c.461G>A (p.Trp154*) is heterozygous in 13 out of 169 non-Hispanic European-American children who had ventilation tube surgery (Table 1). This variant is predicted to be damaging by multiple bioinformatics tools (fathmm, MutationAssessor, MutationTaster, PolyPhen-2, PROVEAN, SIFT) and has overall gnomAD MAF = 0.006, being most frequent in non-Finnish Europeans (0.0097) but rarest in East Asians (0.00005). Additionally, from 140 Minnesota families, of which 133 are European-American, 7 European-American families were positive for rs1800022, which was seen only in specific branches or individuals (Figure S2), suggesting other susceptibility variants per family.[30] In 76 Colorado children, of which 56 are European-American, 1 European-American child is heterozygous for rs1800022 (Table 2). The rs1800022 variant was also heterozygous in 2 out of 105 Finnish probands, but compared with gnomAD Finnish MAF this was non-significant. On the other hand, when comparing the

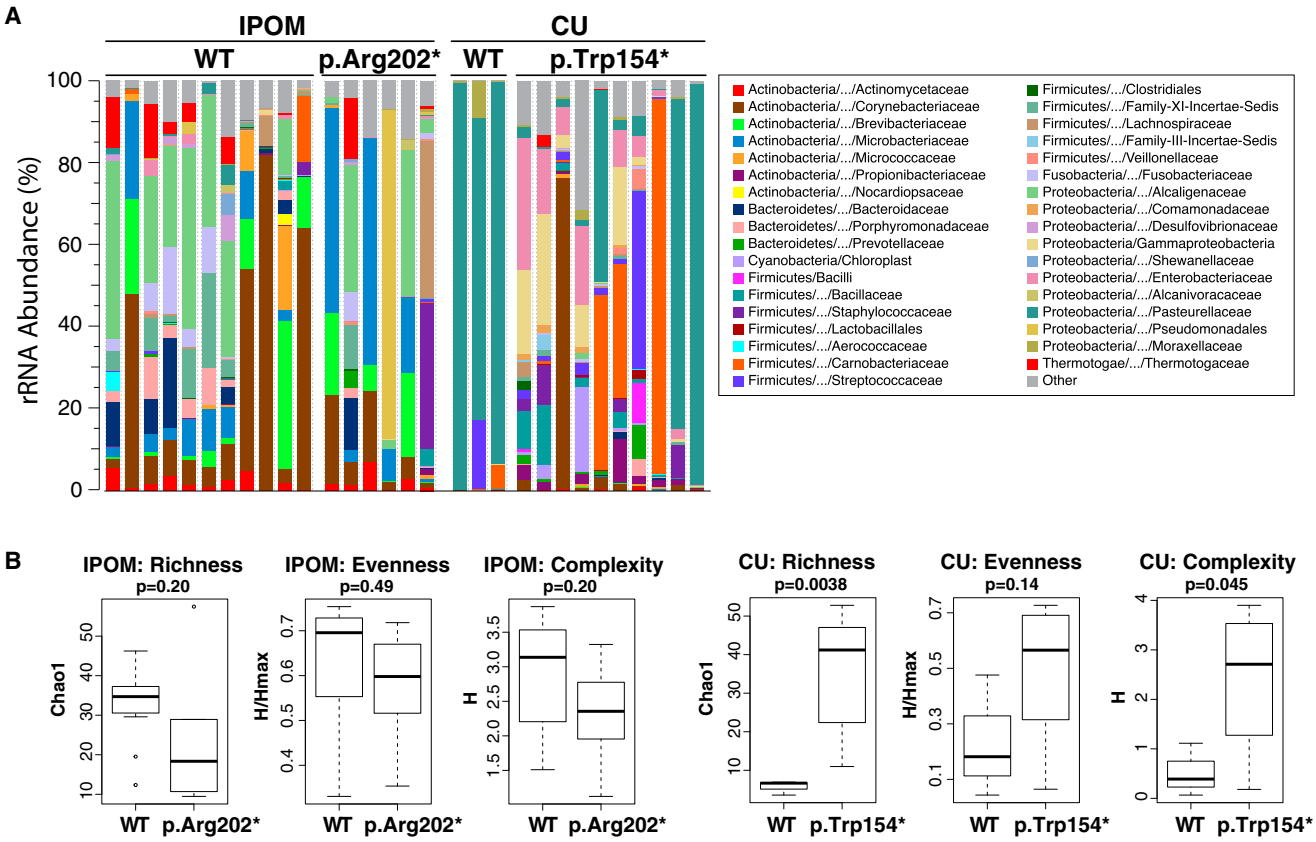

Figure 2. Additional Microbiome Results in the IPOM and CU Cohorts
(A) Inter-individual variability in middle ear samples by cohort and *FUT2* variant.
(B) Boxplots comparing biodiversity indices by cohort and *FUT2* variant: richness measures the number of OTUs per sample (estimated by Chao1), evenness measures the uniformity of OTU distributions (estimated by Shannon H/H_{max}), while complexity (Shannon diversity) combines both richness and evenness. Error bars indicate range of values; boxes show the first and third quartiles.

gnomAD MAF of rs1800022 in non-Finnish Europeans versus 21 heterozygous out of 358 affected European-American children from Colorado, Texas, or Minnesota (MAF = 0.03), the variant was significantly associated with otitis media (Fisher exact OR = 3.1, 95%CI:1.9,4.8; $p = 1.2 \times 10^{-5}$). When we sequenced all DNA samples from both parents of 333 children from Texas and Colorado (Tables 1 and 2), 1 parent of each of the 14 children with the rs1800022 variant also carry the variant, with 2 homozygous and 12 heterozygous parents. On the other hand, in three trios the father carried the rs1800022 variant but the child was wild-type. Thus in US trios the TDT for the rs1800022 variant was significant (p = 0.02). These findings support *FUT2* c.412C>T (p.Arg138Cys) as a novel rare variant conferring high risk for otitis media.

In the Texas trios, in agreement with the previously published GWAS,[2] several variants were in strong LD with stop variant *FUT2* c.461G>A (p.Trp154*; rs601338; Table 1), including the GWAS locus, synonymous variant rs681343, and missense variant c.772G>A (p.Gly258Ser; rs602662). However, compared to gnomAD frequencies using Fisher exact test, these variants were not significantly associated with otitis media in Finnish probands. Additionally, in 140 multi-ethnic Minnesota families with otitis media, both parametric and nonparametric linkage analyses using rs602662 genotypes had negative results. Sanger sequencing of rs601338 in Minnesota and Pakistani families further showed that the variant does not co-segregate with otitis media (Figures S2 and S3). However, the rs601338 variant was over-transmitted in Texas and Colorado trios (180 transmitted, 127 non-transmitted; TDT p = 0.002; Tables 1 and 2), implying that it is not protective against otitis media contrary to previous GWAS findings.[2] Furthermore, microbiome analysis of middle ear samples from 13 individuals with otitis media from Colorado demonstrated that carriage of the rs601338 variant was associated with significant overall shifts in middle ear microbiota composition (PERMANOVA $p < 10^{-7}$) and increased biodiversity (Figure 2). The rs601338 variant was also associated with increased relative abundances of *Enterobacteriaceae*, *Gammaproteobacteria*, and *Propionibacteriaceae* and decreased relative abundance of *Pasteurellaceae* (Figures 1B, 1C, and 2A); however, these were only nominally significant due to small sample size (FDR-adjusted p = 0.43). Taken together, these findings suggest that the *FUT2* c.461G>A (p.Trp154*) variant has a potentially mild pathogenic effect on the middle ear microbiome and is probably overtaken by other stronger genetic effects in familial otitis media.

Table 1. *FUT2* Variants Identified in 257 Texan Children with Otitis Media

Ethnicity		CADD Score	White	Hispanic	Black	Mixed/Other	Transmitted	Non-transmitted	TDT p
N (total)	257	–	169	46	24	18	–	–	–
Variants	**AA Change**								
rs1800021	p.Ile25Val	0.001	1	1	2	1	4	7	0.37
rs681343[a]	synonymous	3.2	118 (46 hom)	27 (3 hom)	20 (10 hom)	11 (3 hom)	–	–	–
rs149356814	p.Ala104Val	23.4	1	0	0	0	1	0	0.32
rs200157007	p.Pro112Leu	24.5	0	0	0	2	2	0	0.16
rs1800022[b]	p.Arg138Cys	28.5	13	0	0	0	11	2	0.01
rs1047781	p.Ile140Phe	23.5	0	0	0	2	1	0	0.32
rs601338[a]	p.Trp154*	37.0	124 (47 hom)	26 (3 hom)	21 (10 hom)	11 (3 hom)	139	100	0.01
rs602662[a]	p.Gly258Ser	23.6	122 (48 hom)	29 (5 hom)	22 (12 hom)	12 (4 hom)	–	–	–
rs485186[a]	synonymous	0.003	124 (48 hom)	29 (5 hom)	21 (13 hom)	12 (4 hom)	–	–	–
All Variants							149[c]	102[d]	0.003
Variants with CADD > 20 only							150	105	0.005

CADD, Combined Annotation Dependent Depletion.
[a]rs681343, rs601338, rs602662, and rs485186 are in strong LD.
[b]rs1800022 is a risk variant for otitis media. Two carriers of this variant have chronic/effusive otitis media while the rest have recurrent acute otitis media. All nine carriers required ventilation tube surgery by age 3 years (median 2 years). One child with chronic otitis media had multiple risk factors including family history of otitis media, no breastfeeding, daycare attendance, additional small children in the household, and exposure to smoking. For two trios, a parent is homozygous for the variant. In two additional trios, the father is heterozygous for rs1800022 but the child is wild-type.
[c]Total transmitted variants were counted based on the sum of transmitted variants (total 158) minus the number of parent-child pairs with the same two heterozygous variants (i.e., 5 pairs with both p.Arg138Cys and p.Trp154* and 1 pair with both p.Ile25Val and p.Trp154*) and minus 3 parent-child pairs where p.Ile25Val is transmitted but p.Trp154*is non-transmitted.
[d]Total non-transmitted variants were counted based on the sum of non-transmitted variants (total 109) minus the number of parents with two non-transmitted heterozygous variants (i.e., 1 parent with both p.Arg138Cys and p.Trp154* non-transmitted and 3 parents with both p.Ile25Val and p.Trp154* non-transmitted) and minus 3 parent-child pairs where p.Ile25Val is transmitted but p.Trp154*is non-transmitted.

Three additional *FUT2* variants were identified in Texan children (Table 1): (1) c.311C>T (p.Ala104Val; rs149356814) was predicted to be damaging (Mutation Assessor, fathmm) and is heterozygous in one European-American child; (2) c.335C>T (p.Pro112Leu; rs200157007) was predicted to be damaging (PolyPhen2, MutationAssessor, fathmm) and is heterozygous in two Asian-American children; and (3) c.418A>T (p.Ile140Phe; rs1047781) was predicted to be damaging (PolyPhen2, MutationAssessor, PROVEAN/SIFT, fathmm) but is annotated in ClinVar as benign. The rs1047781 variant was identified in two Asian-American children from Texas (gnomAD East Asian MAF = 0.47) but was not identified in Minnesota, Finnish, or Pakistani families nor in Colorado or Filipino simplex case subjects (all other gnomAD MAF $\leq$ 0.01). On the other hand, the two variants rs149356814 and rs200157007 are rare in multiple gnomAD populations but common in South Asians with MAF of 0.02 and 0.13, respectively. When 19 Pakistani families with otitis media were sequenced for these two variants, rs200157007 was detected in four families but does not co-segregate with otitis media. On the other hand, in one consanguineous Pakistani family (PKOM11) the low-frequency variant rs149356814 potentially co-segregates with otitis media in one branch as an autosomal-dominant trait, but it does not in family PKOM01 (Figure S3). Putting together the different *FUT2* variants with CADD score > 20 that were identified in Texan and Coloradan trios, *FUT2* is significantly associated with otitis media risk (TDT p = 0.001; Table 2).

As expected for fucosyltransferases, FUT2 protein localizes to the Golgi apparatus (Figure S4). In COS-7 cells transfected with GFP constructs for FUT2 variants p.Ala104Val, p.Arg138Cys, p.Trp154*, and p.Arg202*, FUT2 colocalization to the Golgi apparatus is not affected (Figure S4). However, in mouse middle ear during acute ntHI infection, *Fut2* increases in expression 10×, peaking at 24 hr, then returns to normal by 5 days (Figure 3).

In order to determine cellular levels of FUT2 and A antigen according to variant occurrence, flow cytometry was performed using variant-transfected cells. Control studies using IgM control antibody showed no significant binding with or without variants (Figure S5). Flow cytometry studies on GFP-FUT2 variant-transfected COS-7 cells showed a significant decrease in FUT2 levels due to nonsense variants p.Trp154* (n = 8, p < 0.001) and p.Trp202* (n = 8, p $\leq$ 0.001; Figures 4A and 4B). The expression of p.Ala104Val and p.Arg138Cys variants was not affected. These data were confirmed by western blot analyses (Figure 4C).

Cell surface A blood group epitope levels on GFP-positive cells were analyzed for each variant (Figure 4D). A antigen levels were observed in 42.5% of GFP-FUT2^{WT} cells compared to GFP-transfected cells (21.2%). The

Table 2. *FUT2* Variants Identified in 76 Coloradan Children with Otitis Media

Ethnicity		White	Hispanic	Black	Mixed/Other	Transmitted	Non-transmitted	TDT p
N (total)	76	56	6	0	14	–	–	–
Variants	**AA Change**							
rs1800022	p.Arg138Cys	1	0	0	0	1	1	1
rs601338	p.Trp154*	42 (15 hom)	5 (1 hom)	0	12 (4 hom)	41	27	0.09
Both variants						42	28	0.09
Combined with Texan trios (CADD > 20)						192	133	0.001

accumulation of A antigen at the surface of cells transfected with GFP-FUT2^{Ala104Val}, GFP-FUT2^{Arg138Cys}, GFP-FUT2^{Trp154*}, or GFP-FUT2^{Arg202*} variant constructs was significantly reduced (28.3%, 24%, 11.1%, and 13.2%, respectively). Levels of B and H antigens could not be assessed due to the lack of specific antibody for these antigens. To summarize, *Fut2* expression in wild-type mouse middle ear with acute otitis media is transient, while at the cellular level *FUT2* variants affect A antigen levels if a missense variant and both FUT2 and A antigen levels if a nonsense variant.

We also examined a potential role of *FUT2* in the development of acquired cholesteatoma in the middle ear due to chronic otitis media. Cholesteatoma is a collection of squamous debris encapsulated by keratinized epithelium that develops in the middle ear, which is rarely congenital but more commonly acquired after chronic otitis media and can cause serious complications (e.g., intracranial spread, hearing loss, facial nerve paralysis, vestibular loss) due to erosion of surrounding temporal bone.[31] FUT2 localization was tested in human cholesteatoma matrix, but either there was none or FUT2 was below the threshold level for detection in cholesteatoma tissue (Figure S6). Interestingly, in our Filipino and Colorado cohorts, the carrier frequencies of rs1800028 and rs601338 are increased in individuals with cholesteatoma (Table S4). Differential expression analysis of RNA-sequence data from saliva of 23 individuals with otitis media from Colorado according to carriage of the rs601338 variant revealed only 3 genes with FDR-adjusted p = 0.07 (corresponding log-fold changes as indicated): *TRNV* (10.9), *TRNF* (8.1), and *OSTM1* (−3.9). Of these three genes, *OSTM1* is connected to a subnetwork of OM-related genes (Figure S7).

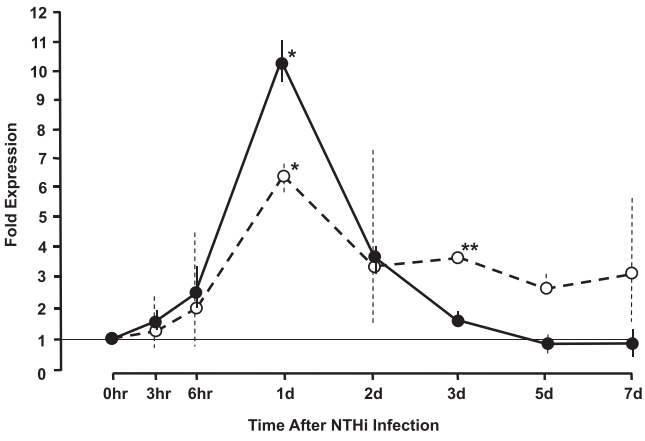

Figure 3. Fold Expression of *Fut2* across Different Time Points after Inoculation of Non-typeable *Haemophilus influenzae* into the Mouse Middle Ear

Affymetrix probes 1434862_at (filled circles) and 1450246_at (open circles), targeting Fut2 transcripts, revealed enhanced expression that peaked at 24 hr post-infection. The 0 hr data point represents uninfected middle ears. Vertical bars represent the range of fold values observed for each probe on all arrays.

Discussion

In this report we identified population-specific *FUT2* variants conferring risk for otitis media susceptibility in multiple cohorts of different ethnicity: (1) the c.604C>T (p.Arg202*; rs1800028) variant that is common in Southeast Asians and co-segregates with otitis media in *A2ML1*-negative Filipinos (Figure 1A); (2) the c.412C>T (p.Arg138Cys; rs1800022) variant which is rare in populations of European ancestry but occurs at MAF = 0.03 in European-American children with otitis media, thus increasing risk for otitis media by >3× (Tables 1 and 2); (3) three additional missense variants in Texan or Pakistani families (Table 1, Figure S3); and (4) the very common c.461G>A (p.Trp154*; rs601338) variant which is significantly associated with familial otitis media and induces shifts in the middle ear microbiome (Tables 1 and 2, Figures 1B, 1C, and 2). All five variants combined are over-transmitted in European-American trios with otitis media (Tables 1 and 2). Moreover, both missense and nonsense variants reduced A antigen expression in epithelial cells (Figure 4). However, we did not find middle ear microbiome differences due to the p.Arg202* variant, which is possibly due to the inclusion of indigenous individuals who also carry *A2ML1* variants that influence the middle ear microbiome[32] (Figures 1B and 2). On the other hand, the c.461G>A (p.Trp154*) variant influences richness and complexity of the middle ear microbiome (Figure 2), which is consistent with findings in the gastrointestinal tract though predominant taxa in the gut may be different.[33]

The transient expression of *Fut2* in the middle ear mucosa during acute otitis media (Figure 3) implies that the mechanism by which *FUT2* variants affect middle ear

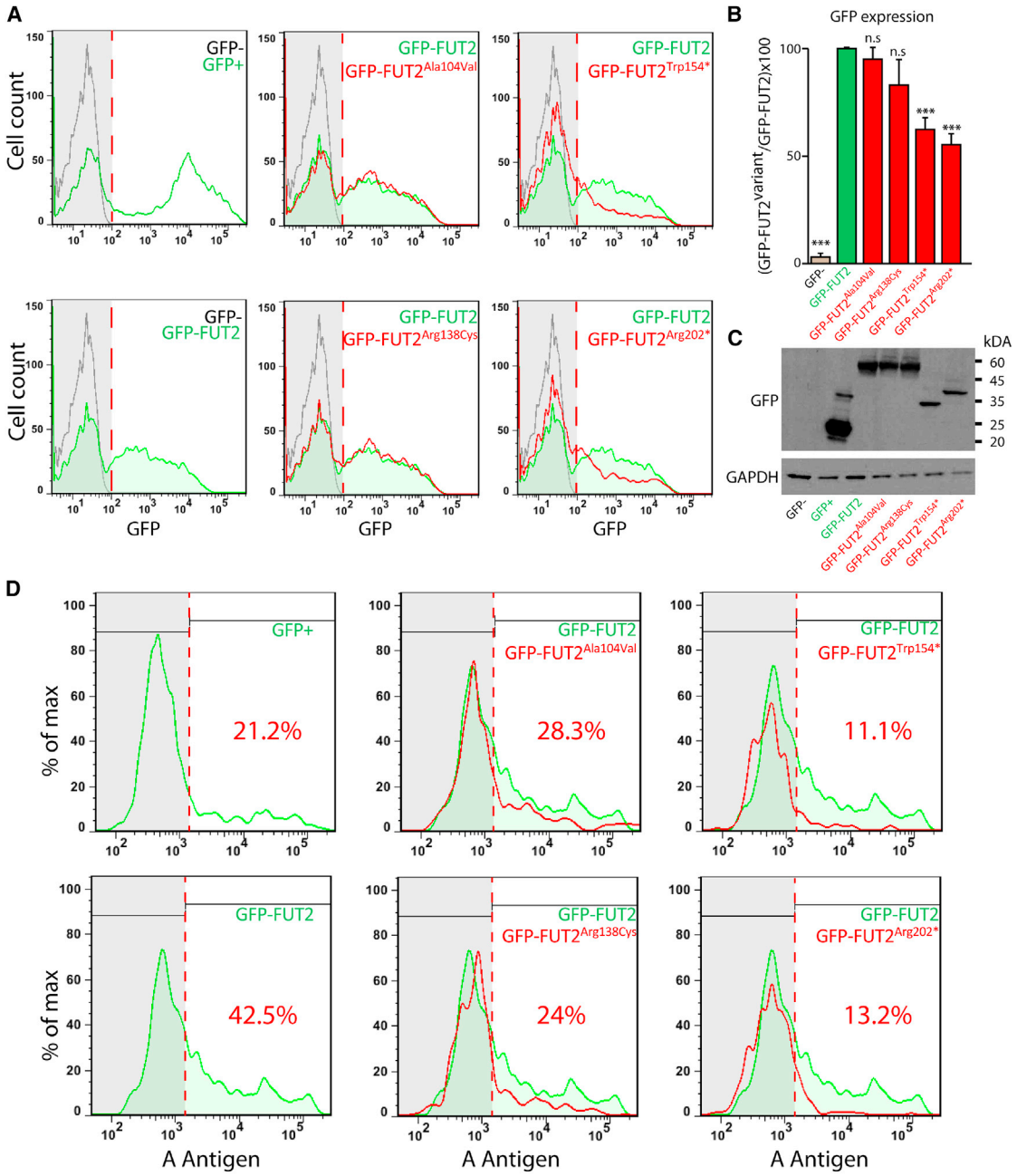

Figure 4. Flow Cytometry Profiles of GFP and Cell Surface A Blood Group Epitope Levels by COS-7 Cells Transfected with Human GFP-FUT2 Variant Constructs

(A) COS-7 cells were transiently transfected for 72 hr with GFP-FUT2^{WT} (green-filled traces) or GFP-FUT2 variant constructs (red traces). GFP-positive (green traces) and non-transfected (gray traces) cells were used as negative controls. GFP-FUT2^{WT}, GFP-FUT2^{Ala104Val}, and GFP-FUT2^{Arg138Cys} transfected cells show the same GFP expression profiles while GFP-FUT2^{Trp154*} and GFP-FUT2^{Arg202*} transfected cells show a decrease of GFP levels as compared to GFP-FUT2^{WT} cells, meaning that those variants are less expressed or degraded quicker in COS-7 cells.

(B) Quantification of the flow cytometry experiment shown in (A). ***$p \leq 0.001$; n.s.: non-significant. Error bars indicate ± SEM.

(C) Western blot analyses of COS-7 cells transfected with GFP-FUT2 variants. GFP and GAPDH blots are shown.

(D) COS-7 cells were transiently transfected for 72 hr with GFP, GFP-FUT2^{WT} (green), or GFP-FUT2 variant constructs (red). Cell surface A blood group epitope levels on GFP-positive cells were analyzed for each condition. 42.5% of GFP-FUT2^{WT} cells have A antigen as compared to GFP-transfected cells (21.2%). The accumulation of A antigen at the surface of cells transfected with GFP-FUT2^{Ala104Val}, GFP-FUT2^{Arg138Cys}, GFP-FUT2^{Trp154*}, or GFP-FUT2^{Arg202*} variant constructs was significantly reduced (28.3%, 24%, 11.1%, and 13.2%, respectively).

function mainly occurs during active infection. Because FUT2 controls surface expression of A and B antigens and their precursor H antigen in mucosal epithelia, we tested levels of these antigens in epithelial cells after variant transfection. These experiments confirmed reduction of A antigen levels due to both nonsense and missense

variants, as well as lower FUT2 levels due to nonsense variants, the latter possibly due to protein degradation and/or reduction of expression (Figure 4). These findings are consistent with the function of A, B, or H antigens as proteins used by bacteria for epithelial binding.[34] A antigen in particular is used by specific *Lactobacilli* strains in binding to intestinal mucosa, while *Escherichia coli* strains that express A-specific adhesins infected only individuals with blood group A.[35,36] These are consistent with our own findings of greater relative abundance of Lactobacillales and Gammaproteobacteria in middle ears of *FUT2* p.Trp154* carriers (Figures 1B, 1C, and S8). Unfortunately, we have middle ear microbiome data from only two individuals who are homozygous for p.Trp154*, but these two individuals have higher relative abundance of Streptococcaceae (Figure S8). Interestingly, virulent strains of *Streptococcus pneumoniae*, a common cause of acute otitis media, use the type I operon to primarily target Lewis antigen which is regulated not only by *FUT2* but also by *FUT3* (MIM: 111100);[37,38] this may imply that nonsecretor status due to homozygosity for *FUT2* p.Trp154* might decrease infection by other bacteria but increases predisposition to Streptococcal infection. In addition to affecting bacterial adherence, FUT2 nonsecretor status has been shown to lead to significantly greater lymphocyte infiltration during infection.[39] Taken together these findings suggest that each *FUT2* genotype may predispose to specific bacterial infections and might also be the reason for the apparent contradiction to previous GWAS findings, such that the heterozygous genotype confers susceptibility to bacteria that are known for chronic otitis media and therefore repeated childhood ear infections, while the homozygous genotype may cause acute otitis media e.g., Streptococcal, albeit recurrent.

Within the indigenous Filipino population, an elderly female who is heterozygous for *FUT2* p.Arg202* was the only person diagnosed with cholesteatoma (Figure 1A). Although numbers of individuals with cholesteatoma are small, we observed an increased prevalence of *FUT2* heterozygous genotypes in those with cholesteatoma due to chronic infection (Table S4). In individuals who carry the *FUT2* p.Trp154* variant, there is lower expression of *OSTM1* (MIM: 607649). In humans, *OSTM1* mutations cause autosomal-recessive, rare, severe osteopetrosis (MIM: 259720), a condition characterized by reduced osteoclast number and activity, bony sclerosis, fractures, reduced bone marrow cavities, pancytopenia, recurrent infections, and early death,[40] consistent with hematopoietic and osteoclast defects in *gl/Ostm1*-mutant mice.[41] Since we identified individuals with *FUT2* stop variants with cholesteatomas and/or downregulated *OSTM1* expression, reduced *FUT2* levels might play a role in cholesteatoma formation and bone pathology in part via decreased *OSTM1* levels.

In summary, we identified *FUT2* variants conferring risk for otitis media in a multi-ethnic cohort of well-phenotyped families. *FUT2* belongs to a select group of genes (e.g., *CFH* for age-related macular degeneration,[42] *CARD14* for psoriasis[43]) documented to have variants with a wide range of MAF and effect sizes in the same direction. The frequency of population-specific *FUT2* variants makes this gene a potential target for preventive screening and future treatments for otitis media, including modulation of the middle ear microbiome. Though *FUT2* upregulation during acute infection is transient, it is likely that the risk variants ultimately predispose to recurrent or chronic otitis media (Figure 1A; Tables 1, 2, S2, and S3). Our families demonstrate marked intra-familial genetic heterogeneity, suggesting that multiple combinations of common and rare variants plus environmental factors influence the individual otitis media phenotype as a complex trait.

Accession Numbers

Demultiplexed paired-end sequence data from middle ear microbiome samples were deposited in the NCBI Short Read Archive under accession number PRJNA439435.

Supplemental Data

Supplemental Data include eight figures and four tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.09.010.

Consortia

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

Acknowledgments

We thank all study participants. We also thank E. Valdivia for data collection efforts at UTMB, E. Leinonen for collection and handling of Finnish samples, the Office of Population Studies Foundation research and data collection teams, and C. Brands for general support. The Cebu Longitudinal Health and Nutrition Survey (CLHNS) was supported by NIH grants DK078150, TW005596, HL085144, and TW008288 and pilot funds from RR020649, ES010126, and DK056350. Exome sequencing was performed at the UWCMG which is funded by the US NIH via grants from the National Human Genome Research Institute and the National Heart, Lung and Blood Institute HG006493 (to D.A.N., M.J.B., and S.M.L.). This work was funded by the Philippine Council for Health Research and Development – Department of Science and Technology via the Balik Scientist Program and the NIH - National Institute on Deafness and Communication Disorders R01 DC015004 (both to R.L.P.S.-C.).

Declaration of Interests

The authors declare no competing interests.

Received: July 25, 2018
Accepted: September 19, 2018
Published: October 25, 2018

Web Resources

Bowtie 2, bowtie-bio.sourceforge.net/bowtie2/
Burrows-Wheeler Aligner, http://bio-bwa.sourceforge.net/
CADD, https://cadd.gs.washington.edu/
ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/
dbSNP, https://www.ncbi.nlm.nih.gov/projects/SNP/
edgeR, https://bioconductor.org/packages/release/bioc/html/edgeR.html
Explicet, http://www.explicet.org/
fathmm v2.3, http://fathmm.biocompute.org.uk/
featureCounts, http://bioinf.wehi.edu.au/featureCounts/
GATK, https://software.broadinstitute.org/gatk/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
gnomAD Browser, http://gnomad.broadinstitute.org/
GTEx Portal, https://gtexportal.org/home/
InnateDB, http://www.innatedb.ca/
MERLIN, http://csg.sph.umich.edu/abecasis/Merlin/
Mutation Assessor, http://mutationassessor.org/
MutationTaster, http://www.mutationtaster.org/
NetworkAnalyst, http://www.networkanalyst.ca/
OMIM, http://www.omim.org/
Phred, http://www.phrap.org/phredphrap/
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
PROVEAN, http://provean.jcvi.org
Python, https://www.python.org/
SeattleSeq, http://snp.gs.washington.edu/SeattleSeqAnnotation150/
Sequence Read Archive (SRA), http://www.ncbi.nlm.nih.gov/sra
silva, SINA Alignment Service, https://www.arb-silva.de/aligner/
STAR Aligner, https://github.com/alexdobin/STAR/releases
Superlink Online SNP, http://cbl-hap.cs.technion.ac.il/superlink-snp/
UCHIME, http://drive5.com/uchime/
UCSC Genome Browser, https://genome.ucsc.edu

References

1. Ferrer-Admetlla, A., Sikora, M., Laayouni, H., Esteve, A., Roubinet, F., Blancher, A., Calafell, F., Bertranpetit, J., and Casals, F. (2009). A natural history of FUT2 polymorphism in humans. Mol. Biol. Evol. *26*, 1993–2003.
2. Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. *48*, 709–717.
3. Tian, C., Hromatka, B.S., Kiefer, A.K., Eriksson, N., Noble, S.M., Tung, J.Y., and Hinds, D.A. (2017). Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat. Commun. *8*, 599.
4. Santos-Cortez, R.L.P., Chiong, C.M., Reyes-Quintos, M.R.T., Tantoco, M.L.C., Wang, X., Acharya, A., Abbe, I., Giese, A.P., Smith, J.D., Allen, E.K., et al.; University of Washington Center for Mendelian Genomics (2015). Rare *A2ML1* variants confer susceptibility to otitis media. Nat. Genet. *47*, 917–920.
5. Sale, M.M., Chen, W.M., Weeks, D.E., Mychaleckyj, J.C., Hou, X., Marion, M., Segade, F., Casselbrant, M.L., Mandel, E.M., Ferrell, R.E., et al. (2011). Evaluation of 15 functional candidate genes for association with chronic otitis media with effusion and/or recurrent otitis media (COME/ROM). PLoS ONE *6*, e22297.
6. Hafrén, L., Kentala, E., Järvinen, T.M., Leinonen, E., Onkamo, P., Kere, J., and Mattila, P.S. (2012). Genetic background and the risk of otitis media. Int. J. Pediatr. Otorhinolaryngol. *76*, 41–44.
7. Adair, L.S., Popkin, B.M., Akin, J.S., Guilkey, D.K., Gultiano, S., Borja, J., Perez, L., Kuzawa, C.W., McDade, T., and Hindin, M.J. (2011). Cohort profile: the Cebu longitudinal health and nutrition survey. Int. J. Epidemiol. *40*, 619–625.
8. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
9. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
10. Fishelson, M., and Geiger, D. (2002). Exact genetic linkage computations for general pedigrees. Bioinformatics *18* (*Suppl 1*), S189–S198.
11. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. *30*, 97–101.
12. Spielman, R.S., McGinnis, R.E., and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. *52*, 506–516.
13. Hara, N., Alkanani, A.K., Ir, D., Robertson, C.E., Wagner, B.D., Frank, D.N., and Zipris, D. (2012). Prevention of virus-induced type 1 diabetes with antibiotic therapy. J. Immunol. *189*, 3805–3814.
14. Markle, J.G., Frank, D.N., Mortin-Toth, S., Robertson, C.E., Feazel, L.M., Rolle-Kampczyk, U., von Bergen, M., McCoy, K.D., Macpherson, A.J., and Danska, J.S. (2013). Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. Science *339*, 1084–1088.
15. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
16. Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. *8*, 175–185.
17. Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. *8*, 186–194.
18. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics *27*, 2194–2200.
19. Schloss, P.D., and Westcott, S.L. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl. Environ. Microbiol. *77*, 3219–3226.
20. Pruesse, E., Peplies, J., and Glöckner, F.O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics *28*, 1823–1829.
21. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596.
22. Robertson, C.E., Harris, J.K., Wagner, B.D., Granger, D., Browne, K., Tatem, B., Feazel, L.M., Park, K., Pace, N.R., and Frank, D.N. (2013). Explicet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. Bioinformatics *29*, 3100–3101.
23. Hsiao, A., Ideker, T., Olefsky, J.M., and Subramaniam, S. (2005). VAMPIRE microarray suite: a web-based platform for

the interpretation of gene expression data. Nucleic Acids Res. *33*, W627-32.
24. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
25. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
26. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.
27. Xia, J., Benner, M.J., and Hancock, R.E. (2014). NetworkAnalyst–integrative approaches for protein-protein interaction network analysis and visual exploration. Nucleic Acids Res. *42*, W167-74.
28. Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S., and Lynn, D.J. (2013). InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res. *41*, D1228–D1233.
29. Chang, J.G., Ko, Y.C., Lee, J.C., Chang, S.J., Liu, T.C., Shih, M.C., and Peng, C.T. (2002). Molecular analysis of mutations and polymorphisms of the Lewis secretor type alpha(1,2)-fucosyltransferase gene reveals that Taiwan aborigines are of Austronesian derivation. J. Hum. Genet. *47*, 60–65.
30. Rehman, A.U., Santos-Cortez, R.L.P., Drummond, M.C., Shahzad, M., Lee, K., Morell, R.J., Ansar, M., Jan, A., Wang, X., Aziz, A., et al.; University of Washington Center for Mendelian Genomics (2015). Challenges and solutions for gene identification in the presence of familial locus heterogeneity. Eur. J. Hum. Genet. *23*, 1207–1215.
31. Maniu, A., Harabagiu, O., Perde Schrepler, M., Cătană, A., Fănuță, B., and Mogoantă, C.A. (2014). Molecular biology of cholesteatoma. Rom. J. Morphol. Embryol. *55*, 7–13.
32. Santos-Cortez, R.L.P., Hutchinson, D.S., Ajami, N.J., Reyes-Quintos, M.R.T., Tantoco, M.L.C., Labra, P.J., Lagrana, S.M., Pedro, M., Llanes, E.G., Gloria-Cruz, T.L., et al. (2016). Middle ear microbiome differences in indigenous Filipinos with chronic otitis media due to a duplication in the *A2ML1* gene. Infect. Dis. Poverty *5*, 97.
33. Gampa, A., Engen, P.A., Shobar, R., and Mutlu, E.A. (2017). Relationships between gastrointestinal microbiota and blood group antigens. Physiol. Genomics *49*, 473–483.
34. Goto, Y., Uematsu, S., and Kiyono, H. (2016). Epithelial glycosylation in gut homeostasis and inflammation. Nat. Immunol. *17*, 1244–1251.
35. Uchida, H., Kinoshita, H., Kawai, Y., Kitazawa, H., Miura, K., Shiiba, K., Horii, A., Kimura, K., Taketomo, N., Oda, M., et al. (2006). Lactobacilli binding human A-antigen expressed in intestinal mucosa. Res. Microbiol. *157*, 659–665.
36. Lindstedt, R., Larson, G., Falk, P., Jodal, U., Leffler, H., and Svanborg, C. (1991). The receptor repertoire defines the host range for attaching *Escherichia coli* strains that recognize globo-A. Infect. Immun. *59*, 1086–1092.
37. Hobbs, J.K., Pluvinage, B., and Boraston, A.B. (2018). Glycan-metabolizing enzymes in microbe-host interactions: the *Streptococcus pneumoniae* paradigm. FEBS Lett. Epub ahead of print. https://doi.org/10.1002/1873-3468.13045.
38. Soejima, M., and Koda, Y. (2005). Molecular mechanisms of Lewis antigen expression. Leg. Med. (Tokyo) *7*, 266–269.
39. Heneghan, M.A., Moran, A.P., Feeley, K.M., Egan, E.L., Goulding, J., Connolly, C.E., and McCarthy, C.F. (1998). Effect of host Lewis and ABO blood group antigen expression on *Helicobacter pylori* colonisation density and the consequent inflammatory response. FEMS Immunol. Med. Microbiol. *20*, 257–266.
40. Ott, C.E., Fischer, B., Schröter, P., Richter, R., Gupta, N., Verma, N., Kabra, M., Mundlos, S., Rajab, A., Neitzel, H., and Kornak, U. (2013). Severe neuronopathic autosomal recessive osteopetrosis due to homozygous deletions affecting OSTM1. Bone *55*, 292–297.
41. Pata, M., Héraud, C., and Vacher, J. (2008). OSTM1 bone defect reveals an intercellular hematopoietic crosstalk. J. Biol. Chem. *283*, 30522–30530.
42. Fritsche, L.G., Igl, W., Bailey, J.N., Grassmann, F., Sengupta, S., Bragg-Gresham, J.L., Burdon, K.P., Hebbring, S.J., Wen, C., Gorski, M., et al. (2016). A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat. Genet. *48*, 134–143.
43. Jordan, C.T., Cao, L., Roberson, E.D., Duan, S., Helms, C.A., Nair, R.P., Duffin, K.C., Stuart, P.E., Goldgar, D., Hayashi, G., et al. (2012). Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis. Am. J. Hum. Genet. *90*, 796–808.