

1 **Genome-wide Screen of Otosclerosis in** 2 **Population Biobanks: 18 Loci and Shared** 3 **Heritability with Skeletal Structure**

4 **Joel T. Rämö¹, Tuomo Kiiskinen¹, Juha Karjalainen^{1,2,3,4}, Kristi Krebs⁵, Mitja Kurki^{1,2,3,4}, Aki S.**
5 **Havulinna⁶, Eija Hämäläinen¹, Paavo Häppölä¹, Heidi Hautakangas¹, FinnGen, Konrad J.**
6 **Karczewski^{1,2,3,4}, Masahiro Kanai^{1,2,3,4}, Reedik Mägi⁵, Priit Palta^{1,5}, Tõnu Esko⁵, Andres Metspalu⁵,**
7 **Matti Pirinen^{1,7,8}, Samuli Ripatti^{1,2,7}, Lili Milani⁵, Antti Mäkitie⁹, Mark J. Daly^{1,2,3,4,10}, and Aarno**
8 **Palotie^{1,2,3,4}**

9 1. Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science (HiLIFE), University of
10 Helsinki, Helsinki, Finland

11 2. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge,
12 Massachusetts, USA

13 3. Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts,
14 USA

15 4. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

16 5. Estonian Genome Center, University of Tartu, Tartu, Estonia, Institute of Molecular and Cell Biology,
17 University of Tartu, Tartu, Estonia

18 6. Finnish Institute for Health and Welfare, Helsinki, Finland

19 7. Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland

20 8. Department of Mathematics and Statistics, Faculty of Science, University of Helsinki, Helsinki, Finland

21 9. Department of Otorhinolaryngology-Head and Neck Surgery, University of Helsinki and HUS Helsinki
22 University Hospital, Helsinki, Finland

23 10. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

24

25 Corresponding author: Aarno Palotie, MD PhD, University of Helsinki, Tukholmankatu 8, FI-00290
26 Helsinki, Finland. Email aarno.palotie@helsinki.fi

27

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

28 **Abstract**

29
30 Otosclerosis is one of the most common causes of conductive hearing loss, affecting 0.3% of the
31 population. It typically presents in adulthood and half of the patients have a positive family history.
32 The pathophysiology of otosclerosis is poorly understood and treatment options are limited. A
33 previous genome-wide association study (GWAS) identified a single association locus in an intronic
34 region of *RELN*. Here, we report a meta-analysis of GWAS studies of otosclerosis in three population-
35 based biobanks comprising 2,413 cases and 762,382 controls. We identify 15 novel risk loci ($p <$
36 5×10^{-8}) and replicate the regions of *RELN* and two previously reported candidate genes (*TGFB1* and
37 *MEPE*). Implicated genes in many loci are essential for bone remodelling or mineralization.
38 Otosclerosis is genetically correlated with height and fracture risk, and the association loci overlap
39 with severe skeletal disorders. Our results highlight TGF β 1 signalling for follow-up mechanistic
40 studies.

41 Introduction

42 Otosclerosis is one of the most common causes of conductive hearing loss. It is a heritable disorder
43 typically characterized by disordered bone growth in the middle ear.^{1,2} In classic otosclerosis, due
44 to fixation of the footplate of the stapes bone, the conduction of sound through the ossicular chain
45 to the inner ear is interrupted. The pathogenesis of otosclerosis is poorly understood, and
46 therapeutic options are mostly limited to hearing aids, prosthesis surgery and cochlear implants.³

47 Clinical otosclerosis has an estimated prevalence of 0.30% to 0.38% in Caucasian populations.⁴
48 Histologic otosclerosis without clinical symptoms is more frequent, with bony overgrowth observed
49 in as many as 2.5% of temporal bone autopsy specimens.⁵ Symptomatic otosclerosis most
50 frequently occurs in working-age individuals between the second and fifth decades.^{6,7} Initial
51 manifestation is often limited to one ear, but eventual bilateral disease is observed in 70-80% of
52 cases.^{6,7}

53 Otosclerosis is highly familial, with a positive family history reported for 50-60% of cases.^{8,9} Based
54 on segregation patterns, early studies classified otosclerosis as an autosomal dominant disease with
55 reduced penetrance.⁹⁻¹¹ However, efforts to identify causative genes have produced inconsistent
56 results, with insufficient evidence for most candidate genes.^{12,13} A genome-wide association study
57 (GWAS) including a total of 1,149 otosclerosis patients identified intronic risk variants within the
58 gene encoding Reelin (*RELN*) in European populations.^{14,15} However, the potential biological
59 function of Reelin and the role of other genes remain unknown.

60 Here, we report to our knowledge the largest GWAS of otosclerosis including a total of 2,413 cases
61 and 762,382 controls from three population-based sample collections: the Finnish FinnGen study,
62 the Estonian Biobank (EstBB), and the UK Biobank (UKBB). We identify 15 novel GWAS loci, and
63 replicate associations in the regions of *RELN* and two previously reported candidate genes, *TGFB1*
64 and *MEPE*.¹⁴⁻¹⁹ Several discovered loci harbor genes involved in the regulation of osteoblast or
65 osteoclast function or biomineralization. We also demonstrate that otosclerosis is genetically
66 correlated with common and rare skeletal traits.

67 Results

68 Identification of otosclerosis cases based on electronic health records

69 Based on International Classification of Diseases (ICD) diagnosis codes (versions 8, 9 and 10), we
70 identified a total of 2,413 otosclerosis cases in the three biobanks, including 1,350 cases in FinnGen,
71 713 in EstBB and 353 in UKBB, for cohort prevalences of 0.64%, 0.52%, and 0.08%, respectively
72 (Table 1). A total of 762,382 individuals had no ICD-based diagnosis of otosclerosis and were
73 assigned control status. A proportion of otosclerosis patients undergo stapes procedures, which
74 reflect disease severity and the validity of the ICD-based diagnoses. Stapes procedures were
75 registered for 544 (40.3%) otosclerosis cases in FinnGen and for 161 (22.6%) otosclerosis cases in
76 EstBB. Among all 1,350 individuals diagnosed with otosclerosis in FinnGen, five (0.37%) had a
77 diagnosis of osteogenesis imperfecta (representing 12% of all 41 individuals with osteogenesis
78 imperfecta), and none had a diagnosis of osteopetrosis.

79 Genomic loci associated with otosclerosis in each cohort

80 In case-control GWASs within the individual cohorts, we observed ten loci associated with
81 otosclerosis at genome-wide significance ($p < 5 * 10^{-8}$) in FinnGen, four in UKBB and one in EstBB
82 (Supplementary Data 1; Supplementary Data 2). Two loci were associated in both FinnGen and
83 UKBB: A chromosome 11 locus tagged by the lead variants rs11601767 and rs144458122,
84 respectively, and a chromosome 17 locus upstream of *FAM20A* tagged by rs11868207 and
85 rs11652821. Among all FinnGen lead variants, six were associated with otosclerosis at nominal
86 significance ($p < 0.05$) in UKBB and five were associated with otosclerosis in EstBB. All three UKBB
87 lead variants were nominally significant in FinnGen. The only genome-wide significant variant in
88 EstBB (rs78942990, $p = 4.2 * 10^{-8}$, MAF = 2.0%) was not associated with otosclerosis in either
89 FinnGen or UKBB at nominal significance level.

90 We calculated pairwise genetic correlation r_g using summary statistics from the different cohorts
91 with the cross-trait LD Score Regression.^{20,21} R_g between FinnGen and UKBB was 1.04, indicating
92 similar genetic effects between the two cohorts.²² The calculation of r_g was not feasible for EstBB
93 due to a negative heritability estimate (observed scale $h^2 = -0.0063$ [S.E. 0.003]), likely reflecting
94 modest polygenic signal in the cohort. However, as the effect estimates for lead variants were

95 largely concordant between the three cohorts (Supplementary Data 2), we proceeded with a fixed-
96 effects meta-analysis.

97 **18 significant loci in the meta-analysis of summary statistics**

98 In the meta-analysis of all three cohorts (including a total of 2,413 cases and 762,382 controls) we
99 identified 1,257 variants associated with otosclerosis at the genome-wide significance level ($p <$
100 5×10^{-8}). The genomic inflation factor calculated from the meta-analysis p -values was not
101 significantly elevated ($\lambda_{GC}=1.03$). The univariate LD Score regression intercept was closer to 1.0 at
102 1.016 (S.E. 0.009), indicating that part of the elevation in λ_{GC} reflected true additive polygenic effects
103 and not spurious associations from population stratification or cryptic relatedness.²¹

104 The significant variants in the meta-analysis clustered in a total of 18 loci (regions at least 500kb
105 apart) (Figure 1; Table 2; Supplementary Data 3; Supplementary Data 4), including the previously
106 reported *RELN* locus and 17 loci not previously reported in a GWAS study of otosclerosis.¹⁴ All lead
107 variants in the 18 meta-analysis loci were common with a cross-cohort effect allele frequency (EAF)
108 of at least 16% (Table 2). In two loci the closest genes (*MEPE* and *TGFB1*) have been implicated in
109 candidate gene studies; to our knowledge, the remaining 15 loci have not been characterized in
110 association with otosclerosis.¹⁵⁻¹⁹

111 **Characterization of the potential susceptibility loci for otosclerosis**

112 *Coding variation*

113 The lead variant in the chromosome 3 locus, rs4917 T>C, is a missense variant in the exon 6 of *AHSG*
114 (EAF = 64%, OR = 0.83 [0.78-0.88], $p = 2.3 \times 10^{-9}$). The variant results in a methionine to threonine
115 conversion and is predicted to be tolerated/benign by the SIFT and PolyPhen algorithms with a
116 scaled CADD score of 9.9. However, it has been reported as a splicing quantitative trait locus (sQTL)
117 for *AHSG* in the liver in the GTEx v8 database ($p = 1.1 \times 10^{-7}$).²³ Among all 1,257 variants which
118 reached genome-wide significance in the meta-analysis, we identified no high-impact variants.
119 Seven missense variants were significantly associated with otosclerosis (Supplementary Data 5):
120 rs1193851 C>G in *PCNX3*, rs11544832 in *YIF1A*, rs7990383 G>A in *COL4A2*, rs7990383 C>A in *IRX5*,
121 rs4918 G>C in *AHSG*, rs2229642 C>G in *ITPR3*, and rs4713668 in *IP6K3*. Rs4713668 (EAF = 0.48, OR
122 = 1.19 [1.12-1.26], $p = 8.8 \times 10^{-9}$) is classified as deleterious and benign by the SIFT and PolyPhen

123 algorithms, respectively, with a scaled CADD score of 19.9. All other missense variants are classified
124 as tolerated or benign.

125 In the chromosome 4 locus near the *MEPE* gene, the rare frameshift variant rs753138805 (EAF 0.3%
126 in controls and 1.3% in cases) was the lead variant in the GWAS in FinnGen, but was not included in
127 the meta-analysis due to poor imputation quality in EstBB (Imputation information score 0.28) and
128 non-inclusion in UKBB genotype data. The rs753138805 variant was strongly associated with
129 otosclerosis in FinnGen (OR = 18.9 [95% CI 8.05-44.4]; $p = 1.5 \times 10^{-11}$). The variant is 2.4-fold enriched
130 in Finnish compared with non-Finnish Europeans based on sequence data in the Genome
131 Aggregation Database.²⁴ We validated the imputation of the variant within the Migraine Family
132 subcohort of FinnGen. Among 65 individuals determined as heterozygous for rs753138805 by
133 imputation, we confirmed the genotype by Sanger sequencing in 56 individuals (86%) and in all five
134 predicted carriers of the reference allele examined as controls.

135 Finally, based on linkage disequilibrium in the FinnGen cohort, we examined 988 variants (including
136 those with AF < 0.001) in high LD ($r > 0.6$) with the lead variants from the meta-analysis, but
137 observed no additional protein-altering variation.

138 *Intronic and intergenic variation*

139 The lead variants were intronic in seven loci. In the locus on chromosome 7, the strongest
140 association was observed for variants in the second and third introns of *RELN* concordant with
141 previous GWA studies (Supplementary Data 4).^{14,15} In the locus on chromosome 19, the association
142 signal overlapped with several genes including *TGFB1*, *AXL*, *HNRNPUL1* and *CCDC97*, with the most
143 significant association for the intronic *TGFB1* variant rs8105161 (EAF = 16%, OR = 0.8 [0.74-0.86], p
144 = 1.76×10^{-8}). In chromosome 20, the associated variants were located in the first intron of *EYA2*. In
145 two of the significant loci, the associated variants were tightly clustered in the region of antisense
146 genes: *COL4A2-AS2* and an intronic region of *COL4A2* in chromosome 13, and *STX-17-AS1* in
147 chromosome 19.

148 The two variants most significantly associated with otosclerosis on chromosome 14, rs1951391
149 (EAF = 33%, OR = 1.38 [1.28-1.49], $p = 7.0 \times 10^{-17}$) and rs62007683 G>T (EAF = 33%, OR = 1.37 [1.27-
150 1.47], $p = 3.2 \times 10^{-16}$), were both located in the same intronic region of *MARK3*. In the association

151 locus on chromosome 6, an apparent association haplotype block spanned the entire *SUPT3H* gene
152 and the first exon and first intron of the *RUNX2* gene; the lead variant rs13192457 (EAF = 44%, OR
153 = 1.25 [1.17-1.34], $p = 2.3 \times 10^{-10}$) is located in an intronic region of *SUPT3H*.

154 **Fine-mapping of the association loci**

155 To identify the most likely causal variants in the otosclerosis association loci, we performed a fine-
156 mapping analysis in FinnGen using the FINEMAP software (v1.4).²⁵ We analyzed ten loci that reached
157 genome-wide significance in the meta-analysis and had suggestive evidence of association in
158 FinnGen ($p < 1 \times 10^{-6}$). The number of variants in the resulting credible sets ranged from three (for
159 the chromosome 4 locus near *MEPE*) to 4,604 reflecting high uncertainty (Supplementary Data 6).
160 In the chromosome 4 locus, the *MEPE* frameshift variant rs753138805 was the most likely causal
161 variant (probability = 66%). In the chromosome 17 locus upstream of *FAM20A*, the credible set
162 included nine variants, with the most likely causal variants being rs11868207 (probability = 52%, AF
163 = 23.6%, OR = 1.39 [1.29-1.49], $p = 2.1 \times 10^{-18}$) and rs8070086 (probability = 32%, AF = 23.6%, OR =
164 1.39 [1.29-1.49], $p = 1.8 \times 10^{-18}$).

165 **Gene and gene set analysis**

166 In a gene-based analysis using MAGMA, a total of 62 genes were significantly associated with
167 otosclerosis ($p < 2.6 \times 10^{-6}$) (Supplementary Data 7). The gene-based associations were significantly
168 enriched ($p < 4.9 \times 10^{-6}$) for the following gene sets based on gene ontology terms: COLLAGEN TYPE IV
169 TRIMER ($p = 4.1 \times 10^{-7}$), BASEMENT MEMBRANE COLLAGEN TRIMER ($p = 3.7 \times 10^{-7}$), TRANSFORMING GROWTH
170 FACTOR BETA RECEPTOR ACTIVITY TYPE I ($p = 1.7 \times 10^{-6}$), POSITIVE REGULATION OF CHOLESTEROL METABOLIC
171 PROCESS ($p = 7.9 \times 10^{-7}$), and BIOMINERALIZATION ($p = 2.2 \times 10^{-6}$).

172 Following the identification of a genome-wide significant association signal overlapping with the
173 *TGFB1* gene and gene-set association with type I TGF β receptor activity, we examined other genes
174 involved in TGF β 1 signalling for subthreshold association signals (Supplementary Data 8).²⁶ We
175 observed a trend towards association in the *TGFBR1* locus but not the *TGFBR2* locus. The lead
176 variant in the *TGFBR1* locus, rs11386616 C>CG (EAF = 24%, OR = 1.18 [1.1-1.26], $p = 2.0 \times 10^{-6}$) is eQTL
177 for *TGFBR1* in whole blood in the GTEx database (NES = -0.12, $p = 2.3 \times 10^{-13}$). The *TGFBR1* gene was
178 also significantly associated with otosclerosis in the gene-based analysis using MAGMA ($p = 1.8 \times 10^{-6}$).
179 Within the regions coding for the SMAD proteins that function as the main signal transducers of

180 the TGF β superfamily, we observed modest and nonsignificant association signals near *SMAD3*
181 (intronic lead variant rs7163381 A>G [AF= 75%, OR = 0.8 [0.80-0.86], $p = 1.0 \cdot 10^{-5}$]) and in the region
182 near *TGFBI* (coding for transforming growth factor-beta-induced) and *SMAD5* (lead variant:
183 rs17169386 T>C; EAF = 0.012, OR = 2.05 [1.43-2.95], $p = 1.3 \cdot 10^{-4}$).

184 **Associations with gene expression and splicing**

185 We examined the association of lead variants with tissue-specific gene expression in the GTEx v8
186 database.²³ Among 18 lead variants, 11 variants were eQTL in at least one tissue at the 5% FDR cut-
187 off provided by the GTEx project (Supplementary Data 11).²³ The 11 variants were eQTL for a total
188 of 53 genes in 43 tissues and in addition, seven lead variants were sQTL for a total of 22 genes in 42
189 tissues (Supplementary Data 12). Eleven genes (*TGFBI*, *CKB*, *RUNX2*, *AHSG*, *TPBG*, *CLCN7*, *SPP1*,
190 *PKD2*, *NEAT1*, *LTBP3*, and *ITPR3*) have been implicated in bone metabolism or skeletal disorders,
191 and are described in more detail in the Discussion section.

192 The eQTL and sQTL signals suggest a regulatory role for many of the identified otosclerosis
193 susceptibility variants. In several loci, the lead variant was associated with the expression or splicing
194 of multiple genes. In the chromosome 14 locus characterized by an association haplotype block
195 spanning *MARK3* and *CKB*, the lead variant rs1951391 T>C (EAF = 33%, OR = 1.38 [1.28-1.49], $p =$
196 $7.0 \cdot 10^{-17}$) was associated with the expression of both *MARK3* (eQTL for 20 tissues and sQTL for 4
197 tissues) and *CKB* (eQTL for 7 tissues and sQTL for 8 tissues); the variant was also eQTL for nine and
198 sQTL for five other genes in close proximity. In the chromosome 6 locus with an association block
199 overlapping *SUPT3H* and part of *RUNX2*, the lead variant rs13192457 (EAF = 44%, OR = 1.25 [1.17-
200 1.34], $p = 2.3 \cdot 10^{-10}$) was eQTL for *SUPT3H* in several tissues (e.g. NES = 0.2 in cultured fibroblasts; p
201 = $9.1 \cdot 10^{-8}$) and an sQTL for *RUNX2* in testes (NES = -0.47, $p = 7.5 \cdot 10^{-15}$). In the chromosome 6 locus
202 near *IP6K3*, the lead variant rs9469583 (EAF = 58%, OR = 1.22 [1.15-1.3], $p = 2.1 \cdot 10^{-10}$) was most
203 strongly associated with the expression (normalized expression score [NES] = -0.7, $p = 8.2 \cdot 10^{-30}$ in
204 pancreas tissue) and splicing (NES = 0.86 in skeletal muscle, $p = 1.1 \cdot 10^{-72}$) of *IP6K3*; however, the
205 variant was also eQTL and sQTL for several other genes, including *ITPR3*, *UQCC2* and *BAK1*.

206 We also sought to identify variants simultaneously associated with reduced susceptibility for
207 otosclerosis and reduced gene expression, as such genes could represent targets for therapeutic
208 inhibition. The intronic rs8105161 lead variant in *TGFBI* was associated with protection from

209 otosclerosis (OR = 0.8 [0.74-0.86], $p = 1.8 \times 10^{-8}$) and reduced expression of *TGFB1* in adrenal glands
210 (NES = -0.76, $p = 1.3 \times 10^{-19}$) and cultured fibroblasts (NES = -0.097, $p = 1.1 \times 10^{-4}$). In the chromosome
211 13 locus, in which the association signal clustered in the region of *COL4A2-AS2* (*COL4A2* antisense
212 2) within the *COL4A2* gene, the lead variant rs9559805 (OR = 0.82 [0.77-0.87] for otosclerosis, $p =$
213 3.4×10^{-11}) was associated with reduced expression of both *COL4A2-AS2* (NES = -0.17 in the
214 muscularis layer of the esophagus, $p = 3.5 \times 10^{-6}$) and *COLA42* (NES = -0.16, $p = 8.3 \times 10^{-5}$ in
215 subcutaneous adipose tissue).

216 Replication of previous otosclerosis candidate genes

217 Candidate genes

218 As two of the genome-wide significant loci from the meta-analysis overlapped with candidate genes
219 (*TGFB1* and *MEPE*), we also examined the association signals in the regions of other candidate genes
220 (Supplementary Data 9; Supplementary Data 10). Nine candidate variants were included in the
221 meta-analysis. Four previously reported susceptibility variants in or near the *COL1A1* gene were
222 associated with otosclerosis at nominal significance in the meta-analysis, but only the upstream
223 gene variant rs11327935 GA>G (EAF = 18%, OR = 1.11 [1.03-1.2], $p = 7.3 \times 10^{-3}$) was nominally
224 significant after multiple correction (OR = 1.12 [1.04-1.21], $p = 0.0041$).²⁷⁻²⁹ None of the previously
225 reported susceptibility variants in *SERPINF1* were included in the current meta-analysis.^{13,30}
226 Observable candidate variants in *BMP2* and *BMP4* were not significant in the meta-analysis.³¹
227 However, we observed a significant association approximately 300kb upstream of *BMP2*
228 overlapping the long non-coding RNA *CASC20*. The lead variant in this locus, rs2225347 A>T (EAF =
229 42%; OR = 0.82 [0.76-0.88]; $p = 6.2 \times 10^{-8}$), is not reported as eQTL or sQTL for *BMP2* or any gene in
230 the GTEx v8 database.²³

231 Linkage loci

232 Among published linkage loci, one (OTSC7, 6q13–16.1) has been narrowed to a region containing
233 only 12 candidate genes.³² In this region, we observed a trend towards association for variants near
234 the gene *CD109*, with the lowest p-value for the intronic *CD109* variant rs10805974 G>C (AF= 36%,
235 OR = 1.16 [1.09-1.23], $p = 3.4 \times 10^{-6}$). *CD109* was significantly associated with otosclerosis in the gene-
236 based analysis using MAGMA ($p = 1.2 \times 10^{-8}$).

237 **Further associations with the 18 otosclerosis lead variants**

238 *Phenome-wide association studies*

239 To assess pleiotropic effects of genetic risk factors to otosclerosis, we first examined the phenome-
240 wide associations of the 18 lead variants from the meta-analysis with a total of 2419 traits in UKBB
241 and 2925 traits in FinnGen. Six lead variants were associated with standing height, and five with heel
242 bone mineral density (BMD) in UKBB ($p < 2.1 \times 10^{-5}$) (Supplementary Data 13). The intronic *MARK3*
243 lead variant rs1951391 was associated with forearm fractures in FinnGen ($p = 3.8 \times 10^{-6}$) and the
244 rs753138805 frameshift variant in *MEPE* was associated with lower leg fractures ($p = 3.8 \times 10^{-6}$).

245 *GWAS Catalog*

246 Among the 18 lead variants, three were reported in the GWAS Catalog as being associated with
247 either height, BMD or BMI-adjusted waist-hip ratio (Supplementary Data 13).³³ Overall, among all
248 the 1,257 genome-wide significant variants associated with otosclerosis in our meta-analysis, 62
249 variants have been reported to be associated with a total of 45 traits (Supplementary Data 14). Eight
250 loci have been associated with BMD, four with height, and two with osteoarthritis. Variants
251 associated with increased risk of otosclerosis were most often associated with increased height and
252 decreased BMD. Additionally, in the chromosome 8 locus near the *EIF3H* gene, an intergenic variant
253 associated with otosclerosis at genome-wide significance (rs13279799, OR = 0.77 [0.71-0.83]; $p =$
254 2.8×10^{-11}) has been previously associated with ossification of the longitudinal ligament of spine, a
255 rare form of heterotopic ossification.³⁴

256 *Shared heritability*

257 To interrogate shared heritability between the traits, we calculated pairwise genetic correlations
258 (r_g) using summary statistics from the otosclerosis meta-analysis and published large GWA
259 studies.^{20,35,36} Otosclerosis was genetically correlated with fracture risk ($r_g = 0.24$ [S.E. 0.08], $p =$
260 0.0034) and height ($r_g = 0.10$ [S.E. 0.04], $p = 0.0177$), but not with BMD ($r_g = -0.088$ [S.E. 0.056], $p =$
261 0.12).

262 Discussion

263 Although otosclerosis is highly heritable its genetic background is still poorly understood. Previous
264 studies have reported one GWAS locus (*RELN*), while results for candidate genes have been
265 inconclusive. Here, we identify 18 loci associated with otosclerosis, of which 15 are novel. We also
266 replicate the *RELN* locus and two candidate gene loci – *TGFB1* and *MEPE*. These results offer further
267 insights to disease pathophysiology.

268 We confirm a polygenic basis to otosclerosis and identify loci harbouring potentially causative genes
269 involved in the regulation of bone structure. Earlier studies assumed a dominant inheritance, but
270 the inability to map otosclerosis to one or a few high-impact genes has suggested a complex nature.
271 Many etiologies to otosclerosis have been proposed, including differences in TGF β , parathyroid
272 hormone or angiotensin II signalling, alterations in collagen type I, inflammation, viral disease, and
273 autoimmunity.³⁷

274 A considerable number of association loci are associated with severe skeletal disorders (Table 3).
275 Disordered or heterotopic bone growth is a common feature e.g. in diaphyseal dysplasia (caused by
276 mutations in *TGFB1*), osteopetrosis subtypes (caused by mutations in *CLCN7*), and ossification of
277 the posterior longitudinal ligament of spine (GWAS signal for rs13279799 in the chromosome 8
278 locus). Otosclerosis presents later in adulthood and represents a more limited and common
279 phenotype compared with severe skeletal dysplasias. Shared heritability with height and
280 appendicular fracture risk across the genome also suggests more cumulative effects on the
281 regulation of bone growth and remodeling.

282 The effects on bone growth and remodeling are likely mediated by several pathways. The known
283 biological functions of many putative risk genes in otosclerosis association loci include regulation of
284 osteoblasts – the cells responsible for bone formation – and osteoclasts – responsible for the
285 breakdown of bone, either through effects on cell differentiation or function (Table 3). Some genes
286 (*AHSG*, *IRX5*, and *MEPE*) are also directly involved in biomineralization, and *AHSG* and the candidate
287 gene *MEPE* have been reported to have dual roles in the regulation of mineralization and osteoclast
288 or osteoblast cell lineages.

289 Recently, Shrauwen and colleagues proposed a model of otosclerosis in which increased bone
290 turnover results from mutations ablating two functional motifs of *MEPE*, leading to both accelerated
291 osteoclast differentiation and enhanced mineralization and thus increased bone turnover.³⁸ In
292 contrast with otosclerosis, they observed frameshift mutations ablating only one functional motif
293 of *MEPE* in individuals with a severe craniofacial defect with thickening of the skull, potentially
294 reflecting an unopposed increase in mineralization. Here, we replicate at genome-wide significance
295 an association between otosclerosis and a frameshift mutation that is likely to ablate both functional
296 motifs in *MEPE*. Reflecting a systemic effect on the skeleton, the rs753138805 variant is also
297 associated with increased leg fracture risk, similarly to an unpublished study in the Norwegian HUNT
298 cohort.³⁹

299 Several genes in the susceptibility loci converge on Transforming growth factor beta-1 (TGFβ1)
300 signalling pathways. *TGFB1* has long been proposed as a susceptibility gene for otosclerosis, but the
301 quality of the supporting evidence has been relatively weak.^{15-18,28,40} Here, for the first time, we
302 report an association at the *TGFB1* locus at genome-wide significance, with the strongest association
303 for the intronic variant rs8105161. The significance of TGFβ1 signalling is supported by gene set
304 enrichment and analysis of other loci.

305 TGFβ1, a member of the transforming growth factor beta superfamily, is a cytokine with an essential
306 role in skeletal development and mature bone remodeling, regulating both osteoblast and
307 osteoclast cell lineages.^{41,42} Gain-of-function mutations in *TGFB1* predispose to diaphyseal
308 dysplasia, and mutations in other TGFβ superfamily members can cause multiple skeletal disorders
309 of varying severity.⁴³

310 Other potentially causative TGFβ-related genes in the otosclerosis susceptibility loci include *RUNX2*
311 in chromosome 3, *AHSG* in chromosome 8, and *LTBP3* and *FAM89B* in chromosome 11, all of which
312 are associated with the expression of the respective lead variants. *RUNX2* is a transcription factor
313 regulated by TGFβ1-signalling via both the canonical Smad and p38 MAPK pathways and has an
314 essential role in osteoblast and chondrocyte differentiation.^{41,42,44} Alpha2-HS-glycoprotein/fetuin
315 (*AHSG*) antagonizes TGFβ1-signalling by binding to TGFβ1 and TGFβ related bone morphogenic
316 proteins (BMPs).^{45,46} It can also affect mineralization directly by inhibiting calcium phosphate
317 precipitation via formation of calcium-fetuin complexes.⁴⁷⁻⁴⁹ *LTBP3* can regulate the latency and
318 activation of TGFβ1 through direct extracellular binding.⁵⁰ The nearby *FAM89B* codes for Leucine

319 repeat adapter protein 25, which may regulate TGF β signaling by preventing the nuclear
320 translocation of Smad2.⁵¹

321 In addition to signaling proteins, *COL1A1* coding for the major subunit of type I collagen is a
322 prominent candidate gene for otosclerosis.^{12,27-29,52} Mutations in *COL1A1* predispose to
323 osteogenesis imperfecta, often also characterized by hearing loss. In osteogenesis imperfecta,
324 similarly to carriers of truncating *MEPE* variants, increased bone turnover in the middle could occur
325 in association with general skeletal fragility. While we do not observe a strong signal for *COL1A1*,
326 similarly to most examined candidate genes, we report an intronic association signal in *COL4A2*,
327 coding for a subunit of collagen IV, located in the basement membrane. Future mechanistic studies
328 could examine whether *COL4A2* has a structural or signaling role in otosclerosis. *COL4A2* is highly
329 conserved across species, and its mutations have been reported in a broad spectrum of organ
330 anomalies, but we are not aware of a previous association with skeletal disorders.⁵³ Although type
331 IV collagens regulate BMPs in *Drosophila*, a similar role in humans is uncertain.⁵⁴

332 Our results highlight several genes and signaling pathways for follow-up mechanistic studies.
333 Genetic discovery has also increasingly preceded or guided therapeutic development. Future
334 sequencing studies could aim to discover loss-of-function mutations within the susceptibility loci to
335 approximate the effects of therapeutic inhibition. In the case of *TGFB1*, loss-of-function mutations
336 are exceedingly rare and functional studies are needed to assess the effect of direct TGF β 1
337 inhibition.²⁴ Of note, an inhibitor of BMP type I receptor kinases has been studied for the treatment
338 of ectopic ossification; however, in our meta-analysis, its targets (coded by *ACVR1* and *BMPRI1A*)
339 were not significantly associated with otosclerosis.⁵⁵ Reflecting the heterogeneity of conditions
340 associated with different TGF β superfamily genes, therapeutic intervention may need to be
341 precisely tailored to each condition.⁴¹

342 Although we present the largest GWAS study of otosclerosis to date, our study has limitations. The
343 identification of otosclerosis cases is based on ICD diagnoses, which we could not verify with
344 audiometric or imaging data. However, the prevalence of stapes procedures among cases is
345 concordant with clinical experience. As bone data are not available within the GTEx consortium, the
346 analyses of gene expression rely on other tissues; associations in bone could differ. The physical
347 proximity of lead variants with biologically relevant genes does not prove causation; identified and
348 unidentified variants in the loci can exert their effects through yet unknown mechanisms.

349 In summary, we determine a polygenic basis to otosclerosis with 18 genome-wide significant
350 susceptibility loci, and shared heritability between otosclerosis and other common and rare skeletal
351 traits. The loci offer multiple potential avenues for understanding disease pathophysiology through
352 genes involved in bone remodeling, mineralization, and basement membrane collagen composition.
353 In particular, our results highlight several TGF β superfamily members for follow-up studies.

354 **Methods**

355 **Cohort descriptions**

356 We identified individuals with ICD-based otosclerosis diagnoses from three national biobank-based
357 cohorts: The Finnish FinnGen cohort, the Estonian Biobank (EstBB) and the UK Biobank (UKBB).

358 The FinnGen data used here comprise 210,932 individuals from FinnGen Data Freeze 5
359 (<https://www.finngen.fi>). The data were linked by unique national personal identification numbers
360 to the national hospital discharge registry (available from 1968) and the specialist outpatient
361 registry (available from 1998). Data comprised in FinnGen Data Freeze 5 are administered by
362 regional biobanks (Auria Biobank, Biobank of Central Finland, Biobank of Eastern Finland, Borealis
363 Biobank, Helsinki Biobank, Tampere Biobank), the Blood Service Biobank, the Terveystalo Biobank,
364 and biobanks administered by the Finnish Institute for Health and Welfare (THL) for the following
365 studies: Botnia, Corogene, FinHealth 2017, FinIPF, FINRISK 1992–2012, GeneRisk, Health 2000,
366 Health 2011, Kuusamo, Migraine, Super, T1D, and Twins). The FinnGen study protocol
367 (HUS/990/2017) was approved by the Ethics Review Board of the Hospital District of Helsinki and
368 Uusimaa.

369 EstBB is a population-based cohort of 200,000 participants with a rich variety of phenotypic and
370 health-related information collected for each individual.⁵⁶ At recruitment, participants have signed
371 a consent to allow follow-up linkage of their electronic health records (EHR), thereby providing a
372 longitudinal collection of phenotypic information. EstBB allows access to the records of the national
373 Health Insurance Fund Treatment Bills (from 2004), Tartu University Hospital (from 2008), and North
374 Estonia Medical Center (from 2005). For every participant there is information on diagnoses in ICD-
375 10 coding and drug dispensing data, including drug ATC codes, prescription status and purchase
376 date (if available).

377 UKBB comprises phenotype data from 500,000 volunteer participants from the UK population aged
378 between 40 and 69 years during recruitment in 2006-2010.⁵⁷ Data for all participants have been
379 linked with national Hospital Episode Statistics.

380 **Identification of otosclerosis cases**

381 Case status was assigned for individuals with any ICD-10 H80* code, or when available, the ICD-9
382 code 387 or ICD-8 code 386 (Table 1). In FinnGen, only ICD codes registered in specialty care settings
383 were used for case definition. Stapes procedures were identified based on the Nomesco codes
384 DDA00 (stapedotomy) and DDB00 (stapedectomy) in FinnGen and the national health insurance
385 treatment service code 61006 (stapedotomy) in EstBB.

386 **Genotyping and imputation of variants**

387 FinnGen samples were genotyped using Illumina and Affymetrix arrays (Illumina Inc., San Diego, and
388 Thermo Fisher Scientific, Santa Clara, CA, USA). Genotype imputation was performed using a
389 population-specific SISu v3 imputation reference panel comprised of 3,775 whole genomes as
390 described (<https://www.protocols.io/view/genotype-imputation-workflow-v3-0-xbgfijw>).

391 The samples from the Estonian Biobank were genotyped at the Genotyping Core Facility of the
392 Institute of Genomics, University of Tartu using the Global Screening Array (GSAv1.0, GSAv2.0, and
393 GSAv2.0_EST) from Illumina. Altogether 155,772 samples were genotyped and PLINK format files
394 were exported using GenomeStudio v2.0.4.⁵⁸ Individuals were excluded from the analysis if their
395 call-rate was < 95% or if the sex defined based on heterozygosity of the X chromosome did not
396 match the sex in phenotype data. Variants were excluded if the call-rate was < 95% and HWE p-
397 value < 1e-4 (autosomal variants only). Variant positions were updated to genome build 37 and all
398 alleles were switched to the TOP strand using tools and reference files provided at
399 <https://www.well.ox.ac.uk/~wrayner/strand/>. After QC the dataset contained 154,201 samples for
400 imputation. Before imputation variants with MAF<1% and indels were removed. Prephasing was
401 done using the Eagle v2.3 software.⁵⁹ The number of conditioning haplotypes Eagle2 uses when
402 phasing each sample was set to: --Kpbwt=20000. Imputation was done using Beagle v.28Sep18.⁷⁹³
403 with effective population size ne=20,000.⁶⁰ An Estonian population specific imputation reference of
404 2297 WGS samples was used.⁶¹

405 UKBB samples have been genotyped and imputed as described previously.⁵⁷ Additional quality
406 controls of version 3 data have been performed and publicly reported by the Pan-UKB team from
407 the Analytic & Translational Genetics Unit (ATGU) at Massachusetts General Hospital and the Broad
408 Institute of MIT and Harvard (<https://pan.ukbb.broadinstitute.org>. 2020). The analysis referenced
409 in this study is based on samples from 419,622 individuals with European ancestry.

410 **Statistical analysis**

411 GWAS for the individual study cohorts was performed using a generalized mixed model with the
412 saddlepoint approximation using SAIGE v0.20, using a kinship matrix as a random effect and
413 covariates as fixed effects.⁶² For all cohorts, variants with a minor allele frequency less than 0.1%
414 were excluded.

415 In FinnGen, samples from individuals with non-Finnish ancestry and twin/duplicate samples were
416 excluded, and GWAS was performed for 1,350 cases and 209,582 controls. Age, sex, 10 PCs and the
417 genotyping batch (for batches with at least 10 cases and controls) were used as covariates. Results
418 were filtered to variants with imputation INFO score > 0.6. In EstBB, a GWAS was performed on 713
419 cases and 136,7933 controls of European ancestry including related individuals and adjusting for the
420 first 10 PCs of the genotype matrix, as well as for birth year and sex. The statistical analysis of UKBB
421 data has been performed and publicly reported by the Pan-UKB team
422 (<https://pan.ukbb.broadinstitute.org>. 2020). The UKBB GWAS results are reproduced here for
423 selected variants for the purpose of cross-cohort replication and inclusion in the meta-analysis.

424 Variant positions in the UKBB and EstBB summary statistics were lifted from GRCh37 to GRCh38
425 using the liftOver v1.12 Bioconductor R package.⁶³ To assess similarity of genetic effects between
426 cohorts, we used the cross-trait LD Score Regression to calculate pairwise genetic correlations (r_g)
427 based on summary statistics from each cohort.^{20,21} Summary statistics from the individual cohorts
428 for 13,762,819 variants present in at least two cohorts with a cross-cohort minor allele frequency >
429 0.1% and imputation INFO score > 0.7 were combined using an inverse-variance weighted fixed-
430 effect meta-analysis with GWAMA v2.2.2.⁶⁴ The following analyses are based on summary statistics
431 from the meta-analysis unless otherwise specified.

432 **Characterization of association loci**

433 For the individual cohorts and meta-analysis, we merged genome-wide significant variants within
434 500kb of each other into association loci. The lead variants from these loci were not in LD with each
435 other based on individual-level data from FinnGen ($r^2 < 1 \cdot 10^{-4}$ for all variant pair comparisons). We
436 annotated the lead variants by mapping their physical position to genes and consequences using
437 the Ensembl Variant Effect Predictor (VEP) based on the GRCh38 genome build.⁶⁵ In addition, we

438 annotated all genome-wide significant variants and all variants in high LD ($r > 0.6$) with the lead
439 variants from the meta-analysis using VEP. As the highest number of association loci was observed
440 in FinnGen, we estimated LD in the FinnGen cohort using PLINK v1.07.^{66,67} LocusZoom was used to
441 visualize the association loci and the regions surrounding other genes of interest.⁶⁸

442 **Fine-mapping**

443 Based on otosclerosis summary statistics from FinnGen, we fine-mapped all regions where the lead
444 variant reached a p -value of $< 1 \times 10^{-6}$ in FinnGen using FINEMAP v1.4.²⁵ We used a 3Mb window (\pm
445 1.5Mb) around each lead variant, and calculated LD between each variant from individual-level
446 FinnGen data. For credible sets with over 10 variants, we report those with a ≥ 0.01 probability of
447 being causal.

448 **eQTL and sQTL mapping**

449 We used the GTEx v8 database (<https://gtexportal.org>) to map otosclerosis susceptibility variants
450 to genes and tissues.²³ We downloaded the data for all 49 tissues, and mapped the lead variants
451 from the meta-analysis, as well as selected variants of interest (e.g. from candidate gene loci), to
452 eQTL (expression quantitative locus) or sQTL (splicing quantitative locus) associations using the 5%
453 false discovery rate cut-off provided by the GTEx Consortium.

454 **Gene and gene set -based analyses**

455 We used MAGMA v1.07 to identify genes and gene sets associated with otosclerosis based on effect
456 estimates from the meta-analysis.⁶⁹ Variants were mapped to 18,910 genes based on their RefSNP
457 numbers. We performed a gene-based analysis using the default SNPwise-mean model and a
458 Bonferroni-corrected p -value threshold ($\alpha = 0.05/18,910$). Based on the results from the gene-based
459 analysis, we then performed a competitive gene set based analysis using 10,182 GO term based
460 gene sets downloaded from the Molecular Signature Database v7.1, using a Bonferroni-corrected
461 p -value threshold ($\alpha = 0.05/10,182$).⁷⁰

462 **Phenome-wide association analyses and genetic correlations**

463 We examined the pleiotropic effects of lead variants based on phenome-wide association studies in
464 FinnGen and UKBB. In FinnGen, GWAS was performed for 2,925 registry-based disease endpoints
465 similarly to the GWAS for otosclerosis. The study participants were linked with national registries
466 covering the whole population for hospital discharges (data available since 1968), deaths (1969–),
467 outpatient specialist appointments (1998–), cancers (1953–), and medication reimbursements
468 (1995–). Disease endpoints were collated based on International Classification of Diseases (ICD)
469 codes (revisions 8–10), International Classification of Diseases for Oncology (ICD-O) third edition
470 codes, NOMESCO procedure codes, Finnish-specific Social Insurance Institute (KELA) drug
471 reimbursement codes, and drug-specific ATC-codes. In addition, specific clinical endpoints
472 combining relevant comorbidities and exclusion criteria have been curated in coordination with
473 clinical expert groups. Currently available FinnGen phenotype definitions are available online at
474 <https://www.finnngen.fi/en/researchers/clinical-endpoints>.

475 For UKBB, we downloaded publicly available association results for 2419 phenotypes including
476 diseases and anthropometric traits (<http://www.nealelab.is/uk-biobank> &
477 <http://pheweb.sph.umich.edu:5000>). Bonferroni-corrected p -value thresholds were used for each
478 cohort ($\alpha = 0.05/2925$ for FinnGen and $\alpha = 0.05/2419$ for UKBB). Additionally, using the GWAS
479 Catalog database (accessed on 22 April 2020), we performed a lookup of previously published
480 association results for all variants significantly associated with otosclerosis in the meta-analysis.³³

481 For three frequently occurring coassociations (height, bone mineral density (BMD), and fracture
482 risk, we estimated the genetic correlations of each trait with otosclerosis using LD Score
483 Regression.^{20,21} We used summary statistics from recently published large GWA studies: the GWAS
484 meta-analysis of height in approximately 700,000 individuals by the GIANT Consortium, and the
485 GWAS of ultrasound-quantified heel BMD and fracture risk in approximately 460,000 UKBB
486 participants by the GEFOS Consortium.^{20,35,36}

487 **Acknowledgements**

488 This research has been conducted using the UK Biobank Resource under Application Number 22627.
489 The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office
490 of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and
491 NINDS. The data used for the gene expression analyses described in this manuscript were obtained
492 from the GTEx Portal in July 2020. We thank the Pan UKBB analysis team from the Analytic &
493 Translational Genetics Unit (ATGU) at Massachusetts General Hospital and the Broad Institute of
494 MIT and Harvard for making UKBB summary statistics publicly available. We thank all study
495 participants for their generous contributions to the biobanks.

496 References

- 497 1 Quesnel, A. M., Ishai, R. & McKenna, M. J. Otosclerosis: Temporal Bone Pathology.
498 *Otolaryngol. Clin. North Am.* **51**, 291-303, doi:10.1016/j.otc.2017.11.001 (2018).
- 499 2 Babcock, T. A. & Liu, X. Z. Otosclerosis: From Genetics to Molecular Biology. *Otolaryngol.*
500 *Clin. North Am.* **51**, 305-318, doi:10.1016/j.otc.2017.11.002 (2018).
- 501 3 Ramaswamy, A. T. & Lustig, L. R. Revision Surgery for Otosclerosis. *Otolaryngol. Clin. North*
502 *Am.* **51**, 463-474, doi:10.1016/j.otc.2017.11.014 (2018).
- 503 4 Declau, F. *et al.* Prevalence of otosclerosis in an unselected series of temporal bones.
504 *Otology and Neurotology* **22**, 596-602, doi:10.1097/00129492-200109000-00006 (2001).
- 505 5 Declau, F. *et al.* Prevalence of histologic otosclerosis: an unbiased temporal bone study in
506 Caucasians. *Adv. Otorhinolaryngol.* **65**, 6-16, doi:10.1159/000098663 (2007).
- 507 6 House, H. P., Hansen, M. R., Al Dakhil, A. A. & House, J. W. Stapedectomy versus
508 stapedotomy: comparison of results with long-term follow-up. *Laryngoscope* **112**, 2046-
509 2050, doi:10.1097/00005537-200211000-00025 (2002).
- 510 7 Del Bo, M., Zaghis, A. & Ambrosetti, U. Some observations concerning 200 stapedectomies:
511 fifteen years postoperatively. *Laryngoscope* **97**, 1211-1213, doi:10.1288/00005537-
512 198710000-00017 (1987).
- 513 8 Cawthorne, T. Otosclerosis. *J. Laryngol. Otol.* **69**, 437-456,
514 doi:10.1017/s0022215100050933 (1955).
- 515 9 Morrison, A. W. Genetic factors in otosclerosis. *Ann. R. Coll. Surg. Engl.* **41**, 202-237 (1967).
- 516 10 Moumoulidis, I., Axon, P., Baguley, D. & Reid, E. A review on the genetics of otosclerosis.
517 *Clin. Otolaryngol.* **32**, 239-247, doi:10.1111/j.1365-2273.2007.01475.x (2007).
- 518 11 Larsson, A. Otosclerosis. A genetic and clinical study. *Acta Otolaryngol. Suppl.* **154**, 1-86
519 (1960).
- 520 12 Bittermann, A. J. *et al.* An introduction of genetics in otosclerosis: a systematic review.
521 *Otolaryngol. Head Neck Surg.* **150**, 34-39, doi:10.1177/0194599813509951 (2014).
- 522 13 Valgaeren, H. *et al.* Insufficient evidence for a role of SERPINF1 in otosclerosis. *Mol. Genet.*
523 *Genomics* **294**, 1001-1006, doi:10.1007/s00438-019-01558-8 (2019).
- 524 14 Schrauwen, I. *et al.* A Genome-wide Analysis Identifies Genetic Variants in the RELN Gene
525 Associated with Otosclerosis. *The American Journal of Human Genetics* **84**, 328-338,
526 doi:<https://doi.org/10.1016/j.ajhg.2009.01.023> (2009).
- 527 15 Mowat, A. J. *et al.* Evidence of distinct RELN and TGFB1 genetic associations in familial and
528 non-familial otosclerosis in a British population. *Hum. Genet.* **137**, 357-363,
529 doi:10.1007/s00439-018-1889-9 (2018).
- 530 16 Priyadarshi, S., Hansdah, K., Ray, C. S., Biswal, N. C. & Ramchander, P. V. Otosclerosis
531 Associated with a De Novo Mutation -832G > A in the TGFB1 Gene Promoter Causing a
532 Decreased Expression Level. *Sci. Rep.* **6**, 29572, doi:10.1038/srep29572 (2016).
- 533 17 Priyadarshi, S. *et al.* Genetic association and gene expression profiles of TGFB1 and the
534 contribution of TGFB1 to otosclerosis susceptibility. *J. Bone Miner. Res.* **28**, 2490-2497,
535 doi:10.1002/jbmr.1991 (2013).
- 536 18 Thys, M. *et al.* Detection of Rare Nonsynonymous Variants in TGFB1 in Otosclerosis
537 Patients. *Ann. Hum. Genet.* **73**, 171-175, doi:10.1111/j.1469-1809.2009.00505.x (2009).
- 538 19 Schrauwen, I. *et al.* Association of Bone Morphogenetic Proteins With Otosclerosis. *J. Bone*
539 *Miner. Res.* **23**, 507-516, doi:doi:10.1359/jbmr.071112 (2008).
- 540 20 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits.
541 *Nat. Genet.* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).

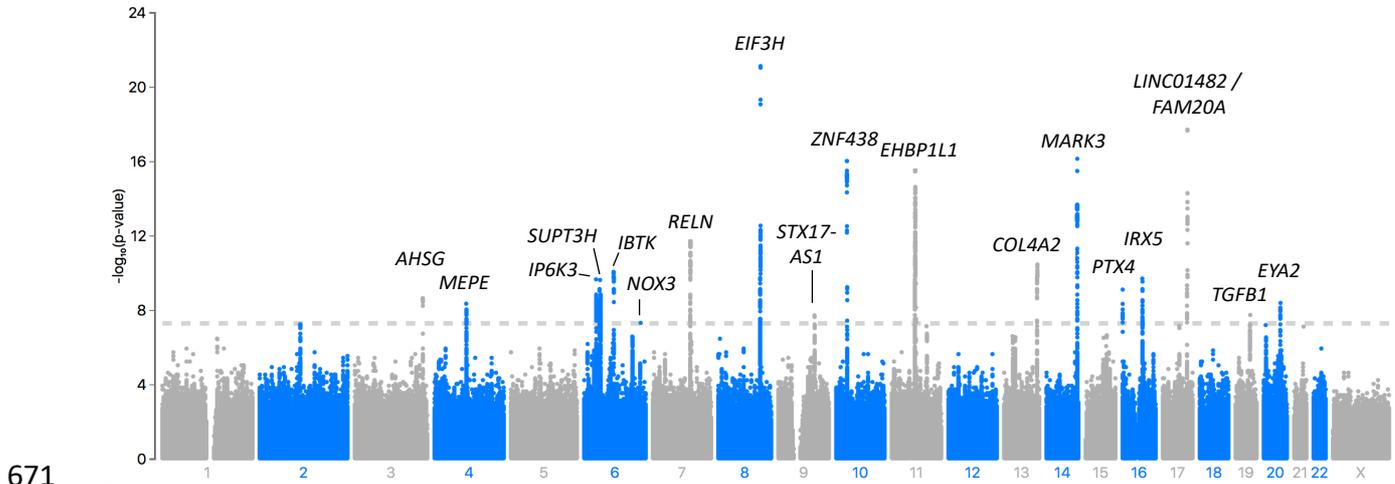
- 542 21 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity
543 in genome-wide association studies. *Nat. Genet.* **47**, 291-295, doi:10.1038/ng.3211 (2015).
- 544 22 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits.
545 *Nat. Genet.* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).
- 546 23 Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human
547 tissues. *bioRxiv*, 787903, doi:10.1101/787903 (2019).
- 548 24 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in
549 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).
- 550 25 Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-
551 wide association studies. *Bioinformatics* **32**, 1493-1501,
552 doi:10.1093/bioinformatics/btw018 (2016).
- 553 26 Massagué, J. TGF β signalling in context. *Nature Reviews Molecular Cell Biology* **13**, 616-630,
554 doi:10.1038/nrm3434 (2012).
- 555 27 Schrauwen, I. *et al.* COL1A1 association and otosclerosis: A meta-analysis. *American*
556 *Journal of Medical Genetics Part A* **158A**, 1066-1070, doi:10.1002/ajmg.a.35276 (2012).
- 557 28 Khalfallah, A. *et al.* Association of COL1A1 and TGFB1 Polymorphisms with Otosclerosis in a
558 Tunisian Population. *Ann. Hum. Genet.* **75**, 598-604, doi:10.1111/j.1469-1809.2011.00665.x
559 (2011).
- 560 29 Chen, W. *et al.* Single-nucleotide polymorphisms in the COL1A1 regulatory regions are
561 associated with otosclerosis. *Clin. Genet.* **71**, 406-414, doi:10.1111/j.1399-
562 0004.2007.00794.x (2007).
- 563 30 Ziff, J. L. *et al.* Mutations and altered expression of SERPINF1 in patients with familial
564 otosclerosis. *Hum. Mol. Genet.* **25**, 2393-2403, doi:10.1093/hmg/ddw106 (2016).
- 565 31 Schrauwen, I. *et al.* Association of bone morphogenetic proteins with otosclerosis. *J. Bone*
566 *Miner. Res.* **23**, 507-516, doi:10.1359/jbmr.071112 (2008).
- 567 32 Pauw, R. J. *et al.* Phenotype description of a Dutch otosclerosis family with suggestive
568 linkage to OTSC7. *American Journal of Medical Genetics Part A* **143A**, 1613-1622,
569 doi:10.1002/ajmg.a.31807 (2007).
- 570 33 Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association
571 studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005-d1012,
572 doi:10.1093/nar/gky1120 (2019).
- 573 34 Nakajima, M. *et al.* A genome-wide association study identifies susceptibility loci for
574 ossification of the posterior longitudinal ligament of the spine. *Nat. Genet.* **46**, 1012-1016,
575 doi:10.1038/ng.3045 (2014).
- 576 35 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass
577 index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641-3649,
578 doi:10.1093/hmg/ddy271 (2018).
- 579 36 Morris, J. A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nat.*
580 *Genet.* **51**, 258-266, doi:10.1038/s41588-018-0302-x (2019).
- 581 37 Rudic, M. *et al.* The pathophysiology of otosclerosis: Review of current research. *Hear. Res.*
582 **330**, 51-56, doi:10.1016/j.heares.2015.07.014 (2015).
- 583 38 Schrauwen, I. *et al.* Variants affecting diverse domains of MEPE are associated with two
584 distinct bone disorders, a craniofacial bone defect and otosclerosis. *Genet. Med.* **21**, 1199-
585 1208, doi:10.1038/s41436-018-0300-5 (2019).
- 586 39 Van Hout, C. V. *et al.* Whole exome sequencing and characterization of coding variation in
587 49,960 individuals in the UK Biobank. *bioRxiv*, 572347, doi:10.1101/572347 (2019).

- 588 40 Schrauwen, I. *et al.* A genome-wide analysis identifies genetic variants in the RELN gene
589 associated with otosclerosis. *Am. J. Hum. Genet.* **84**, 328-338,
590 doi:10.1016/j.ajhg.2009.01.023 (2009).
- 591 41 Wu, M., Chen, G. & Li, Y.-P. TGF- β and BMP signaling in osteoblast, skeletal development,
592 and bone formation, homeostasis and disease. *Bone Research* **4**, 16009,
593 doi:10.1038/boneres.2016.9 (2016).
- 594 42 Crane, J. L. & Cao, X. Bone marrow mesenchymal stem cells and TGF- β signaling in bone
595 remodeling. *The Journal of clinical investigation* **124**, 466-472, doi:10.1172/JCI70050
596 (2014).
- 597 43 Kinoshita, A. *et al.* Domain-specific mutations in TGFB1 result in Camurati-Engelmann
598 disease. *Nat. Genet.* **26**, 19-20, doi:10.1038/79128 (2000).
- 599 44 Komori, T. Regulation of Proliferation, Differentiation and Functions of Osteoblasts by
600 Runx2. *Int. J. Mol. Sci.* **20**, doi:10.3390/ijms20071694 (2019).
- 601 45 Demetriou, M., Binkert, C., Sukhu, B., Tenenbaum, H. C. & Dennis, J. W. Fetuin/alpha2-HS
602 glycoprotein is a transforming growth factor-beta type II receptor mimic and cytokine
603 antagonist. *J. Biol. Chem.* **271**, 12755-12761, doi:10.1074/jbc.271.22.12755 (1996).
- 604 46 Szweras, M. *et al.* alpha 2-HS glycoprotein/fetuin, a transforming growth factor-beta/bone
605 morphogenetic protein antagonist, regulates postnatal bone growth and remodeling. *J.*
606 *Biol. Chem.* **277**, 19991-19997, doi:10.1074/jbc.M112234200 (2002).
- 607 47 Schinke, T. *et al.* The serum protein alpha2-HS glycoprotein/fetuin inhibits apatite
608 formation in vitro and in mineralizing calvaria cells. A possible role in mineralization and
609 calcium homeostasis. *J. Biol. Chem.* **271**, 20789-20796, doi:10.1074/jbc.271.34.20789
610 (1996).
- 611 48 Heiss, A. *et al.* Structural basis of calcification inhibition by alpha 2-HS glycoprotein/fetuin-
612 A. Formation of colloidal calciprotein particles. *J. Biol. Chem.* **278**, 13333-13341,
613 doi:10.1074/jbc.M210868200 (2003).
- 614 49 Price, P. A. & Lim, J. E. The inhibition of calcium phosphate precipitation by fetuin is
615 accompanied by the formation of a fetuin-mineral complex. *J. Biol. Chem.* **278**, 22144-
616 22152, doi:10.1074/jbc.M300744200 (2003).
- 617 50 Robertson, I. B. *et al.* Latent TGF- β -binding proteins. *Matrix Biol.* **47**, 44-53,
618 doi:10.1016/j.matbio.2015.05.005 (2015).
- 619 51 Kokura, K. *et al.* The Ski-binding Protein C184M Negatively Regulates Tumor Growth
620 Factor- β Signaling by Sequestering the Smad Proteins in the Cytoplasm. *J. Biol. Chem.* **278**,
621 20133-20139, doi:10.1074/jbc.M210855200 (2003).
- 622 52 McKenna, M. J., Kristiansen, A. G., Bartley, M. L., Rogus, J. J. & Haines, J. L. Association of
623 COL1A1 and otosclerosis: evidence for a shared genetic etiology with mild osteogenesis
624 imperfecta. *Am. J. Otol.* **19**, 604-610 (1998).
- 625 53 Meuwissen, M. E. *et al.* The expanding phenotype of COL4A1 and COL4A2 mutations:
626 clinical data on 13 newly identified families and a review of the literature. *Genet. Med.* **17**,
627 843-853, doi:10.1038/gim.2014.210 (2015).
- 628 54 Wang, X., Harris, R. E., Bayston, L. J. & Ashe, H. L. Type IV collagens regulate BMP signalling
629 in *Drosophila*. *Nature* **455**, 72-77, doi:10.1038/nature07214 (2008).
- 630 55 Yu, P. B. *et al.* BMP type I receptor inhibition reduces heterotopic [corrected] ossification.
631 *Nat. Med.* **14**, 1363-1369, doi:10.1038/nm.1888 (2008).
- 632 56 Leitsalu, L., Alavere, H., Tammesoo, M. L., Leego, E. & Metspalu, A. Linking a population
633 biobank with national health registries-the estonian experience. *Journal of personalized*
634 *medicine* **5**, 96-106, doi:10.3390/jpm5020096 (2015).

- 635 57 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
636 *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
- 637 58 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
638 datasets. *GigaScience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).
- 639 59 Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel.
640 *Nat. Genet.* **48**, 1443-1448, doi:10.1038/ng.3679 (2016).
- 641 60 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data
642 inference for whole-genome association studies by use of localized haplotype clustering.
643 *Am. J. Hum. Genet.* **81**, 1084-1097, doi:10.1086/521987 (2007).
- 644 61 Mitt, M. *et al.* Improved imputation accuracy of rare and low-frequency variants using
645 population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum.*
646 *Genet.* **25**, 869-876, doi:10.1038/ejhg.2017.51 (2017).
- 647 62 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in
648 large-scale genetic association studies. *Nat. Genet.* **50**, 1335-1341, doi:10.1038/s41588-
649 018-0184-y (2018).
- 650 63 Bioconductor Package Maintainer (2020). liftOver: Changing genomic coordinate systems
651 with rtracklayer::liftOver. R package version 1.12.0,
652 <https://www.bioconductor.org/help/workflows/liftOver/>.
- 653 64 Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis.
654 *BMC Bioinformatics* **11**, 288, doi:10.1186/1471-2105-11-288 (2010).
- 655 65 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122,
656 doi:10.1186/s13059-016-0974-4 (2016).
- 657 66 PLINK v1.07. URL: <http://pngu.mgh.harvard.edu/purcell/plink/>.
- 658 67 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
659 linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575, doi:10.1086/519795 (2007).
- 660 68 Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan
661 results. *Bioinformatics (Oxford, England)* **26**, 2336-2337,
662 doi:10.1093/bioinformatics/btq419 (2010).
- 663 69 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set
664 analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219,
665 doi:10.1371/journal.pcbi.1004219 (2015).
- 666 70 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set
667 collection. *Cell systems* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).
- 668

669 **Figures and tables**

670 **Figure 1**



672 **Figure 1. Meta-analysis of the genome-wide association studies of Otosclerosis in FinnGen, EstBB**
673 **and UKBB.** In the Manhattan plot, chromosomal positions are indicated on the x-axis and $-\log_{10}(p$ -
674 $value)$ is presented on the y-axis for each variant. Results are presented for a fixed-effect meta-
675 analysis of effect estimates from GWASs in the three cohorts, including a total of 2,413 cases and
676 762,382 controls. The included variants were present in at least two cohorts with a cross-cohort
677 minor allele frequency $> 0.1\%$ and imputation INFO score > 0.7 . Eighteen loci reached genome-wide
678 significance ($p < 5 \times 10^{-8}$, marked by the dashed line). The loci are annotated by the names of the
679 genes nearest to the lead variants.

680

681 **Table 1. ICD and procedure codes corresponding to otosclerosis in each study cohort. Stapes**
 682 *procedures were identified based on the Nomesco codes DDA00 (stapedotomy) and DDB00*
 683 *(stapedectomy) in FinnGen and the national health insurance treatment service code 61006*
 684 *(stapedotomy) in EstBB.*

685

Classification	Code	Definition	FinnGen	EstBB	UKBB
ICD-10	H80	Otosclerosis (any)		256	318
	H80.0	.. Involving oval window, nonobliterative	655	246	10
	H80.1	.. Involving oval window, obliterative	185	63	3
	H80.2	.. Cochlear	45	32	3
	H80.8	.. Other	133	95	6
	H80.9	.. Unspecified	628	379	301
ICD-9	387	Otosclerosis (any)	272		36
ICD-8	386	Otosclerosis (any)	284		
Procedures		Stapedotomy	458	161	
Procedures		Stapedectomy	113		
Total cases			1350	713	350
Total controls			209582	136793	416007
Prevalence (%)			0.64	0.52	0.08

686

687 **Table 2. Lead variants for significant association loci in the meta-analysis of otosclerosis.** Data (in
688 cursive) are also presented for the rs753138805 variant in *MEPE*, observed only in FinnGen and not
689 included in the meta-analysis. *Chr* = chromosome, *EA* = effect allele, *EAF* = effect allele frequency,
690 *NEA* = non-effect allele, *OR* = odds ratio, *CI* = confidence interval.

Rsid	Chr	Position	EA	NEA	EAF	Nearest Gene	Consequence	OR (95% CI)	P-value
rs4876361	8	116554149	A	G	0.71	<i>EIF3H</i>	Intergenic	0.73 (0.68-0.78)	7.2*10 ⁻²²
rs8070086	17	68682159	A	G	0.24	<i>LINC01482 / FAM20A</i>	Intronic / Intergenic	1.39 (1.29-1.49)	1.8*10 ⁻¹⁸
rs1951391	14	103431976	C	T	0.33	<i>MARK3</i>	Intronic	1.38 (1.28-1.49)	7.0*10 ⁻¹⁷
rs947091	10	30765257	A	G	0.42	<i>ZNF438</i>	Regulatory Region	0.77 (0.73-0.82)	9.2*10 ⁻¹⁷
rs67924081	11	65575510	G	A	0.23	<i>EHBP1L1</i>	Upstream Gene	0.72 (0.67-0.78)	2.5*10 ⁻¹⁶
rs39352	7	103827282	T	G	0.41	<i>RELN</i>	Intronic	1.24 (1.17-1.32)	1.8*10 ⁻¹²
rs9559805	13	110462733	A	C	0.48	<i>COL4A2, COL4A2-AS2</i>	Synonymous	0.82 (0.77-0.87)	3.4*10 ⁻¹¹
rs10943838	6	81969549	C	G	0.61	<i>IBTK</i>	Upstream Gene	0.82 (0.77-0.87)	8.4*10 ⁻¹¹
rs3922616	16	55008491	C	A	0.43	<i>IRX5</i>	Intergenic	1.21 (1.14-1.29)	1.9*10 ⁻¹⁰
rs9469583	6	33749993	C	T	0.58	<i>IP6K3</i>	Upstream Gene	1.22 (1.15-1.3)	2.1*10 ⁻¹⁰
rs13192457	6	44887654	A	C	0.44	<i>SUPT3H</i>	Intronic	1.25 (1.17-1.34)	2.3*10 ⁻¹⁰
rs67284550	16	1480948	T	C	0.42	<i>PTX4</i>	Downstream Gene	1.2 (1.13-1.28)	7.6*10 ⁻¹⁰
rs4917	3	186619924	C	T	0.64	<i>AHSG</i>	Missense	0.83 (0.78-0.88)	2.3*10 ⁻⁹
rs6066131	20	46941020	C	T	0.45	<i>EYA2</i>	Intronic	1.19 (1.12-1.26)	4.0*10 ⁻⁹
rs7683897	4	87888898	A	G	0.41	<i>MEPE</i>	Intergenic	1.19 (1.12-1.26)	4.3*10 ⁻⁹
rs753138805	4	87845066	G	GGAAA	0.0032	<i>MEPE</i>	Frameshift	18.9 (8.05-44.4)	1.5*10 ⁻¹¹
rs8105161	19	41333726	C	T	0.16	<i>TGFB1</i>	Intronic	0.8 (0.74-0.86)	1.8*10 ⁻⁸
rs80339979	9	99889430	G	GA	0.64	<i>STX17-AS1</i>	Intronic	1.19 (1.12-1.26)	1.9*10 ⁻⁸
rs9397842	6	155810724	G	T	0.78	<i>NOX3</i>	Intergenic	0.79 (0.73-0.86)	4.7*10 ⁻⁸

691

692

693 **Table 3: Biological functions of genes in selected association loci.** In 13 association loci, genes
694 implicated in bone metabolism are in close proximity with the lead variant or associated with the
695 expression of the lead variant. eQTL = expression quantitative trait locus, sQTL = splicing
696 quantitative trait locus. References to individual studies are listed in **Supplementary Data 15.**

697

698

----- Table shown on next page -----

Rsid	Chr	Position	Nearest Gene	Consequence	Biological role(s) of selected gene(s) in locus	Severe skeletal phenotypes associated with variants in loci (ClinVar & GWAS Catalog)
rs4876361	8	116554149	<i>EIF3H</i>	Intergenic Variant	<i>TRPS1</i> (750kb from rs4876361): Trichorhinophalangeal syndrome type I. <i>TNFRSF11B</i> (2.3Mb from rs4876361) : Osteoprotegerin, inhibitor of osteoclast differentiation and function.	rs13279799: Ossification of the posterior longitudinal ligament of the spine. <i>TRPS1</i> : Trichorhinophalangeal syndrome types I and III, Langer-Giedion syndrome
rs1951391	14	103431976	<i>MARK3</i>	Intron Variant	<i>MARK3</i> (eQTL and sQTL) : Potential regulator of BMD. <i>CKB</i> (eQTL and sQTL): Regulator of the bone-resorbing function of osteoclasts.	
rs67924081	11	65575510	<i>EHBP1L1</i>	Upstream Gene Variant	<i>LTBP3</i> (eQTL and sQTL): Latent Transforming Growth Factor Beta Binding Protein 3 (LTBP3), a component of the extracellular matrix. Regulates TGFβ activity via extracellular binding. <i>FAM89B</i> (eQTL): Regulates TGFβ signalling by preventing the nuclear translocation of Smad2. <i>NEAT1</i> (eQTL): Long noncoding RNA. Overexpression has been reported both in osteoporosis through osteoclastogenesis stimulation and in osteosarcoma as a stimulator of cell proliferation.	<i>LTBP3</i> : Acromioclavicular dysplasia, Brachyolmia-amelogenesis imperfecta syndrome, Geleophysic dysplasia. <i>BANF1</i> (sQTL) : Nestor-Guillermo progeria syndrome
rs9559805	13	110462733	<i>COL4A2</i> , <i>COL4A2-AS2</i>	Synonymous Variant	<i>COL4A2</i> (eQTL): Alpha-2 chain of basement membrane collagen.	
rs10943838	6	81969549	<i>IBTK</i>	Upstream Gene Variant	<i>TPBG</i> (eQTL): Binds osteoclast-derived collagen triple helix repeat containing 1, a stimulator of osteoblast differentiation.	
rs3922616	16	55008491	<i>IRX5</i>	Intergenic Variant	<i>IRX5</i> : Regulator of mineralization in cranial bones.	
rs9469583	6	33749993	<i>IP6K3</i>	Upstream Gene Variant	<i>ITPR3</i> (eQTL & sQTL): Regulator of osteoclastogenesis; mediates the activation of IRE1α via calcium influx.	<i>IRX5</i> : Hamamy syndrome (with features including skeletal abnormalities, short stature and hearing loss)
rs13192457	6	44887654	<i>SUPT3H</i>	Intron Variant	<i>RUNX2</i> (sQTL): Transcription factor, regulator of osteoblast and chondrocyte differentiation. <i>RUNX2</i> and <i>SUPT3H</i> show synteny across several species and the <i>SUPT3H</i> promoter is a potential regulator of the bone-specific Runx2-P1 promoter.	<i>RUNX2</i> : Metaphyseal dysplasia with maxillary hypoplasia and brachydactyly, cleidocranial dysostosis
rs67284550	16	1480948	<i>PTX4</i>	Downstream Gene Variant	<i>CLCN7</i> (eQTL and sQTL): Regulator of osteoblast function.	<i>CLCN7</i> : Osteopetrosis (several types)
rs4917	3	186619924	<i>AHSG</i>	Missense Variant	<i>AHSG</i> (sQTL): Binds to TGFβ1 and BMPs. Regulator of mineralization through inhibition of calcium phosphate precipitation. <i>AHSG</i> knockout mice exhibit increased bone formation with age and ectopic bone formation in response to TGFβ1/BMP signalling.	
rs7683897	4	87888898	<i>MEPE</i>	Intergenic Variant	<i>MEPE</i> : Candidate gene for otosclerosis. Regulator of osteoclast differentiation and mineralization. <i>SPP1</i> (eQTL): Osteopontin, bone matrix protein implicated in bone remodeling and osteotropic cancers. <i>PKD2</i> (eQTL): <i>PKD2</i> deficiency is associated with reduced expression of Runx2 and impaired osteoblastic differentiation in mouse bone.	
rs8105161	19	41333726	<i>TGFB1</i>	Intron Variant	<i>TGFB1</i> (eQTL and sQTL): Candidate gene for otosclerosis. Essential cytokine for skeletal development and bone remodeling. Regulator of osteoblast and osteoclast cell lineages. Gain-of-function mutations predispose to Camurati-Engelman disease.	<i>TGFB1</i> : Diaphyseal dysplasia