SUPPLEMENTARY MATERIALS

Supplementary Methods

Haplotype analysis

We performed haplotype analysis of short tandem repeat (STR) marker genotypes to map meiotic recombination breakpoints defining the smallest region of overlap (SRO) of chromosome 7 co-segregating with EVA in six unrelated pairs of affected siblings.[1, 2]

Massively parallel sequencing

The dual-indexed, enriched library fragments were sequenced on an Illumina HiSeq1500 instrument using SBSv4 Paired-End (126 x 126) chemistry. The reads were de-multiplexed and fastq files were generated with bcl2fastq (v 2.17). Fastq files were transferred to the NIH Biowulf High-Performance Computing Cluster and aligned to the hg19 version of the human genome using Burrows-Wheeler Aligner (BWA v 0.7.13).[3] Annotated variants were loaded into a custom FilemakerPro database to sort and filter for shared variants among family members and zygosity.

Linkage disequilibrium mapping

We downloaded a Variant Call Format (VCF) file (https://vcftools.github.io/index.html) of a 1-Mb region of chromosome 7q (Chr7:106,500,000-107,500,000, hg19) in 503 Caucasian-European individuals (CEU, FIN, GBR, IBS, TSI) from the 1000 Genomes Project (Phase 3).[4] The genotype data were already phased into haplotypes at the time of download. We indexed the VCF files using Tabix[5] (http://www.htslib.org/doc/tabix.html), retained bi-allelic SNPs with minor allele frequency (MAF) > 0.01 using VCF tools[6] and converted the VCF file to a linkage pedigree file (PED) and a marker information file (INFO) using PLINK[7] (http://zzz.bwh.harvard.edu/plink/). We imported the PED file and INFO file into Haploview[8] (https://www.broadinstitute.org/haploview/haploview) to calculate pairwise values of D' and r^2.

Statistical comparisons of EVA chromosomes

We used genotype-haplotype analysis of the large region with modest LD (r^2>0.6) corresponding to the CEVA haplotype to determine the origin and segregation of chromosomes within families. When we could not resolve the number of independent chromosomal segments in a family, we determined the higher (non-conservative) and lower (conservative) possible numbers for that family. We used these numbers to calculate the cumulative highest and lowest possible numbers of chromosomes in the M0 and M2 cohorts.

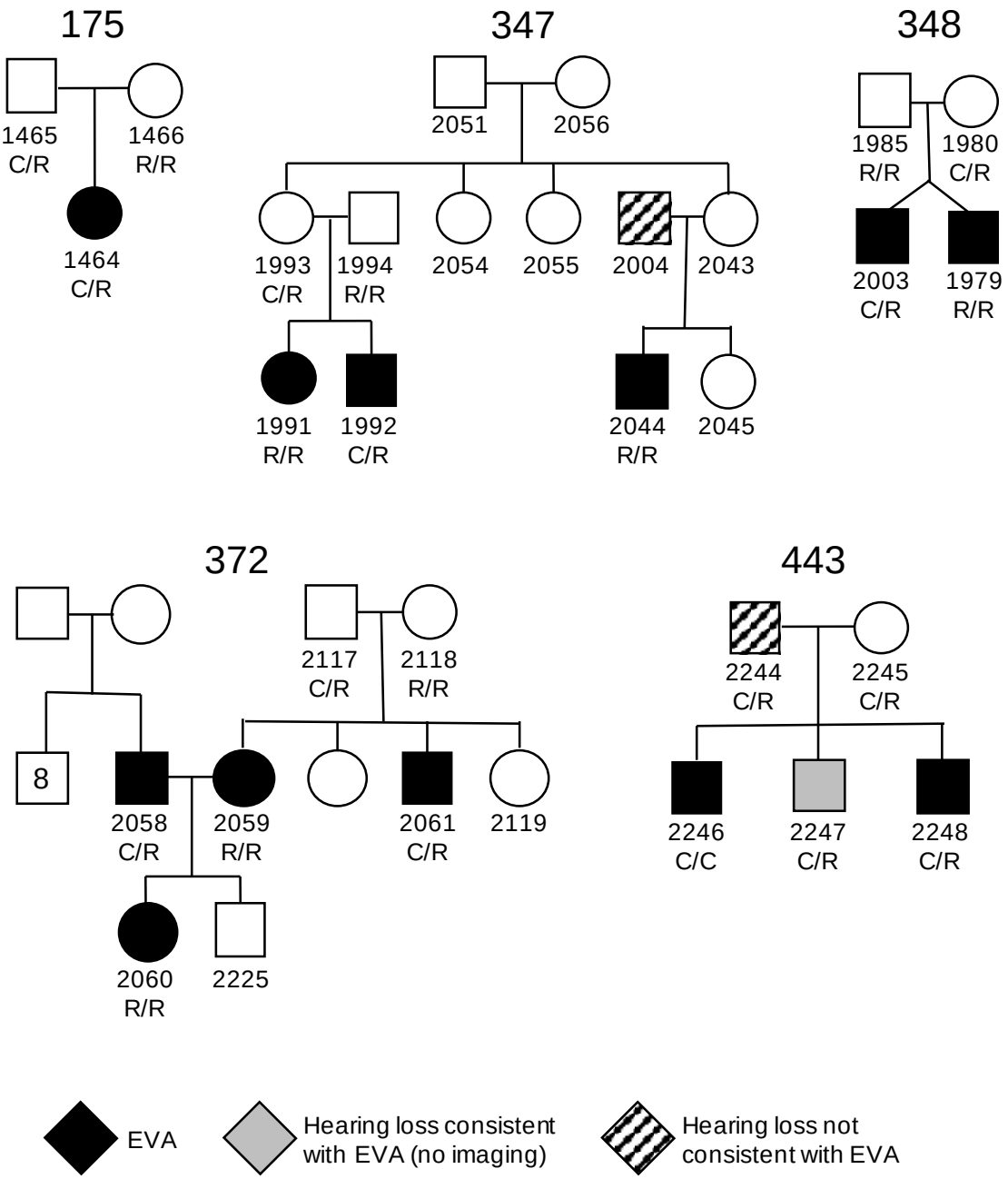
Supplementary Results

	rs17424561	rs79579403	rs17425867	rs117113959	rs17349280	rs117386523	rs80149210	rs199667576	rs9649298	rs117714350	rs199915614	rs150942317
rs17424561		0.96	0.93	0.91	0.91	0.89	0.91	0.91	0.91	0.94	0.94	0.94
rs79579403	0.89		0.98	0.96	0.96	0.93	0.96	0.95	0.95	0.97	0.97	0.97
rs17425867	0.85	0.96		0.93	0.93	0.91	0.93	0.93	0.93	0.94	0.94	0.94
rs117113959	0.79	0.89	0.85		1	0.98	1	1	1	1	1	1
rs17349280	0.79	0.89	0.85	1		0.98	1	1	1	1	1	1
rs117386523	0.75	0.85	0.81	0.96	0.96		0.98	1	1	0.97	0.97	0.97
rs80149210	0.79	0.89	0.85	1	1	0.96		1	1	1	1	1
rs199667576	0.77	0.87	0.83	0.98	0.98	0.98	0.98		1	0.97	0.97	0.97
rs9649298	0.77	0.87	0.83	0.98	0.98	0.98	0.98	1		0.97	0.97	0.97
rs117714350	0.62	0.68	0.63	0.74	0.74	0.69	0.74	0.71	0.71		1	1
rs199915614	0.62	0.68	0.63	0.74	0.74	0.69	0.74	0.71	0.71	1		1
rs150942317	0.62	0.68	0.63	0.74	0.74	0.69	0.74	0.71	0.71	1	1	

D' (upper triangle)

r^2 (lower triangle)

Supplementary Table 1 Pairwise linkage disequilibrium values between *SLC26A4*-linked variants. r^2 and D' values were calculated by Haploview[8] (https://www.broadinstitute.org/haploview/haploview) for 503 Caucasian-European individuals in the 1000 Genomes (phase 3) database.



Supplementary Figure 1 Segregation of EVA and the Caucasian EVA haplotype in M0 families**.** Families 347, 348, 372 and 443 show discordant inheritance of EVA and the CEVA haplotype. R, reference haplotype; C, Caucasian EVA (CEVA) haplotype.

REFERENCES

1 Pryor SP, Madeo AC, Reynolds JC, Sarlis NJ, Arnos KS, Nance WE, Yang Y, Zalewski CK, Brewer CC, Butman JA, Griffith AJ. SLC26A4/PDS genotype-phenotype correlation in hearing loss with enlargement of the vestibular aqueduct (EVA): evidence that Pendred syndrome and non-syndromic EVA are distinct clinical and genetic entities. *J Med Genet* 2005;42(2):159-65.

2 Choi BY, Stewart AK, Madeo AC, Pryor SP, Lenhard S, Kittles R, Eisenman D, Kim HJ, Niparko J, Thomsen J, Arnos KS, Nance WE, King KA, Zalewski CK, Brewer CC, Shawker T, Reynolds JC, Butman JA, Karniski LP, Alper SL, Griffith AJ. Hypo-functional SLC26A4 variants associated with nonsyndromic hearing loss and enlargement of the vestibular aqueduct: genotype-phenotype correlation or coincidental polymorphisms? *Hum Mutat* 2009;30(4):599-608.

3 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14):1754-60.

4 Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* 2015;526(7571):68-74.

5 Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* 2011;27(5):718-9.

6 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G. The variant call format and VCFtools. *Bioinformatics* 2011;27(15):2156-8.

7 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81(3):559-75.

8 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21(2):263-5.