Population-scale analysis of common and rare genetic variation associated with hearing loss in adults

Kavita Praveen[1], Lee Dobbyn[1], Lauren Gurski[1], Ariane H. Ayer[1], Jeffrey Staples[1], Shawn Mishra[2], Yu Bai[2], Alexandra Kaufman[2], Arden Moscati[1], Christian Benner[1], Esteban Chen[1], Siying Chen[1], Alexander Popov[1], Janell Smith[2], GHS-REGN DiscovEHR collaboration[*], Regeneron Genetics Center[*], Decibel-REGN collaboration[*], Olle Melander[3,4], Marcus Jones[1], Jonathan Marchini[1], Suganthi Balasubramanian[1], Brian Zambrowicz[2], Meghan Drummond[2], Aris Baras[1], Goncalo R. Abecasis[1], Manuel A. Ferreira[1], Eli A. Stahl[1], Giovanni Coppola[1]

From:

[1]Regeneron Genetics Center, Tarrytown, NY 10591, USA.

[2]Regeneron Pharmaceuticals, Inc., Tarrytown, NY 10591, USA.

[3]Lund University, Department of Clinical Sciences Malmö, Malmö, Sweden.

[4]Skåne University Hospital, Department of Emergency and Internal Medicine, Malmö, Sweden.

*A complete list of contributing authors to the Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration and the Decibel-REGN Collaborations is provided in the Supplementary Appendix.

Correspondence to giovanni.coppola@regeneron.com.

ABSTRACT

Understanding the genetic underpinnings of disabling hearing loss, which affects ~466 million people worldwide, can provide avenues for new therapeutic target development. We performed a genome-wide association meta-analysis of hearing loss with 125,749 cases and 469,497 controls across five cohorts, including UK Biobank, Geisinger DiscovEHR, the Malmö Diet and Cancer Study, Mount Sinai's Bio*Me* Personalized Medicine Cohort, and FinnGen. We identified 53 loci affecting hearing loss risk, 15 of which are novel, including common coding variants in *COL9A3 and TMPRSS3*. Through exome-sequencing of 108,415 cases and 329,581 controls from the same cohorts, we identified hearing loss associations with burden of rare coding variants in *FSCN2* (odds ratio [OR] = 1.14, $P = 1.9 \times 10^{-15}$) and burden of predicted loss-of-function variants in *KLHDC7B* (OR = 2.14, $P = 5.2 \times 10^{-30}$). We also observed single-variant and gene-burden associations with 11 genes known to cause Mendelian forms of hearing loss, including an increased risk in heterozygous carriers of mutations in the autosomal recessive hearing loss genes *GJB2* (Gly12fs; OR = 1.21, $P = 4.2 \times 10^{-11}$) and *SLC26A5* (gene burden; OR = 1.96, $P = 2.8 \times 10^{-17}$). Our results suggest that loss of KLHDC7B function increases risk for hearing loss, and show that Mendelian hearing loss genes contribute to the burden of hearing loss in the adult population, suggesting a shared etiology between common and rare forms of hearing loss. This work illustrates the potential of large-scale exome sequencing to elucidate the genetic architecture of common traits in which risk is modulated by both common and rare variation.

INTRODUCTION

The loss of hearing can have a debilitating impact on quality of life, requiring major adjustments to day-to-day activities. Significant comorbidities are also associated with hearing loss including social isolation, depression, cognitive impairment, and dementia, which further deteriorate quality of life[1]. Disabling hearing loss is common, with ~466 million people affected worldwide (World Health Organization: https://www.who.int/health-topics/hearing-loss). While existing technologies – such as hearing aids and cochlear implants – can ameliorate hearing loss, their use is limited by barriers to access, including cost, health policies and regulations, and social stigma associated with device use[2]. Furthermore, while these assistive devices typically provide some benefit, they do not address the chief complaint associated with acquired hearing loss: lack of hearing clarity, particularly in social and work environments (https://www.hearingloss.org/hlaa-pfdd/). The Lancet Commission on dementia prevention, intervention and care has identified untreated hearing loss in middle age as the top modifiable risk factor for dementia, but it is estimated that 67-86% of adults who may benefit from hearing aids do not use them[1,3]. These challenges, combined with a lack of therapeutics to stop or slow hearing loss progression, have contributed to its status as a growing global health issue. Novel therapies based on genetic evidence, therefore, will be crucial in addressing this unmet need.

Hearing loss affects individuals of all ages, but its prevalence increases with age. Approximately 1-2/1,000 babies are born with hearing loss[4]. Mutations in over 150 genes (hereditaryhearingloss.com) account for over 50% of the cases. While autosomal recessive hearing loss is generally pre-lingual and non-progressive, autosomal dominant forms are mostly post-lingual (including adult onset) and progressive. The prevalence of hearing loss increases to 2.8/1,000 in primary school-age children and 3.5/1,000 in adolescents[4,5]. The National Institute on

Deafness and other Communication Diseases calculates that by the age of 45 ~2% of individuals have a disabling hearing loss, and this number increases to 50% in individuals over the age of 75. This increase in prevalence with age reflects a combination of late-onset hearing loss mutations, the cumulative effects of environmental factors such as exposure to noise and ototoxic drugs and, in aging individuals, the degenerative effects of age on the cochlea. These genetic and environmental insults primarily damage the structures of the inner ear, resulting in sensorineural hearing loss[6,7].

Heritability estimates for age-related hearing loss range as high as 36 and 70%[8–12] suggesting that genetics, along with environmental factors, play a significant role in determining an individual's risk for developing hearing loss. Genome-wide association studies (GWAS) of hearing loss in adults have identified 61 common variant loci associated with the trait in Europeans[13–18]. While the majority of these studies have established a common variant contribution to adult hearing loss, there are few reports addressing rare and low-frequency variant contribution[19].

Recently, Ivarsdottir et al[18] published association results with hearing loss on ~50K Icelandic individuals with whole-genome sequence and ~50K individuals from UKB with exome sequence data, with imputation of larger samples into these variant sets. We have now expanded the rare variant analysis to exome-sequences from ~295K individuals in UKB and ~143K individuals from three other datasets. Here, we report findings from genome- and exome-wide association meta analyses with a total of 125,749 cases and 469,497 controls. Our analyses have identified 15 novel susceptibility loci and 15 rare variant associations that provide novel insights into the biology of hearing loss in adults.

RESULTS

To study common variants, we performed association meta-analyses using genotyping and imputation across five cohorts: Geisinger DiscovEHR study (GHS), the Malmö Diet and Cancer study from Malmö, Sweden (MALMO), Mount Sinai's Bio*Me* Personalized Medicine Cohort from Mount Sinai Health System, New York (SINAI), UK Biobank (UKB) and an additional study from Finland, FinnGen, for a total of 125,749 cases and 469,497 controls. To study rare and ultra-rare variants, we also generated exome sequence data and performed combined GWAS and exome-wide association study (ExWAS) on a subset of 108,415 cases and 329,581 controls across GHS, MALMO, SINAI, and UKB. Phenotypes were derived from ICD-10 diagnosis codes in GHS, MALMO, SINAI and FinnGen, and combined self-report and ICD-10 codes in UKB (see Methods and Supplementary Table 1 for details). Our genome-wide association analyses included 15,881,489 variants with frequency > 0.1% that were genotyped or imputed with r2 > 0.3 in at least one study, and 2,923,124 coding or essential splice site variants with minor allele count at least five from exome sequencing (111,588 of which overlapped with the imputed).

Novel common variant associations

We identified 53 independent loci harboring genome-wide significant ($P < 5 \times 10^{-8}$) common (MAF ≥ 1%) variants associated with hearing loss (Figure 1A, Supplementary Table 2, Supplementary Figure 1), 38 of which are shared with the previously reported 61 hearing loss-associated loci. We observed a genomic control lambda_GC = 1.36 but an LD score regression intercept 1.054 (standard error 0.008), indicating that inflation was largely due to polygenic signal for the hearing loss phenotype.

107 Among the lead variants at the 15 novel loci (Figure 1A, Table 1, Supplementary Figure
108 2) is rs117887149 that maps close to *GJB2* (Supplementary Figure 2H), which is a predominant
109 cause of congenital hearing loss. While majority of the lead variants in the novel loci lay in
110 intergenic or downstream/upstream regions of genes, at two loci they were within the introns of
111 the following genes: *KCTD10,* a member of the potassium channel tetramerization domain family
112 that is implicated in cardiac development[20,21] and *MAP7D2,* an axonal cargo transport protein
113 predominantly expressed in the brain[22].

114 Top variants at two other novel loci were missense changes and thus also implicate specific
115 genes: in *COL9A3* (Arg103Trp; MAF = 0.066) and *TMPRSS3* (Ala90Thr; MAF = 0.055).
116 Mutations in *COL9A3*, which is highly expressed in the ear, have been implicated in the autosomal
117 recessive Stickler syndrome, in which hearing loss is prominent[23], and tentatively in non-
118 syndromic hearing loss[24,25]. High-throughput mouse knockout characterization from the
119 International Mouse Phenotyping Consortium (IMPC) indicates that *Col9a3*-null mice also have
120 hearing loss[26]. *TMPRSS3* is a type II membrane serine protease that localizes to the endoplasmic
121 reticulum and plasma membranes, and is expressed in hair cells and supporting cells in the organ
122 of Corti, the spiral ganglion, and the stria vascularis in the ear[27–29]. Mutations in *TMPRSS3* cause
123 congenital and childhood-onset autosomal recessive hearing loss[30] but there is also evidence for
124 hearing deficits in heterozygous carriers[18].

125 The number of associated variants at the 50 autosomal loci ranged from eight to 1,326. In
126 order to assess the presence of multiple independent causal variants at each locus, we ran
127 conditional analyses using GCTA-COJO[31], which indicated the presence of secondary association
128 signals (joint $P < 10^{-5}$) at eight loci (Supplementary Table 3). We also ran FINEMAP Bayesian
129 causal variant inference[32] to prioritize associated variants as credibly causal at thirty loci that were

genome-wide significant in RGC data (excluding FinnGen). The COJO and FINEMAP methods disagreed on the presence of single vs. multiple association signals within five loci, three of which showed subthreshold ($10^{-5} < P < 10^{-4}$) secondary associations in COJO. FINEMAP prioritized ten or fewer variants in top causal variant 95% credible sets for the top causal variant at eleven loci, including single putatively causal variants at four loci (Supplementary Table 3): missense variants in *CDH23* (Ala371Thr; rs143282422) and *KLHDC7B* (Val504Met; rs36062310), and intronic variants in *CTBP2* (rs10901863) and *PAFAH1B1* (rs12938775).

Colocalization with GTEx eQTLs identifies candidate genes driving GWAS signals.

To identify genes for which expression regulation might drive the observed association signals, we tested for colocalization of our hearing loss-associated loci with expression quantitative trait loci (eQTL) data for 48 tissues from the Genotype-Tissue Expression (GTEx) project (see URLs) using coloc2[33]. Across all GTEx eQTL tissues tested, we identified 19 genes mapping to 15 loci with evidence (posterior probability of colocalization, PPH4 ≥ 0.5) for colocalization between the hearing loss association and an eQTL signal, in at least one tissue (Supplementary Table 4). Only two of the 15 loci with GWAS-eQTL overlap had multiple eQTL signals for more than one gene (*NUCKS1* and *RAB29* in locus 1-4; *ACADVL*, *DLG4*, *CTDNEP1* and *CLDN7* in locus 17-2) making it difficult to prioritize causal genes at these loci based on eQTL data. Since GTEx did not include tissues from the ear, we used single-cell RNA sequencing data that were generated in-house from mouse cochleae to check the expression of the 19 genes in the ear (Supplementary Table 5). Thirteen of the 19 genes showed evidence of expression across 26 inner-ear cell-types and, of these, 12 were expressed in the hair cells. While the majority of genes showed broad expression across the 26 cell types, we noted a subset that were specific to only a few,

153 including *CRIP3* in inner and outer hair cells, and *TCF19* in neurons and immune cells
154 (Supplementary Tables 4 & 5).
155
156 **Rare-variant association analysis identifies novel large-effect hearing loss variants in known**
157 **hearing loss genes.**
158 We identified significant ($P < 5 \times 10^{-8}$) rare variant (MAF < 1%) associations in 25 genes,
159 15 of which had nonsynonymous variant or gene burden associations (Figure 1B, Table 2,
160 Supplementary Tables 6 & 7). Five of the 15 genes also had significant common variant
161 associations within 1 Mb *(KLHDC7B, SYNJ2, GJB2, EYA4,* and *CDH23)*. After conditioning on
162 independent common variants at each locus (Supplementary Tables 6, 7 & 8), the rare variant and
163 gene burdens remained associated with hearing loss (maximum conditional $P \leq 3 \times 10^{-5}$) except
164 for the *CDH23* Asn1103Ser association (conditional $P = 0.79$). Of note, the lead common variant
165 in *CDH23* is also a missense (Ala371Thr) variant (MAF = 1.1%) and is pinpointed by FINEMAP
166 as the only causal variant at that locus with high confidence. Overall, our conditional analyses
167 suggest that rare variant association signals are usually independent of nearby common variant
168 associations.
169
170 *Associations with Mendelian hearing loss genes*
171 Of the 14 genes with independent, rare nonsynonymous and/or burden associations, 11
172 were previously identified as causes of Mendelian forms of hearing loss. These include
173 associations with two genes (*GJB2* and *SLC26A5)* that cause recessive hearing loss, burden
174 associations in seven genes (*MYO6, COCH, TECTA, SIX1, CEACAM16, POU4F3*, and *EYA4*) and
175 single-variant associations in two genes (*TBC1D24* and *COL11A2*) that cause autosomal dominant

176 hearing loss (reviewed in Shearer et. al., 1999[34]). In *MYO6* and *COCH*, we also observed genome-
177 wide significant single-variant associations of large effect size (*MYO6* His246Arg, OR=30.7;
178 *COCH* Cys542Phe, OR=81.4; Supplementary Table 6). Both variants have previously been
179 characterized as pathogenic in family-based genetic analyses[35,36]. Only a minority of variants
180 included in burden tests have been classified as pathogenic by ClinVar (Supplementary Table 9)
181 suggesting that our analysis has detected additional novel, risk-associated variants with variable
182 penetrance in Mendelian hearing loss genes. We also identified single-variant associations in two
183 genes that cause autosomal dominant hearing loss: *TBC1D24* Asn307Ser, also recently reported
184 by Ivarsdottir et al[18], was implicated as pathogenic in two unrelated families[37], and *COL11A2*
185 Phe80Ser, which has not yet been classified as pathogenic but is predicted deleterious, lies in a
186 domain (Laminin G-like/NC4) that harbors other mutations causing non-syndromic hearing loss[38].
187 We performed our analysis under an additive model, which assumes risk effects in
188 heterozygous carriers as well as homozygotes; therefore, it is not surprising to see that the majority
189 of Mendelian genes (9/11) identified can cause hearing loss in heterozygote carriers. However, we
190 also detect associations in two genes that have previously been implicated in recessive hearing
191 loss: *GJB2* Gly12fs (OR = 1.21; $P = 4 \times 10^{-11}$), and *SLC26A5* Leu46Pro (OR = 1.3; $P = 3 \times 10^{-14}$)
192 as well as the burden of *SLC26A5* predicted loss-of-function (pLOF) and strict deleterious
193 missense variants (excluding Leu46Pro) (OR = 1.96; $P = 3 \times 10^{-17}$) (Figure 2). We also observed
194 a suggestive association with *GJB2* Leu90Pro (OR = 1.51, $P = 4.4 \times 10^{-5}$), another known
195 pathogenic variant in recessive hearing loss. The associations persisted after excluding
196 homozygous carriers and compound heterozygous carriers of rare, coding variants in these genes
197 (Supplementary Table 10), suggesting a previously unappreciated increase in risk for hearing loss
198 in heterozygous carriers of loss-of-function *GJB2,* and missense and pLOF *SLC26A5* variants.

199

200 *Rare variant associations in KLHDC7B, FSCN2,* and *SYNJ2*

201 The most significant rare coding association in our analysis was the aggregate of 68 pLOF
202 variants *(*43 frameshift, 23 stop-gain and 2 stop-loss, (Supplementary Table 9)) in *KLHDC7B*
203 (Kelch-like domain containing 7B), with an approximately two-fold increase in risk for hearing
204 loss (OR = 2.14, $P = 5 \times 10^{-30}$). The main contributors to the gene burden were two frameshift
205 variants, Gly302fs and Lys181fs, that are predicted to truncate the protein near the start of the
206 Kelch domains. These variants were also significantly associated in single-variant tests (Figure 3),
207 and Gly302fs was recently reported in an analysis that included UKB[18]. The association with pLOF
208 variants remained significant after repeating the pLOF burden test conditioning on Gly302fs and
209 Lys181fs ($P = 8 \times 10^{-8}$; Supplementary Table 7). In addition, we observed a common (MAF = 4%)
210 missense variant (Val504Met) within the last Kelch domain of *KLHDC7B* associated with
211 increased risk for hearing loss (OR = 1.14, $P = 4 \times 10^{-26}$). This variant was also prioritized as the
212 sole causal variant at this locus by FINEMAP (Supplementary Table 3).

213 We also identified rare coding associations in fascin actin-bundling protein 2 (*FSCN2)* and
214 synaptojanin 2 (*SYNJ2)* with increased risk for hearing loss, both of which were recently observed
215 in UKB[18] (Figure 4). In *FSCN2*, an actin cross-linking protein[39–41], an aggregate of pLOF and
216 deleterious missense variants was associated with hearing loss, with majority of the carriers in the
217 burden having the His138Tyr variant. Conditioning on His138Tyr attenuated the significance of
218 the burden test ($P = 3 \times 10^{-6}$ after conditioning on His138Tyr; Supplementary Table 7) but did not
219 eliminate the signal, suggesting that other variants in *FSCN2* may increase the risk for hearing
220 loss. Mice homozygous for loss-of-function mutations in *Fscn2* present progressive hearing loss
221 starting at 3 weeks and near deafness by 24 weeks due to degeneration of the outer hair cells in the

222 cochlea[41]. In *SYNJ2*, we identified an association with a missense (Thr656Met)[18] variant that lies
223 in the catalytic domain of this inositol polyphosphate 5-phosphatase[42]. Mice harboring
224 homozygous mutations in *Synj2* that are predicted to reduce protein levels or the 5-phosphatase
225 catalytic activity show progressive high-frequency hearing loss and a degeneration of hair cells
226 that is most profound in the outer hair cells[43,44].
227
228 **GWAS/ExWAS supports a highly polygenic architecture of adult hearing loss.**
229 Given that this is the largest sequencing study to date of adult hearing loss, and given the
230 presence of Mendelian hearing loss genes among our common and rare-variant associations, we
231 sought to explore the distribution of effect sizes across allele frequencies (Figure 5). Hearing loss-
232 associated variants span the frequency spectrum and, while we observe a few rare variants of large
233 effect (e.g. *COCH* Cys542Phe and *MYO6* His246Arg), we do not observe any common variants
234 of large effect. We further estimated phenotypic variance explained by the genetic data, or
235 heritability h^2_{Tot}, using LD score regression (LDSC) partitioning into functional categories and
236 stratifying by minor allele frequency[45,46]. We estimated a total heritability (h^2_{Tot}) of 0.089, with
237 contributions from common variation (h^2_{CV}, MAF > 0.05) and low-frequency variation (h^2_{LFV},
238 0.001 < MAF ≤ 0.05) of 0.074 and 0.015, respectively (Supplementary Figure 3, Supplementary
239 Table 11). These results indicate that, while the bulk of SNP heritability is derived from common
240 variation, low-frequency variation contributes 16.8% of total SNP heritability.

241 **DISCUSSION**

242 We performed combined GWAS and ExWAS using exome sequencing data and identified
243 53 independent associations, of which 15 were novel. Novel associations included a missense lead
244 variant in *TMPRSS3*, a known cause of Mendelian hearing loss, adding to the tally of Mendelian
245 deafness genes (*EYA4, CDH23, TRIOBP*) showing common coding-variant associations with
246 hearing loss in adult humans. We also identified coding variant/gene burden associations with 15
247 genes through exome sequencing. We estimated that low-frequency variation contributes a non-
248 negligible portion (16.8%) of SNP heritability for adult hearing loss, consistent with a highly
249 polygenic genetic architecture with rare, low-frequency and common genetic variation for adult
250 hearing loss, and where variants of large effect would be subject to purifying selection.

251 Notably, the majority of genes implicated by rare-variant and burden associations are
252 already known to cause Mendelian forms of hearing loss; however, the odds ratios for these
253 associations are wide-ranging. At the lower end of the effect spectrum are single-variant
254 associations in *GJB2* and *SLC26A5,* with ORs close to typical for low-frequency GWAS findings
255 (1.2-1.3), followed by *COL11A2* (2.2-7) and, at the highest end of the distribution, *COCH* and
256 *MYO6* (31-81). *COCH* Cys542Phe and *MYO6* His246Arg are known pathogenic variants, and
257 consistent with this, we see almost all carriers presenting hearing loss. The two control carriers for
258 each variant (Supplementary Table 6) could be explained by imperfect ascertainment of the
259 phenotype or, as these mutations cause late-onset and progressive hearing loss, could also reflect
260 a difference in expressivity. While some dilution of the effect sizes may be expected when working
261 with self-reported as opposed to objective measures of hearing loss, the broad range in ORs that
262 we observe suggests that for several of these Mendelian hearing loss genes we are identifying
263 carriers with incompletely penetrant, risk-increasing variants. These results are consistent with the

264 hypothesis that there is a continuum between Mendelian and common forms of hearing loss with
265 the same genes harboring mutations 'causal' for the former, and risk-increasing for the latter.
266 Furthermore, compared to epidemiological risk factors for common hearing loss, including noise
267 exposure (OR~1.5-3[47]), odds ratios of low-frequency and rare genetic factors may be large enough
268 to provide mechanistic insights and to have implications for precision medicine.

269 Ultimately, to understand the biology of hearing loss and develop testable hypotheses from
270 our association results, we need to identify the genes driving the observed associations at each
271 locus. For common variants, we note the high resolution of FINEMAP to prioritize causal variants
272 in several loci, including pinpointing single variants in four genes, three of which are related to
273 hearing loss (*CDH23*), hearing function (*CTBP2*), or have significant rare variant associations with
274 hearing loss (*KLHDC7B*). While FINEMAP helps prioritize variants, the gene that is impacted by
275 those variants is not necessarily obvious, so we looked for colocalization with GTEx eQTL data
276 and identified 19 genes whose expression may influence hearing loss risk (Supplementary Table
277 4). Analysis of single-cell expression data from mouse ears showed that 13 of the 19 genes are
278 expressed in the ear. Of note, among these are *LMO7* and *SPTBN1* (BetaII-spectrin). Both genes
279 code for components of the cuticular plate in the hair cells of the ear and knockout mice for either
280 gene develop hearing loss[48,49]. *LMO7* colocalizes with eQTLs for the gene in GTEx skeletal muscle
281 and *SPTBN1* in thyroid tissue, such that decreased expression of these genes is correlated with
282 increased risk for hearing loss in humans.

283 One of the most compelling ways in which human genetics can establish a role for a gene
284 in disease is when a diverse set of rare nonsynonymous or loss-of-function variants in the same
285 gene show consistent association with disease. Coding variant and gene burden associations that
286 we identified included associations with known pathogenic (e.g. in *MYO6* and *COCH*) and novel

287 variants (*COL11A2* Phe80Ser), in genes that can cause autosomal dominant Mendelian hearing
288 loss. We also identified variants and genes that cause recessive hearing loss (*GJB2* Gly12fs and
289 *SLC26A5* pLOF+strict deleterious missense mutations), associated with increased risk (OR~1.2-
290 1.3) for hearing loss in heterozygous carriers. Missense mutations in *GJB2* can cause dominant,
291 syndromic hearing loss[50] but loss-of-function mutations, including Gly12fs, cause recessive non-
292 syndromic hearing loss[51–53]. Consistent with our findings, studies have detected hearing deficiency
293 at high frequencies and a possibly more prominent effect in female adult heterozygous carriers of
294 Gly12fs[53–55]. Mutations in *SLC26A5* cause recessive hearing loss in humans[56] and mice
295 homozygous for *Slc26a5* null mutations develop hearing loss as early as 5-7 weeks of age[57].
296 *Slc26a5* heterozygous null mice have a hearing deficiency that is intermediate between the controls
297 and knockouts. While we cannot rule out the possibility of compound heterozygosity with
298 variation in regulatory/promoter regions in heterozygous carriers, our results offer the possibility
299 that heterozygous carriers of loss-of-function variants in *GJB2* and *SLC26A5* may also have
300 increased risk for adult-onset hearing loss.

301 Three of the 15 coding and burden associations were in genes that have not previously been
302 implicated in Mendelian hearing loss in humans: *KLHDC7B, FSCN2*, and *SYNJ2*. In *KLHDC7B,*
303 we confirmed a previously reported association of the common Val504Met and a rare frameshift
304 variant with increased risk for hearing loss[18,58], and identified associations of increased risk with a
305 burden of additional pLOF variants. While the biological effect of the missense is difficult to
306 interpret, the association of putative LOFs with increased risk for hearing loss suggests that loss
307 of KLHDC7B function is deleterious for hearing function. Based on the smaller effect of
308 Val504Met (OR = 1.14) compared to the pLOF burden (OR=2.14), we would hypothesize that this
309 more common missense (MAF = 4%) is a hypomorph. *KLHDC7B* (Kelch-like domain containing

310 7B) is a relatively understudied gene; it has been characterized as a 594-amino-acid protein
311 containing a Kelch domain that is hypermethylated and upregulated in breast cancer cell lines and
312 may influence cell proliferation in MCF-7 cells lines[59,60]. In the mouse ear, RNAseq expression
313 profiling[61] and our real-time PCR data (Supplementary Figure 4) show *Klhdc7B* expression in the
314 cochlea with enrichment in the outer hair cells[62]. Consistent with the hypothesis from our genetic
315 findings that loss of KLHDC7B function increases risk for hearing loss, initial characterization of
316 *Klhdc7B* null mice by IMPC showed that homozygous carriers have hearing loss[26]. Mouse models
317 *Fscn2* and *Synj2* also develop hearing loss[41,43,44]. Based on our association results, it would also be
318 interesting to test heterozygous null *Klhdc7b*, *Fscn2*, and *Synj2* animals for increased susceptibility
319 to hearing loss with age or environmental insults such as noise-exposure.

320 We recognize that our study has several limitations. Our phenotype includes (in UKB) self-
321 reported hearing loss among adults, which is likely to be a heterogeneous mix of early-onset, late-
322 onset, age-related, as well as hearing loss due to environmental insults. In general, greater
323 phenotype precision, including environmental exposure measures, should help future genetic
324 analyses of adult hearing loss. We note that our colocalization analyses utilized GTEx eQTL data
325 across tissues not including ear expression data. Given the high degree of sharing of genetic
326 regulation of expression across tissues[63,64], the results are likely to point to the causal genes in many
327 instances. eQTL data for ear cell types should help with the interpretation of genetic analyses of
328 adult hearing loss.

329 In summary, this work contributes to connecting the two ends of the genetic architecture
330 of hearing loss by detecting a common signal in genes known to cause hearing loss in Mendelian
331 fashion and by detecting an additive signal (i.e. increased risk in heterozygous carriers) in genes
332 known to cause autosomal recessive hearing loss. This latter finding also connects young- and

333 adult-onset hearing loss in a single phenotypic spectrum with complex genetic underpinnings,
334 including contributions from rare and common variation.

335 **METHODS**

336 **Participating cohorts and phenotype data**

337 We performed meta-analysis for hearing loss on a total of 125,749 cases and 469,495
338 controls of European ancestry from the following cohorts: United Kingdom Biobank (UKB)[65], the
339 MyCode Community Health Initiative cohort from Geisinger Health System (GHS)[66], the Mount
340 Sinai Biome cohort (SINAI) (see URLs), the Malmö Diet and Cancer study (MALMO)[67], and
341 FinnGen R3 (URLs). Hearing loss case-control status in GHS, MALMO and SINAI was defined
342 using ICD-10 code diagnoses from the EHR. Cases were individuals with ICD10 H90.3-H90.8,
343 H91.1or H91.9 diagnoses. Controls were individuals who did not meet the case criteria and did
344 not have a diagnosis for ICD-10 Q16 (congenital malformations of ear causing hearing
345 impairment) or ICD-10 H93.1 (tinnitus). In UKB the phenotype was defined using ICD-10 codes
346 as above or by self-reported hearing loss or complete deafness on the touchscreen questionnaire
347 (Field IDs: 2247 and 2257). Controls in UKB were individuals who did not have an ICD-10
348 diagnosis for hearing loss or tinnitus and did not self-report hearing loss, deafness or tinnitus (Field
349 IDs: 4803 and non-cancer illness code 1597). The FinnGen analysis used was
350 finngen_r3_H8_CONSENHEARINGLOSS, conductive or sensorineural hearing loss defined by
351 ICD10 codes H90[.0-8], versus controls excluding other ear disorders (H91-H95). Further details
352 are given in the Supplementary Note.

353

354 **Genetic data and association analyses**

355 High-coverage whole-exome sequencing was performed at the Regeneron Genetics Center
356 as previously described[68,69]. For SINAI and MALMO, DNA from participants was genotyped on
357 the Global Screening Array (GSA), and for GHS genotyping was done on either the Illumina
358 OmniExpress Exome (OMNI) or GSA array; the datasets (stratified by array for GHS) were

359 imputed to the TOPMed (GHS) or HRC (MALMO and SINAI) reference panels using the
360 University of Michigan Imputation Server or the TOPMed Imputation Server (URLs). Additional
361 details are given in the Supplementary Note. We tested for association with hearing loss genetic
362 variants or their gene burdens using REGENIE v1.0.43[70]. Analyses were adjusted for age, age2,
363 sex, an age-by-sex interaction term, experimental batch-related covariates, and genetic principal
364 components. Cohort-specific statistical analysis details are provided in the Supplementary Note.
365 Results across cohorts were pooled using inverse-variance weighted meta-analysis.
366
367 **Fine mapping and follow-on genetic analyses.**
368 We defined genome-wide significant loci in our analysis by linkage disequilibrium ($r^2 >$
369 0.1) with lead variants. We defined previously associated loci by their index variants reported in
370 previous hearing loss GWAS[13–17,58,71–75], and excluded 1 Mb regions surrounding them in the
371 identification of novel loci in our analysis. LD score (LDSC) regression[76] was used to assess
372 inflation (LDSC intercept) accounting for polygenic signal. Power calculations determined
373 genotype relative risks (GRRs) providing 80 and 50 percent power given specified risk allele
374 frequencies (RAFs, from 10^{-6} to 1) and the numbers of cases and controls in our meta-analysis.
375 Fine mapping analyses included forward stepwise conditional analyses carried out in every locus
376 with GCTA-COJO using a UK Biobank subsample LD reference panel, with independent
377 associations determined using a joint P-value threshold of 1×10^{-5} and r^2 cutoff of 0.9, and
378 FINEMAP[33] Bayesian causal variant inference, using LD from available individual level data.
379 For genes whose cis regions overlapped genome-wide significant hearing loss loci,
380 coloc2[75,76] was used to assess evidence for colocalization between our hearing loss GWAS and
381 GTEx (release v8) cis-eQTL data derived from 48 tissues (URLs), using GWAS and eQTL
382 summary statistics for all common (MAF ≥ 0.01) variants within each gene's cis-region. Genes

383 with posterior probability of colocalization H4 ≥ 0.5 were determined as having evidence for
384 colocalization, and were visually inspected in eQTL+GWAS regional association plots
385 (Supplementary Figure 5).

386 Conditional analyses were performed for each cohort using REGENIE's Firth-corrected
387 logistic regression and the resulting summary statistics were meta-analyzed as described above.
388 Rare variant association analyses conditional on the common variant signal were carried out for
389 four loci with both common (MAF ≥ 0.01) and single rare variant (MAF < 0.01) genome-wide
390 significant signals, by including as covariates the dosages of variants identified in fine mapping
391 analyses. Burden analyses conditional on rare variants were carried out for five genes with
392 significant single rare variant and burden associations.

393 Assessment of heterozygous effects used association analyses excluding homozygotes as
394 well as individuals carrying pairs of exome sequenced variants (MAF < 2% and MAC > 1) that
395 were called as compound heterozygous mutations (CHMs) or potential CMHs (i.e. unknown
396 phase). CHMs were called by SHAPEIT4 phasing of merged genotype and exome data with
397 scaffolds based on inferred close relatives[77,78].

398 Heritability derived from variants partitioned into seven functional categories (coding-
399 synonymous, coding-nonsynonymous, 5-prime-UTR, 3-prime-UTR, splice site, intronic,
400 intergenic) with each category further stratified into a low frequency (0.001 < MAF ≤ 0.05) and
401 common (MAF > 0.05) minor allele frequency bin as in stratified LD score regression was
402 estimated using LD score regression (LDSC) of hearing loss association statistics on LD scores.
403 LD scores were generated from a reference panel of N = 10,000 random UK Biobank European-
404 ancestry samples' merged imputed and exome data.

405

406 **Single-Cell RNA Sequencing and Analysis**

407 All protocols were approved by the Institutional Animal Care and Use Committee in
408 accordance with the Regeneron's Institutional Animal Care and Use Committee (IACUC).
409 Cochlea and utricles from C57BL/6 mice at post-natal day 7 were micro-dissected and dissociated.
410 Suspensions of 200 cells / uL were subjected to Chromium Single Cell (10xGenomics) library
411 preparation and were sequenced on Illumina NextSeq 500. Cell Ranger Single-Cell Software Suite
412 (10X Genomics, v2.0.0) was used to perform sample de-multiplexing, alignment to MM10
413 Genome assembly with UCSC gene models, filtering, and UMI counting. PCA, UMAP and
414 clustering analyses used Seurat V3.2 (URLs)[79]. Cluster marker genes as well as canonical cell type-
415 specific genes were used to manually label the cell type for each cluster.

416

417 **URLs**

418 International Mouse Phenotyping Consortium: https://www.mousephenotype.org/.

419 gEAR portal: https://umgear.org/

420 GTEx: https://www.gtexportal.org/home/

421 [referred to in METHODS]

422 Mount Sinai Bio*Me*: https://icahn.mssm.edu/research/ipm/programs/biome-biobank/pioneering

423 FinnGen: https://www.finngen.fi/

424 University of Michigan Imputation Server: https://imputationserver.sph.umich.edu/index.html

425 TOPMed Imputation Server: https://imputation.biodatacatalyst.nhlbi.nih.gov/

426 MakeScaffold: https://github.com/odelaneau/makeScaffold

427 SHAPEIT4: https://github.com/odelaneau/shapeit4

428 GTEx v8 eQTL data: https://www.gtexportal.org/

Seurat V3.2: https://github.com/satijalab/seurat

ACKNOWLEDGEMENTS

The authors would like to thank everyone who made this work possible. In particular: the UK Biobank team, their funders, the dedicated professionals from the member institutions who contributed to and supported this work, and most especially the UK Biobank participants, without whom this research would not be possible. The exome sequencing was funded by the UK Biobank Exome Sequencing Consortium (i.e., Bristol Myers Squibb, Regeneron, Biogen, Takeda, Abbvie, Alnylam, AstraZeneca and Pfizer). This research has been conducted using the UK Biobank Resource under application number 2604. We also want to acknowledge the participants and investigators of the FinnGen study. We thank the MyCode Community Health Initiative participants for taking part in the DiscovEHR collaboration.

Author contributions: All authors contributed to critical review of the manuscript for important intellectual content, and final approval of submission of the manuscript for publication.

Conceptualization: KP, EAS, MD, BZ, AB, GC

Data curation: MNC

Analysis: KP, LD, LG, MAF, AHA, JS, AP, AM, SC, EAS

Single-cell sequencing data generation and analysis: SM, YB, JS

Mouse experiments: AK

Project administration: EC, MJ

Datasets: OM

Supervision: MAF, GA, EAS, SB, BZ, GC, AB

Writing – original draft: KP, LD, EAS, GC

Competing interests:

Regeneron authors receive salary from and own options and/or stock of the company. Decibel authors receive salary and may own options and/or stock of the company.

Data availability statement

All whole-exome sequencing, genotyping chip, and imputed sequence for UKB described in this report are publicly available to registered researchers via the UK Biobank data access protocol. Additional information about registration for access to the data is available at http://www.ukbiobank.ac.uk/register-apply/. Further information about the whole exome sequence is available at http://www.ukbiobank.ac.uk/wp-content/uploads/2019/03/Access_064-UK-Biobank-50k-Exome-Release-FAQ-v3.pdf Detailed information about the chip and imputed sequence is available at: http://www.ukbiobank.ac.uk/wp-content/uploads/2018/03/UKB-Genotyping-and-Imputation-Data-Release-FAQ-v3-2-1.pdf. Geisinger DiscovEHR, Malmo Diet and Cancer study and Mt. Sinai Biome exome sequencing and genotyping data can be made available to qualified, academic, non-commercial researchers upon request via a Data Transfer Agreement with the respective institutions. Summary statistics for FinnGen r3 can be downloaded from https://www.finngen.fi/en/access_results.

FIGURES AND TABLES

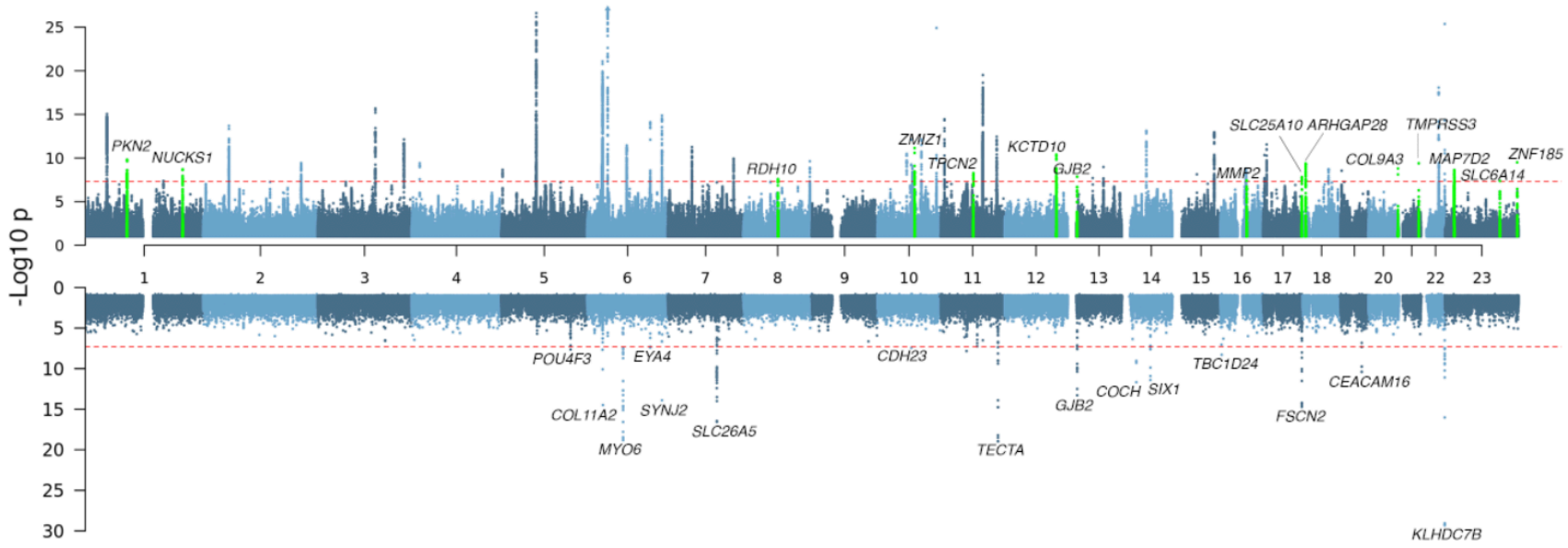


Figure 1: Common (MAF ≥ 0.01) variant (A), and rare (MAF < 0.01) coding single-variant and gene burden (B) associations with hearing loss. The gene labels refer to the nearest gene and colored in green are loci that have not previously been associated with hearing loss in the common variant results (A).

Table 1. Lead variants at the 15 novel loci associated with hearing loss. The alternate allele frequency (AAF) refers to the allele listed second in the 'Position' column. The direction of the effect in each single study (in the order: MALMO, UKB, GHS, SINAI, FinnGen) in the meta-analysis is given in the 'Direction' column where '+' indicates increased risk, '-' indicates decreased risk, and '?' indicates that the variant was not present or tested. The nSNPs column indicates the number of genome-wide significant SNPs within each locus.

Position	rsID	OR (LCI, UCI)	p-value	AAF	nSNPs	Nearest Gene	Direction
10:78760556:T:C	rs11596052	0.959 (0.947, 0.970)	7.14E-12	0.2162	94	*ZMIZ1*	+----
12:109460927:A:G	rs1558804	1.033 (1.023, 1.044)	4.11E-11	0.4468	620	*KCTD10*	-++++
1:88282464:G:A	rs475788	1.036 (1.024, 1.047)	1.56E-10	0.7145	462	*PKN2*	-++++
23:152898837:T:C	rs186256023	1.082 (1.056, 1.109)	3.14E-10	0.0315	16	*ZNF185*	+++-?
21:42388983:C:T	rs45598239	1.070 (1.048, 1.093)	4.05E-10	0.0549	18	*TMPRSS3*	+++++
18:6678716:T:A	rs8090563	1.031 (1.021, 1.042)	4.29E-10	0.4901	210	*ARHGAP28*	+++++
20:62819980:C:T	rs61734651	1.062 (1.041, 1.083)	1.60E-09	0.0654	57	*COL9A3*	+++++
1:205751355:G:A	rs823116	0.971 (0.961, 0.980)	1.99E-09	0.5493	209	*NUCKS1*	+----
23:20054286:C:T	rs7055595	1.033 (1.022, 1.044)	2.44E-09	0.2007	488	*MAP7D2*	++++?
11:69208057:T:C	rs7926098	1.030 (1.020, 1.041)	5.43E-09	0.6182	204	*TPCN2*	-++-+
13:20186225:A:T	rs117887149	1.114 (1.073, 1.156)	1.40E-08	0.0178	10	*GJB2*	+++-+
17:81711135:C:T	rs62077192	1.140 (1.089, 1.193)	1.55E-08	0.0117	92	*SLC25A10*	+++-+
8:73332501:T:C	rs4738323	1.035 (1.023, 1.048)	2.59E-08	0.1972	42	*RDH10*	-++++
16:55456852:T:A	rs17300627	0.967 (0.955, 0.978)	3.88E-08	0.2034	105	*MMP2*	+----
23:116435671:G:A	rs3788766	1.025 (1.016, 1.034)	4.35E-08	0.6386	66	*SLC6A14*	+++-?

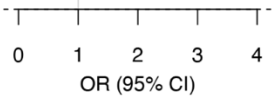
Table 2: **Nonsynonymous rare (minor allele frequency, MAF < 0.01) variants and gene burdens associated ($P < 5 \times 10^{-8}$) with hearing loss in meta-analysis**. The direction of the effect in each single study (in the order: MALMO, UKB, GHS, SINAI, FinnGen) in the meta-analysis is given in the 'Direction' column where '+' indicates increased risk, '-' indicates decreased risk and '?' indicates that the variant was not present or tested.

Gene	Top Burden/SNV	OR (LCI, UCI)	p-value	Direction
KLHDC7B	pLOF only (MAF ≤ 0.01)	2.145 (1.881, 2.446)	5.22E-30	-+++?
TECTA	pLOF + strict deleterious missense (MAF ≤ 0.01)	1.358 (1.271, 1.451)	1.10E-19	++++?
MYO6	pLOF + strict deleterious missense (MAF ≤ 0.0001)	1.565 (1.420, 1.725)	1.51E-19	+++??
FSCN2	pLOF + all missense (MAF ≤ 0.01)	1.144 (1.107, 1.183)	1.90E-15	+++-?
COL11A2	6:33189182:A:G; Phe80Ser	6.926 (4.280, 11.208)	3.24E-15	?+???
SYNJ2	6:158071628:C:T; Thr656Met	1.306 (1.221, 1.398)	1.25E-14	+++++
SLC26A5	pLOF and strict deleterious missense (MAF ≤ 0.0001)	1.956 (1.674, 2.284)	2.75E-17	-++??
COCH	pLOF + strict deleterious missense (MAF ≤ 0.0001)	1.719 (1.449, 2.039)	5.22E-10	-++??
SIX1	pLOF + strict deleterious missense (MAF ≤ 0.00001)	4.252 (2.825, 6.400)	3.95E-12	?++??
CEACAM16	pLOF + deleterious missense (MAF ≤ 0.01)	1.187 (1.128, 1.249)	4.01E-11	-++-?
GJB2	13:20189546:AC:A; Gly12fs	1.214 (1.146, 1.286)	4.23E-11	-+++?
TBC1D24	16:2497068:A:G; Asn307Ser	4.140 (2.571, 6.668)	5.10E-09	?++??
POU4F3	pLOF + all missense (MAF ≤ 0.00001)	1.923 (1.529, 2.418)	2.31E-08	?++??

CDH23	10:71712737:A:G; Asn1103Ser	1.204 (1.127, 1.287)	3.70E-08	++++?
EYA4	pLOF only (MAF ≤ 0.001)	3.077 (2.061, 4.595)	3.87E-08	+++??

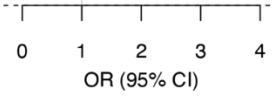
A. *GJB2 (Gly12fs)*

Study	Cases	Controls	MAF	OR (LCI I UCI)	p–value
GHS	9,296 I 193 I 4	97,381 I 2,012 I 1	0.0105	1.10 (0.94 I 1.29)	2.5E–01
UKB	104,059 I 1,760 I 15	207,696 I 2,944 I 0	0.0075	1.24 (1.16 I 1.32)	1.6E–11
SINAI	307 I 7 I 0	9,857 I 163 I 0	0.0082	1.38 (0.57 I 3.35)	4.8E–01
MALMO	497 I 3 I 0	26,707 I 386 I 0	0.0070	0.54 (0.26 I 1.14)	1.0E–01
Meta	**114,159 I 1,963 I 19**	**341,641 I 5,505 I 1**	**0.0081**	**1.21 (1.15 I 1.29)**	**4.2E–11**



B. *SLC26A5 (Leu46Pro)*

Study	Cases	Controls	MAF	OR (LCI I UCI)	p–value
GHS	8,935 I 92 I 0	95,478 I 831 I 2	0.0044	1.15 (0.91 I 1.45)	2.5E–01
UKB	104,469 I 1,359 I 6	208,576 I 2,052 I 12	0.0054	1.32 (1.23 I 1.42)	2.2E–14
SINAI	312 I 2 I 0	9,940 I 82 I 0	0.0041	0.97 (0.24 I 4.00)	9.7E–01
MALMO	493 I 7 I 0	26,761 I 349 I 2	0.0063	1.19 (0.51 I 2.79)	6.9E–01
FinnGen	9,500 I 106 I 0	121,070 I 1,253 I 3	0.0016	1.09 (0.75 I 1.58)	6.6E–01
Meta	**123,709 I 1,566 I 6**	**461,825 I 4,567 I 19**	**0.0052**	**1.30 (1.21 I 1.39)**	**3.1E–14**



C. *SLC26A5* aggregate of pLOF and strict deleterious missense

Study	Cases	Controls	MAF	OR (LCI I UCI)	p–value
GHS	9,002 I 25 I 0	96,103 I 208 I 0	0.0011	1.51 (0.92 I 2.47)	1.0E–01
UKB	98,267 I 307 I 0	195,806 I 330 I 0	0.0011	2.03 (1.73 I 2.40)	2.5E–17
MALMO	500 I 0 I 0	27,056 I 56 I 0	0.0010	NA	3.4E–01
Meta	**107,769 I 332 I 0**	**318,965 I 594 I 0**	**0.0011**	**1.96 (1.67 I 2.28)**	**2.8E–17**

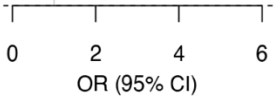
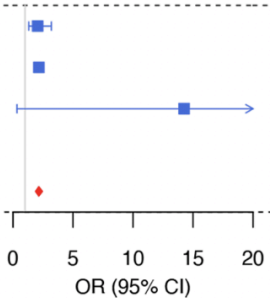


Figure 2: Association of *GJB2* Gly12fs (A), *SLC26A5* Leu46Pro (B) and *SLC26A5* pLOF and strict deleterious missense, MAF ≤ 0.0001 (C) with hearing loss.

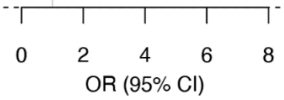
A. *KLHDC7B* aggregate of predicted loss-of-function variants (MAF ≤ 0.01)

Study	Cases	Controls	MAF	OR (LCI | UCI)	p–value
GHS	8,998 | 29 | 0	96,149 | 162 | 0	0.0009	2.05 (1.32 | 3.19)	1.5E–03
UKB	98,123 | 451 | 0	195,687 | 449 | 0	0.0015	2.15 (1.88 | 2.47)	9.3E–28
SINAI	313 | 1 | 0	10,007 | 11 | 0	0.0006	14.26 (0.33 | 619.33)	1.7E–01
MALMO	500 | 0 | 0	27,092 | 20 | 0	0.0004	NA	5.7E–01
Meta	**107,934 | 481 | 0**	**328,935 | 642 | 0**	**0.0013**	**2.14 (1.88 | 2.45)**	**5.2E–30**



B. *KLHDC7B* (Lys181fs)

Study	Cases	Controls	MAF	OR (LCI | UCI)	p–value
GHS	9,018 | 8 | 0	96,274 | 34 | 0	0.0002	2.90 (1.19 | 7.08)	1.9E–02
UKB	98,450 | 118 | 0	196,002 | 121 | 0	0.0004	2.14 (1.64 | 2.78)	1.9E–08
Meta	**107,468 | 126 | 0**	**292,276 | 155 | 0**	**0.0004**	**2.19 (1.70 | 2.82)**	**1.4E–09**



C. *KLHDC7B* (Gly302fs)

Study	Cases	Controls	MAF	OR (LCI | UCI)	p–value
GHS	9,013 | 12 | 0	96,259 | 52 | 0	0.0003	2.69 (1.32 | 5.51)	6.6E–03
UKB	98,377 | 195 | 0	195,956 | 178 | 0	0.0006	2.34 (1.89 | 2.90)	6.9E–15
MALMO	313 | 1 | 0	10,013 | 9 | 0	0.0005	49.78 (0.64 | 3,887.84)	7.9E–02
Meta	**107,703 | 208 | 0**	**302,228 | 239 | 0**	**0.0005**	**2.39 (1.94 | 2.93)**	**9.5E–17**

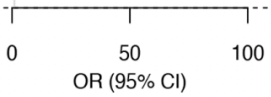
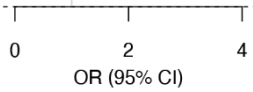


Figure 3: Association of rare variants in *KLHDC7B* with increased risk for hearing loss in meta-analysis. A. Association of an aggregate of predicted loss of function variants (MAF ≤ 0.01) in *KLHDC7B* with risk for hearing loss. **B and C.** Two variants were the predominant contributors to the *KLHDC7B* loss-of-function gene burden aggregate, Lys181fs (B) and Gly302fs (C).

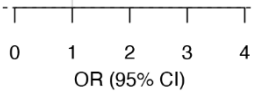
A. *FSCN2* aggregate of pLOF and missense (MAF ≤ 0.01)

Study	Cases	Controls	MAF	OR (LCI I UCI)	p–value
GHS	8,446 I 579 I 2	90,863 I 5,435 I 13	0.0287	1.15 (1.05 I 1.27)	2.6E–03
UKB	93,239 I 5,321 I 14	186,674 I 9,446 I 16	0.0252	1.14 (1.10 I 1.18)	4.8E–13
SINAI	299 I 15 I 0	9,436 I 579 I 3	0.0290	0.96 (0.57 I 1.63)	8.9E–01
MALMO	466 I 34 I 0	25,745 I 1,363 I 4	0.0254	1.43 (0.99 I 2.05)	5.7E–02
Meta	**102,450 I 5,949 I 16**	**312,718 I 16,823 I 36**	**0.0261**	**1.14 (1.11 I 1.18)**	**1.9E–15**



B. *FSCN2 (His138Tyr)*

Study	Cases	Controls	MAF	OR (LCI I UCI)	p–value
GHS	9,280 I 211 I 2	97,546 I 1,836 I 12	0.0095	1.21 (1.04 I 1.41)	1.6E–02
UKB	104,019 I 1,805 I 10	207,659 I 2,975 I 6	0.0076	1.24 (1.16 I 1.31)	1.2E–11
SINAI	179 I 3 I 0	4,832 I 109 I 1	0.0113	1.00 (0.27 I 3.65)	1.0E+00
MALMO	482 I 18 I 0	26,653 I 458 I 1	0.0087	2.22 (1.28 I 3.86)	4.8E–03
FinnGen	9,432 I 175 I 0	120,516 I 1,804 I 6	0.0082	1.24 (1.05 I 1.47)	1.2E–02
Meta	**123,392 I 2,212 I 12**	**457,206 I 7,182 I 26**	**0.0080**	**1.24 (1.18 I 1.31)**	**4.1E–15**



C. *SYNJ2 (Thr656Met)*

Study	Cases	Controls	MAF	OR (LCI I UCI)	p–value
GHS	9,362 I 130 I 1	98,229 I 1,162 I 3	0.0061	1.19 (0.97 I 1.46)	9.5E–02
UKB	97,319 I 1,238 I 17	194,164 I 1,969 I 3	0.0055	1.31 (1.22 I 1.41)	4.8E–13
SINAI	308 I 6 I 0	9,947 I 75 I 0	0.0039	2.66 (1.00 I 7.04)	4.9E–02
MALMO	495 I 5 I 0	26,902 I 210 I 0	0.0039	1.26 (0.47 I 3.38)	6.5E–01
FinnGen	9,524 I 83 I 0	121,546 I 779 I 1	0.0022	1.36 (0.97 I 1.92)	7.3E–02
Meta	**117,008 I 1,462 I 18**	**450,788 I 4,195 I 7**	**0.0050**	**1.31 (1.22 I 1.40)**	**1.3E–14**

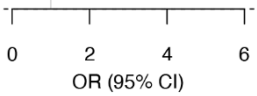


Figure 4 Association of human hearing loss with genes previously implicated in hearing loss in mice. A. pLOF and missense (MAF ≤ 0.01) burden association in *FSCN2*. **B**. The His138Tyr variant is the major contributor to the burden. **C.** *SYNJ2* (Thr656Met) association with increased risk for hearing loss.

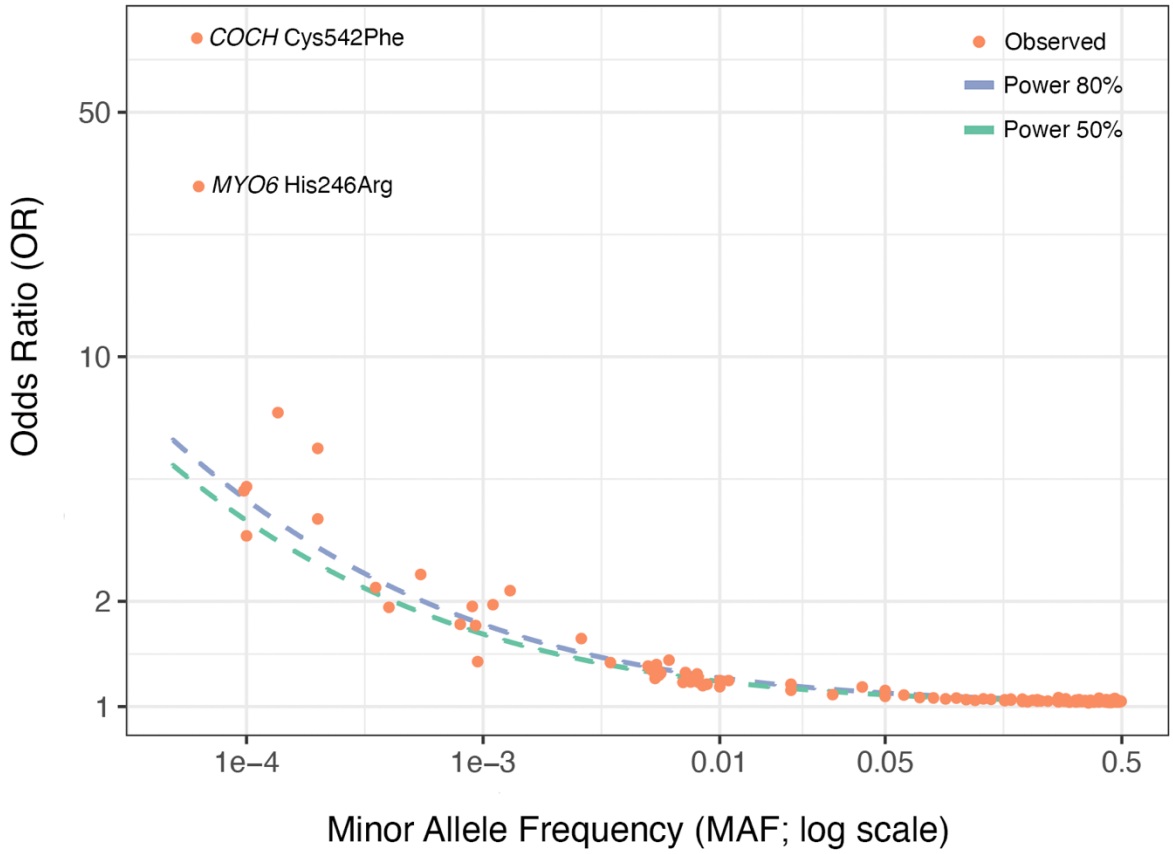


Figure 5: Effect size and allele frequency for variants associated with adult hearing loss. Plotted are odds ratio estimates (on log scale) and minor allele frequencies of genome-wide significant variants and gene burdens (Supplementary Tables 2, 6 & 7). 50% and 80% power curves for the present study are plotted as dotted lines. Notably, almost all points (particularly low-frequency and common variants, MAF > 0.0005) lie very close to the dotted power curves.

REFERENCES

1. Cunningham, L. L. & Tucci, D. L. Hearing Loss in Adults. *N. Engl. J. Med.* **377**, 2465–2473 (2017).
2. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy & Committee on Accessible and Affordable Hearing Health Care for Adults. *Hearing Health Care for Adults: Priorities for Improving Access and Affordability*. (National Academies Press, 2016).
3. Livingston, G. *et al*. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* **396**, 413–446 (2020).
4. Morton, C. C. & Nance, W. E. Newborn hearing screening--a silent revolution. *N. Engl. J. Med.* **354**, 2151–2164 (2006).
5. Korver, A. M. H. *et al*. Congenital hearing loss. *Nat Rev Dis Primers* **3**, 16094 (2017).
6. Liberman, M. C. Noise-induced and age-related hearing loss: new perspectives and potential therapies. *F1000Res*. **6**, 927 (2017).
7. Wu, P.-Z., O'Malley, J. T., de Gruttola, V. & Liberman, M. C. Age-related hearing loss is dominated by damage to inner ear sensory cells, not the cellular battery that powers them. *J. Neurosci*. **40**, 6357–6366 (2020).
8. Karlsson, K. K., Harris, J. R. & Svartengren, M. Description and primary results from an audiometric study of male twins. *Ear Hear*. **18**, 114–120 (1997).
9. Christensen, K., Frederiksen, H. & Hoffman, H. J. Genetic and environmental influences on self-reported reduced hearing in the old and oldest old. *J. Am. Geriatr. Soc*. **49**, 1512–1517 (2001).
10. Kvestad, E., Czajkowski, N., Krog, N. H., Engdahl, B. & Tambs, K. Heritability of hearing loss. *Epidemiology* **23**, 328–331 (2012).

11. Hendrickx, J.-J. *et al*. Familial aggregation of pure tone hearing thresholds in an aging European population. *Otol. Neurotol*. **34**, 838–844 (2013).

12. Viljanen, A. *et al*. Genetic and environmental influences on hearing in older women. *J. Gerontol. A Biol. Sci. Med. Sci*. **62**, 447–452 (2007).

13. Friedman, R. A. *et al*. GRM7 variants confer susceptibility to age-related hearing impairment. *Hum. Mol. Genet*. **18**, 785–796 (2009).

14. Hoffmann, T. J. *et al*. A Large Genome-Wide Association Study of Age-Related Hearing Impairment Using Electronic Health Records. *PLoS Genet*. **12**, e1006371 (2016).

15. Nagtegaal, A. P. *et al*. Genome-wide association meta-analysis identifies five novel loci for age-related hearing impairment. *Sci. Rep*. **9**, 15192 (2019).

16. Fransen, E. *et al*. Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. *Eur. J. Hum. Genet*. **23**, 110–115 (2015).

17. Van Laer, L. *et al*. A genome-wide association study for age-related hearing impairment in the Saami. *Eur. J. Hum. Genet*. **18**, 685–693 (2010).

18. Ivarsdottir, E. V. *et al*. The genetic architecture of age-related hearing impairment revealed by genome-wide association analysis. *Commun Biol* **4**, 706 (2021).

19. Lewis, M. A. *et al*. Whole exome sequencing in adult-onset hearing loss reveals a high load of predicted pathogenic variants in known deafness-associated genes and identifies new candidate genes. *BMC Med. Genomics* **11**, 77 (2018).

20. Tong, X. *et al*. Kctd10 regulates heart morphogenesis by repressing the transcriptional activity of Tbx5a in zebrafish. *Nat. Commun*. **5**, 3153 (2014).

21. Ren, K. *et al*. KCTD10 is involved in the cardiovascular system and Notch signaling during early embryonic development. *PLoS One* **9**, e112275 (2014).

22. Pan, X. *et al*. MAP7D2 Localizes to the Proximal Axon and Locally Promotes Kinesin-1-Mediated Cargo Transport into the Axon. *Cell Rep*. **26**, 1988-1999.e6 (2019).

23. Hanson-Kahn, A. *et al*. Autosomal recessive Stickler syndrome resulting from a COL9A3 mutation. *Am. J. Med. Genet. A* **176**, 2887–2891 (2018).

24. Asamura, K., Abe, S., Fukuoka, H., Nakamura, Y. & Usami, S.-I. Mutation analysis of COL9A3, a gene highly expressed in the cochlea, in hearing loss patients. *Auris Nasus Larynx* **32**, 113–117 (2005).

25. Wonkam, A., Manyisa, N., Bope, C. D., Dandara, C. & Chimusa, E. R. Whole Exome Sequencing Reveals Pathogenic Variants in MYO3A, MYO15A and COL9A3, and Differential Frequencies in Ancestral Alleles in Hearing Impairment Genes Among Individuals from Cameroon. *Hum. Mol. Genet*. (2020) doi:10.1093/hmg/ddaa225.

26. Dickinson, M. E. *et al*. High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).

27. Guipponi, M. *et al*. An integrated genetic and functional analysis of the role of type II transmembrane serine proteases (TMPRSSs) in hearing loss. *Hum. Mutat*. **29**, 130–141 (2008).

28. Fasquelle, L. *et al*. Tmprss3, a transmembrane serine protease deficient in human DFNB8/10 deafness, is critical for cochlear hair cell survival at the onset of hearing. *J. Biol. Chem*. **286**, 17383–17397 (2011).

29. Tang, P.-C. *et al*. Defective Tmprss3-associated hair cell degeneration in inner ear organoids. *Stem Cell Reports* **13**, 147–162 (2019).

30. Scott, H. S. *et al*. Insertion of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. *Nat. Genet*. **27**, 59–63 (2001).

31. Yang, J. *et al*. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet*. **44**, 369–75, S1-3 (2012).

32. Benner, C. *et al*. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).

33. Dobbyn, A. *et al*. Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am. J. Hum. Genet*. **102**, 1169–1184 (2018).

34. Shearer, A. E., Hildebrand, M. S. & Smith, R. J. H. Hereditary Hearing Loss and Deafness Overview. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1999).

35. Mohiddin, S. A. *et al*. Novel association of hypertrophic cardiomyopathy, sensorineural deafness, and a mutation in unconventional myosin VI (MYO6). *J. Med. Genet*. **41**, 309–314 (2004).

36. Street, V. A. *et al*. A novel DFNA9 mutation in the vWFA2 domain of COCH alters a conserved cysteine residue and intrachain disulfide bond formation resulting in progressive hearing loss and site-specific vestibular and central oculomotor dysfunction. *Am. J. Med. Genet. A* **139A**, 86–95 (2005).

37. Parzefall, T. *et al*. A Novel Variant in the TBC1D24 Lipid-Binding Pocket Causes Autosomal Dominant Hearing Loss: Evidence for a Genotype-Phenotype Correlation. *Front. Cell. Neurosci*. **14**, 585669 (2020).

38. Chakchouk, I. *et al*. Novel mutations confirm that COL11A2 is responsible for autosomal recessive non-syndromic hearing loss DFNB53. *Mol. Genet. Genomics* **290**, 1327–1334 (2015).

39. Hashimoto, Y., Kim, D. J. & Adams, J. C. The roles of fascins in health and disease. *J. Pathol.* **224**, 289–300 (2011).

40. Perrin, B. J. *et al*. β-Actin and fascin-2 cooperate to maintain stereocilia length. *J. Neurosci.* **33**, 8114–8121 (2013).

41. Liu, X. *et al*. Null Mutation of the Fascin2 Gene by TALEN Leading to Progressive Hearing Loss and Retinal Degeneration in C57BL/6J Mice. *G3* **8**, 3221–3230 (2018).

42. Rusk, N. *et al*. Synaptojanin 2 functions at an early step of clathrin-mediated endocytosis. *Curr. Biol.* **13**, 659–663 (2003).

43. Manji, S. S. M. *et al*. A mutation in synaptojanin 2 causes progressive hearing loss in the ENU-mutagenised mouse strain Mozart. *PLoS One* **6**, e17607 (2011).

44. Martelletti, E. *et al*. Synaptojanin2 mutation causes progressive high-frequency hearing loss in mice. *Front. Cell. Neurosci.* **14**, 561857 (2020).

45. Finucane, H. K. *et al*. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

46. Gazal, S. *et al*. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet.* **50**, 1600–1607 (2018).

47. Gong, R. *et al*. Hearing loss prevalence and risk factors among older adults in China. *Int. J. Audiol.* **57**, 354–359 (2018).

48. Du, T.-T. *et al*. LMO7 deficiency reveals the significance of the cuticular plate for hearing function. *Nat. Commun*. **10**, 1117 (2019).

49. Liu, Y. *et al*. Critical role of spectrin in hearing development and deafness. *Sci Adv* **5**, eaav7803 (2019).

50. Lee, J. R. & White, T. W. Connexin-26 mutations in deafness and skin disease. *Expert Rev. Mol. Med*. **11**, e35 (2009).

51. Kelsell, D. P. *et al*. Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature* **387**, 80–83 (1997).

52. Kelley, P. M. *et al*. Novel mutations in the connexin 26 gene (GJB2) that cause autosomal recessive (DFNB1) hearing loss. *Am. J. Hum. Genet*. **62**, 792–799 (1998).

53. Rabionet, R. *et al*. Molecular basis of childhood deafness resulting from mutations in the GJB2 (connexin 26) gene. *Hum. Genet*. **106**, 40–44 (2000).

54. Engel-Yeger, B. *et al*. The effects of a connexin 26 mutation--35delG--on oto-acoustic emissions and brainstem evoked potentials: homozygotes and carriers. *Hear. Res*. **163**, 93–100 (2002).

55. Groh, D. *et al*. Hearing function in heterozygous carriers of a pathogenic GJB2 gene mutation. *Physiol. Res*. **62**, 323–330 (2013).

56. Mutai, H. *et al*. Diverse spectrum of rare deafness genes underlies early-childhood hearing loss in Japanese patients: a cross-sectional, multi-center next-generation sequencing study. *Orphanet J. Rare Dis*. **8**, 172 (2013).

57. Liberman, M. C. *et al*. Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier. *Nature* **419**, 300–304 (2002).

58. Wells, H. R. R. *et al*. GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. *Am. J. Hum. Genet*. **105**, 788–802 (2019).

59. Jeong, G. *et al*. A Kelch domain-containing KLHDC7B and a long non-coding RNA ST8SIA6-AS1 act oppositely on breast cancer cell proliferation via the interferon signaling pathway. *Sci. Rep*. **8**, 12922 (2018).

60. Martín-Pardillos, A. & Cajal, S. R. Y. Characterization of Kelch domain-containing protein 7B in breast tumours and breast cancer cell lines. *Oncol. Lett*. **18**, 2853–2860 (2019).

61. Rudnicki, A. *et al*. Next-generation sequencing of small RNAs from inner ear sensory epithelium identifies microRNAs and defines regulatory pathways. *BMC Genomics* **15**, 484 (2014).

62. Liu, H. *et al*. Cell-specific transcriptome analysis shows that adult pillar and Deiters' cells express genes encoding machinery for specializations of cochlear hair cells. *Front. Mol. Neurosci*. **11**, 356 (2018).

63. GTEx Consortium *et al*. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

64. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet*. **51**, 187–195 (2019).

65. Sudlow, C. *et al*. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. **12**, e1001779 (2015).

66. Carey, D. J. *et al*. The Geisinger MyCode community health initiative: an electronic health record–linked biobank for precision medicine research. *Genet. Med*. **18**, 906–913 (2016).

67. Berglund, G., Elmstähl, S., Janzon, L. & Larsson, S. A. The Malmo Diet and Cancer Study. Design and feasibility. *J. Intern. Med.* **233**, 45–51 (1993).

68. Dewey, F. E. *et al*. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**, aaf6814 (2016).

69. Van Hout, C. V. *et al*. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).

70. Mbatchou, J. *et al*. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* (2021) doi:10.1038/s41588-021-00870-7.

71. Girotto, G. *et al*. Hearing function and thresholds: a genome-wide association study in European isolated populations identifies new loci and pathways. *J. Med. Genet.* **48**, 369–374 (2011).

72. Nolan, L. S. *et al*. Estrogen-related receptor gamma and hearing function: evidence of a role in humans and mice. *Neurobiol. Aging* **34**, 2077.e1–9 (2013).

73. Wolber, L. E. *et al*. Salt-inducible kinase 3, SIK3, is a new gene associated with hearing. *Hum. Mol. Genet.* **23**, 6407–6418 (2014).

74. Vuckovic, D. *et al*. Genome-wide association analysis on normal hearing function identifies PCDH20 and SLC28A3 as candidates for hearing function and loss. *Hum. Mol. Genet.* **24**, 5655–5664 (2015).

75. Kalra, G. *et al*. Biological insights from multi-omic analysis of 31 genomic risk loci for adult hearing difficulty. *PLoS Genet.* e1009025 (2019) doi:10.1371/journal.pgen.1009025.

76. Bulik-Sullivan, B. K. *et al*. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

77. Staples, J. *et al*. PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am. J. Hum. Genet*. **95**, 553–564 (2014).

78. Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T. Accurate, scalable and integrative haplotype estimation. *Nat. Commun*. **10**, 5436 (2019).

79. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol*. **36**, 411–420 (2018).