

Multiethnic GWAS Reveals Polygenic Architecture of Earlobe Attachment

John R. Shaffer,^{1,27} Jinxi Li,^{2,27} Myoung Keun Lee,³ Jasmien Roosenboom,³ Ekaterina Orlova,¹ Kaustabh Adhikari,⁴ 23andMe Research Team,⁵ Carla Gallo,⁶ Giovanni Poletti,⁶ Lavinia Schuler-Faccini,⁷ Maria-Cátira Bortolini,⁷ Samuel Canizales-Quinteros,⁸ Francisco Rothhammer,^{9,10} Gabriel Bedoya,¹¹ Rolando González-José,¹² Paige E. Pfeffer,¹³ Christopher A. Wollenschlaeger,¹⁴ Jacqueline T. Hecht,¹⁵ George L. Wehby,¹⁶ Lina M. Moreno,¹⁷ Anan Ding,² Li Jin,^{2,18} Yajun Yang,¹⁸ Jenna C. Carlson,^{1,19} Elizabeth J. Leslie,³ Eleanor Feingold,^{1,19} Mary L. Marazita,^{1,3,20,21} David A. Hinds,⁵ Timothy C. Cox,^{22,23,24} Sijia Wang,^{2,18,*} Andrés Ruiz-Linares,^{4,18,25} and Seth M. Weinberg^{1,3,26,*}

The genetic basis of earlobe attachment has been a matter of debate since the early 20th century, such that geneticists argue both for and against polygenic inheritance. Recent genetic studies have identified a few loci associated with the trait, but large-scale analyses are still lacking. Here, we performed a genome-wide association study of lobe attachment in a multiethnic sample of 74,660 individuals from four cohorts (three with the trait scored by an expert rater and one with the trait self-reported). Meta-analysis of the three expert-rater-scored cohorts revealed six associated loci harboring numerous candidate genes, including *EDAR*, *SP5*, *MRPS22*, *ADGRG6* (*GPR126*), *KIAA1217*, and *PAX9*. The large self-reported 23andMe cohort recapitulated each of these six loci. Moreover, meta-analysis across all four cohorts revealed a total of 49 significant ($p < 5 \times 10^{-8}$) loci. Annotation and enrichment analyses of these 49 loci showed strong evidence of genes involved in ear development and syndromes with auricular phenotypes. RNA sequencing data from both human fetal ear and mouse second branchial arch tissue confirmed that genes located among associated loci showed evidence of expression. These results provide strong evidence for the polygenic nature of earlobe attachment and offer insights into the biological basis of normal and abnormal ear development.

Introduction

Earlobe attachment (MIM: 128900) is often presented as an example of a readily observable Mendelian phenotype in educational materials¹ and continues to be studied as a Mendelian phenotype in the contemporary primary literature (for example, see Ordu et al.²). Yet, as early as 1937, Wiener³ pointed out that earlobe attachment is likely to be a polygenic trait exhibiting a continuous phenotypic distribution. Although earlobe attachment is a neutral morphological trait, understanding its genetic etiology is valuable in that it offers a glimpse into the biological basis

of ear development, improving our understanding of genes potentially involved in developmental defects. Moreover, it serves as an instructive example of simple versus polygenic inheritance in an accessible trait.

Recent genome-wide association studies (GWASs) have investigated earlobe attachment^{4,5} and reported significant associations with variants in *EDAR* (MIM: 604095) and *SP5* (MIM: 609391).⁴ Although promising, these and other suggestive associations have yet to be replicated in independent samples. Of note, ethnic differences in the frequency of lobe attachment are well documented,⁶ suggesting that genetic heterogeneity might underlie the trait

¹Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; ²Chinese Academy of Sciences Key Laboratory of Computational Biology, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China; ³Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA 15219, USA; ⁴Department of Genetics, Evolution and Environment, University College London, London, UK; ⁵23andMe Inc., 899 West Evelyn Avenue, Mountain View, CA 94041, USA; ⁶Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, 430 Cercado de Lima, Peru; ⁷Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre 90040-060, Brazil; ⁸Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, Universidad Nacional Autónoma de México, Instituto Nacional de Medicina Genómica, Mexico City 4510, Mexico; ⁹Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile; ¹⁰Facultad de Medicina, Universidad de Chile, Santiago 8320000, Chile; ¹¹Grupo Genética Molecular GENMOL, Universidad de Antioquia, Medellín 050003, Colombia; ¹²Instituto Patagónico de Ciencias Sociales y Humanas, Centro Científico Tecnológico, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas, Puerto Madryn U9120, Argentina; ¹³Center for Advanced Dental Education, Orthodontics Program, Saint Louis University, St. Louis, MO 63104, USA; ¹⁴Department of Orthodontics, University of Pittsburgh, Pittsburgh, PA 15261, USA; ¹⁵Department of Pediatrics, McGovern Medical School, University of Texas, Houston, TX 77030, USA; ¹⁶Department of Health Management and Policy, University of Iowa, Iowa City, IA 52246, USA; ¹⁷Department of Orthodontics, University of Iowa, Iowa City, IA 52242, USA; ¹⁸Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200433, China; ¹⁹Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261, USA; ²⁰Clinical and Translational Science Institute, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; ²¹Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; ²²Center for Developmental Biology & Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA 98101, USA; ²³Craniofacial Medicine, Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; ²⁴Department of Anatomy & Developmental Biology, Monash University, Clayton, VIC 3800, Australia; ²⁵Laboratory of Biocultural Anthropology, Law, Ethics, and Health, Centre National de la Recherche Scientifique and Etablissement Français du Sang, UMR 7268, Aix-Marseille University, Marseille 13284, France; ²⁶Department of Anthropology, University of Pittsburgh, Pittsburgh, PA 15260, USA

²⁷These authors contributed equally to this work

*Correspondence: wangsijia@pitt.edu (S.W.), smwst46@pitt.edu (S.M.W.)

<https://doi.org/10.1016/j.ajhg.2017.10.001>

© 2017 The Authors. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

and that deciphering its genetic architecture might require trans-ethnic studies. This notion is supported by the fact that one of the two previously reported associations was with a missense *EDAR* variant that is common in Asian and American populations but absent or infrequent in European and African populations.

We propose that large-scale genetic studies of normal human morphological traits can provide insights into the genes and pathways involved in developmental malformations. For example, the external human ear exhibits a highly complex morphology that develops from the first and second branchial arches⁷ and requires precise spatial and temporal coordination of tissue proliferation, fusion, and apoptosis. Disruption of these processes can cause birth defects, such as nonsyndromic microtia (congenital under-development of the external ear [MIM: 600674]), which is a fairly common developmental defect that differs in frequency across populations.⁸ Moreover, because the jaw and associated masticatory musculature are also derived from the branchial arches, a number of craniofacial syndromes involving arch deficiencies, such as hemifacial macrosomia (MIM: 164210)⁹ and Treacher Collins syndrome (MIM: 154500),¹⁰ are characterized by external ear abnormalities. Understanding the genetic factors that contribute to normal structural variation in human ears could provide critical insights into ear morphogenesis, as well as morphogenetic processes in general. In this report, we focus on one aspect of external ear morphology: the lobe.

Specifically, we scanned the genome for variants associated with lobe attachment in 74,660 individuals from four independent and ethnically distinct samples. These included (1) European American ($n = 1,791$), Latin American ($n = 5,062$), and Chinese ($n = 2,857$) cohorts in which lobe attachment was scored as a tripartite (free, partially attached, or fully attached) phenotype by an objective expert rater and (2) a European-ancestry cohort ($n = 64,950$) comprising research participants who were customers of 23andMe, a personal genomics company, and self-reported lobe attachment as a binary (free or attached) phenotype. We performed two nested meta-analyses to combine results across the three expert-rater-scored cohorts and subsequently across all four cohorts. All participants were genotyped on an Illumina genome-wide array and imputed to the 1000 Genomes Project reference panel, and a GWAS was performed after adjustment for necessary covariates and principal components (PCs) of ancestry (see [Table S1](#)). Manhattan and quantile-quantile plots for the GWAS results in individual cohorts are shown in [Figure S1](#).

Subjects and Methods

Study Design

We used a nested genome-wide meta-analysis approach to identify and replicate variants associated with ear attachment. A GWAS

was performed separately in four independent cohorts representing populations with distinct ancestries. For three of these cohorts (European American, Latin American, and Chinese), lobe attachment was scored by objective raters as a tripartite phenotype. For the fourth cohort (European-ancestry individuals in the 23andMe sample), lobe attachment was self-reported as a dichotomous phenotype. Because of differences in phenotype definition (tripartite versus dichotomous) and collection method (rater-scored versus self-reported), we performed two genome-wide meta-analyses of (1) the three rater-scored cohorts and (2) all four cohorts.

Data Access

Summary statistics for the 10,000 most significant SNPs from the meta-analyses are provided in [Table S2](#). The individual-level genetic data for the expert-rated European American cohort are available from dbGaP: phs000949.v1.p1. Full summary statistics for all SNPs in the European American, Latin American, and Chinese cohorts are available upon request. Summary statistics for the 23andMe cohort can be requested directly from 23andMe and will be made available to qualified researchers under the terms of a data-transfer agreement with 23andMe to protect the privacy of the participants. Please contact David Hinds for more information and to apply to access the data.

Recruitment and Phenotyping

The European American cohort consisted of 1,791 unrelated European-ancestry individuals aged 3–49 years and recruited from Pittsburgh, Seattle, Houston, and Iowa City as part of a 3D Facial Norms Project.^{11,12} Participants were screened for conditions affecting craniofacial morphology, including a history of congenital malformations, trauma, and surgery. The Latin American cohort comprised 5,062 participants from the Consortium for the Analysis of the Diversity and Evolution of Latin America (CANDELA)¹³ recruited from Brazil, Chile, Colombia, Mexico, and Peru. The Chinese cohort comprised 2,857 ethnic Han Chinese participants recruited from Taizhou in the Jiangsu Province of China as part of the Taizhou Longitudinal Study.¹⁴ All participants provided informed consent, and all study protocols were approved by the institutional review boards of the pertinent research institutions.

In the European American, Latin American, and Chinese cohorts, earlobes were classified as free, partially attached, or attached. An individual was considered to possess attached earlobes if at least one ear was rated as attached. For the European American cohort, two independent observers examined the ears of participants from 3D craniofacial surface images captured by digital stereophotogrammetry. For the Latin American cohort, the same rater scored lobe attachment according to digital photography (Nikon) of the right side (45° angle) and front of the face. For the Chinese cohort, two independent observers scored lobe attachment from 2D digital photography (Canon EOS 600D) of both sides at a 45° angle.

An additional cohort, composed of research participants from the consumer base of 23andMe as previously described,^{5,15} was included. For this study, the 23andMe sample comprised 64,950 unrelated individuals of European ancestry, and data on ear attachment were collected via self-reporting in online surveys using a dichotomous (attached or free) phenotype definition. Example imagery of attached and free earlobes was provided to participants for reference. 23andMe research participants provided informed consent and answered surveys online according

to a human subjects protocol approved by Ethical and Independent Review Services, an external institutional review board.

Genotyping, Quality Control, and Imputation

Genotyping was performed separately in the four cohorts. For the European American cohort, DNA was extracted from saliva samples and genotyped along with 72 HapMap control samples for 964,193 SNPs on the Illumina HumanOmniExpress+Exome v.1.2 array by the Center for Inherited Disease Research. Genetic data cleaning and quality control have been described in detail previously.¹¹ In brief, samples were interrogated for sex, chromosomal aberrations, relatedness, genotype call rate, and batch effects. SNPs were interrogated for call rate, discordance among 70 duplicate samples, Mendelian errors among HapMap controls (parent-offspring trios), deviations from Hardy-Weinberg equilibrium, and sex differences in allele frequencies and heterozygosity. For the Latin American cohort, DNA was extracted from blood samples obtained by a certified phlebotomist and genotyped for 730,525 SNPs on the Illumina HumanOmniExpress array. Quality filters included genotyping call rates per participant and per SNP and minor allele frequency (MAF). Because of admixture within the sample, filters for Hardy-Weinberg equilibrium were not implemented. For the Chinese cohort, DNA was extracted from peripheral-blood samples and genotyped for 887,270 SNPs on the Illumina HumanOmniZhonghua-8 array. SNP-level quality filters were applied for missing call rate, MAF, deviation of genotype frequencies from Hardy-Weinberg equilibrium, and technical filters (shown in Table S3).

For the 23andMe cohort, DNA extraction and genotyping were performed on saliva samples by Laboratory Corporation of America clinical laboratories certified by the Clinical Laboratory Improvement Amendments and accredited by the College of American Pathologists. Samples from this cohort were genotyped on one of four Illumina platforms: two versions of the HumanHap550 chip plus 25,000 custom SNPs, the HumanOmniExpress plus custom content to increase overlap with the HumanHap550 platforms, or a fully custom-designed array. Participants with samples that failed to reach 98.5% call rates were re-contacted for a replacement sample and were re-analyzed. Quality filters were applied for genotype call rate, MAF, Hardy-Weinberg equilibrium, and artifact effects by date.

For all studies, unobserved variants were imputed with haplotypes from the 1000 Genomes Project as the reference (phase 1 for the Latin American and 23andMe cohorts and phase 3 for the European American and Chinese cohorts). Pre-phasing (using SHAPEIT¹⁶ for the rater-scored cohorts and the company's own tool according to the Beagle¹⁷ algorithm for the 23andMe cohort) was performed before imputation. Imputation was performed with IMPUTE2¹⁸ for rater-scored cohorts and with Minimac2¹⁹ for the 23andMe cohort. For the European and Latin American cohorts, masked variant analysis, in which genotyped SNPs were imputed for assessment of imputation quality, indicated high accuracy of imputation. Table S4 shows imputation quality filters.

Population Structure

To assess population structure, we performed principal-component analysis (PCA) within each cohort by using subsets of uncorrelated SNPs. Plots of the top PCs of ancestry for the European American, Chinese, and 23andMe cohorts are shown in Figure S5. The complex population structure in the Latin American sample was the focus of a previous investigation.¹³ On the

basis of scatterplots of the PCs and scree plots of the eigenvalues, we determined that adjustment for 4, 5, 0, and 5 PCs was necessary for the European American, Latin American, Chinese, and 23andMe cohorts, respectively. In the 23andMe cohort, we compared phased genomic segments with reference data across 31 populations to assign the mostly likely ancestry source of each segment.²⁰ We aggregated local-ancestry assignments to determine the overall proportions of ancestry of each individual. Of the 68,965 consenting 23andMe participants with available phenotypes, 64,950 individuals were determined to have >97.5% European ancestry and were included in this study. In general, genetically determined European ancestry closely matched the self-reported ancestry of the 23andMe participants.

Association Analyses

Earlobe attachment was analyzed as a semiquantitative phenotype (coded 0, 1, or 2 for free, partially attached, or attached earlobes, respectively) separately in the European American, Latin American, and Chinese cohorts. We tested genetic association while adjusting for necessary covariates (such as age, sex, height, and body mass index; see Table S1) and PCs of ancestry by using linear regression under the additive genetic model. For the analysis of the X chromosome, we coded genotypes as 0, 1, or 2 per the additive genetic model for females and as 0 or 2 for males in order to maintain the same scale between sexes. GWAS results for each study were combined via inverse-variance-weighted meta-analysis.

For the 23andMe cohort, earlobe attachment was analyzed as a binary phenotype (coded 0 for free or 1 for attached). We used logistic regression including adjustment for age, sex, genotyping platform, and the top five PCs to test for genetic association under the additive genetic model. Results across all four cohorts were combined via Stouffer's method²¹ of inverse-weighted meta-analysis (based on p values, direction of effect, and sample size). This method of meta-analysis was chosen because it is robust to differences in the scale of the effect estimates among the expert-rater-scored cohorts and between the expert-rater-scored and self-reported 23andMe cohort as a result of differences in phenotype assessments. Compared with meta-analysis methods that use effect sizes and standard errors, Stouffer's method results in only a small loss of efficiency,²¹ which is outweighed by its robustness to known and unknown phenotype differences across cohorts, and it does not require an assumption that effects are the same across cohorts. We used the binomial test (i.e., sign test) to model the consistency of direction of expert-rater-scored effects with the 23andMe cohort.

Functional Annotation

We used HaploReg²² to query the lead SNP (i.e., the SNP with the smallest p value) at each associated locus in order to extract evidence of functional variation (promotor and enhancer histone marks, DNase hypersensitivity, expression quantitative trait locus [eQTL] information) for all SNPs in linkage disequilibrium (LD; $r^2 > 0.8$) with the lead SNP. 351 genes of interest were defined on the basis of a physical proximity of 500 kb to the lead SNP at each locus. These genes were queried in a number of online databases. We used Mouse Genome Informatics (MGI)²³ to annotate expression in relevant tissues and phenotypic consequences and used the VISTA Enhancer Browser²⁴ to annotate active enhancer elements in relevant tissues. We used OMIM, PubMed, DECIPHER,²⁵ and ClinVar²⁶ to annotate human phenotypic

information. We performed genomic enrichment analyses by using the Genomic Regions Enrichment of Annotations Tool (GREAT).²⁷

Tissue Collection, RNA Isolation, and Sequencing

Branchial arch two tissue was dissected from wild-type embryonic day 10.5 mice, as well as equivalently staged homozygous *sbse* and *dmbo* embryos, and snap frozen on dry ice. RNA was isolated with the QIAGEN RNeasy Mini Kit (74104, QIAGEN), and its quality and concentration were assessed with an Agilent 2200 TapeStation system at the Genomics Core of the Fred Hutchinson Cancer Research Center. All samples had an RNA integrity number greater than 8. For each genotype, RNA was pooled from three male pups, and a total of 1 µg RNA of each genotype was sent to the Genomic Services Lab at the Hudson Alpha Institute for Biotechnology for preparation of indexed directional libraries and ribosomal reduction RNA sequencing (RNA-seq). Samples were paired-end sequenced with 250 million reads on an Illumina HiSeq v.4 PE100. The Cufflinks software suite was used for transcriptome assembly and differential expression analysis. We used the DESeq2 package to weight expression with count data and the Integrative Genomics Viewer to visualize sequences.^{28,29}

For the analyses of human fetal RNA, ear tissue was obtained (after informed parental consent was provided) from material collected by the Birth Defects Research Laboratory (under approval by the institutional review board of the University of Washington). The gestational age of conceptuses, reported as the number of days after fertilization, was estimated from fetal foot length. The tissue was snap frozen, and RNA was processed as described above.

Results

Earlobe-Attachment Loci Observed in Trans-ethnic Meta-analysis

Rates of lobe attachment differed across the cohorts (Table S1), which was expected given the known differences across ethnic groups. Meta-analysis of the GWAS results from the three expert-rater-scored cohorts (see Figures 1A and 1B) yielded six loci that were significantly (i.e., $p < 5 \times 10^{-8}$) associated with earlobe attachment (Table 1; Figure 2): 2q13, 2q31.1, 3q23, 6q24.2, 10p12.2, and 14q13.1. These loci included the genes *EDAR* (2q13; lead SNP: rs3827760), *SP5* (2q31.1; lead SNP: rs6756973), *MRPS22* (MIM: 605810; 3q23; lead SNP: rs9866054), *ADGRG6* (*LOC153910* or *GPR126* [MIM: 612243]; 6q24.2; lead SNP: rs58122955), *KIAA1217* (MIM: 617367; 10p12.2; lead SNP: rs7096127), and *PAX9* (MIM: 167416; 14q13.1; lead SNP: rs1950357). One missense variant, rs3827760 in *EDAR*, was observed among the six lead SNPs and other variants in high LD ($r^2 > 0.8$) with the lead SNPs. The others were either intronic or intergenic, and as summarized by HaploReg,²² several showed evidence of DNase hypersensitivity and histone marks indicative of enhancer or promoter regulatory elements in skin and other cell types. None of the lead SNPs or variants in high LD with the lead SNPs were known eQTLs.

Each of these six loci also showed significant evidence of association with earlobe attachment in the self-rated 23andMe cohort (Table 1). Moreover, in the meta-analysis across all four cohorts (Figures 1C and 1D), a total of 49 significant loci were observed (Figure S2; Table S5), which included the six loci observed in the meta-analysis of expert-rater-scored cohorts. Of the 49 loci, 15 showed significant associations (i.e., $p < 5 \times 10^{-8}$) in the 23andMe sample and replication-level p values of < 0.001 in at least one additional cohort or in the meta-analysis across the three expert-rater-scored cohorts. Another 24 loci were driven primarily by significant associations observed in the 23andMe cohort; 14 of the 24 showed consistent directions of effects between 23andMe and all three expert-rater-scored cohorts (sign test $p = 1.29 \times 10^{-7}$), and 23 of 24 showed consistent directions of effects between 23andMe and at least two of the three expert-rater-scored cohorts (sign test $p = 1.49 \times 10^{-6}$). The remaining 4 of the 49 loci were significant only in the meta-analysis across all four cohorts but not in any individual cohort.

Functional Annotation

351 genes were located within 500 kb of the lead SNP across the 49 associated loci. Using public databases (see Subjects and Methods), we queried these 351 genes for documented expression and activity of enhancer elements in relevant tissues, as well as any known roles associated with ear phenotypes in human disorders or mouse models. Table S6 enumerates the genes at associated loci and the evidence substantiating their biological roles related to ear morphology. In total, 71 (20%) of the 351 genes were expressed in relevant tissues in mice (12 [3%] in the outer ear, 58 [17%] in the inner ear, and 39 [11%] in the branchial arches). Likewise, 21 (6%) of the genes were implicated in human syndromes manifesting with ear phenotypes, 22 (6%) were implicated in ear phenotypes in mouse models, and 16 (5%) were flanking active enhancer elements in relevant tissues. Overall, several plausible candidate genes were identified across the 49 associated loci identified via meta-analysis.

We performed genomic enrichment analysis by using GREAT²⁷ to determine whether the gene set comprising the two nearest genes across the 49 associated loci (indicated by the position of the lead SNP) was enriched with relevant annotations across several ontologies. Of most relevance, we observed significant enrichment of over 16-fold for the human phenotype annotations “microtia,” “aplasia/hypoplasia of the external ear,” and “aplasia/hypoplasia of the ear” (p values $< 1.5 \times 10^{-5}$ for all), as well as 5-fold enrichment for “low-set ears” and “abnormal location of ears” (p values < 0.0001 for both). Significant enrichment was also observed for several embryonic-development- and morphogenesis-related Gene Ontology biological processes, many mouse morphology (including ear) terms, and mouse expression in many relevant tissues (notably branchial arch and ear). Detailed enrichment results are shown in Figure S3.

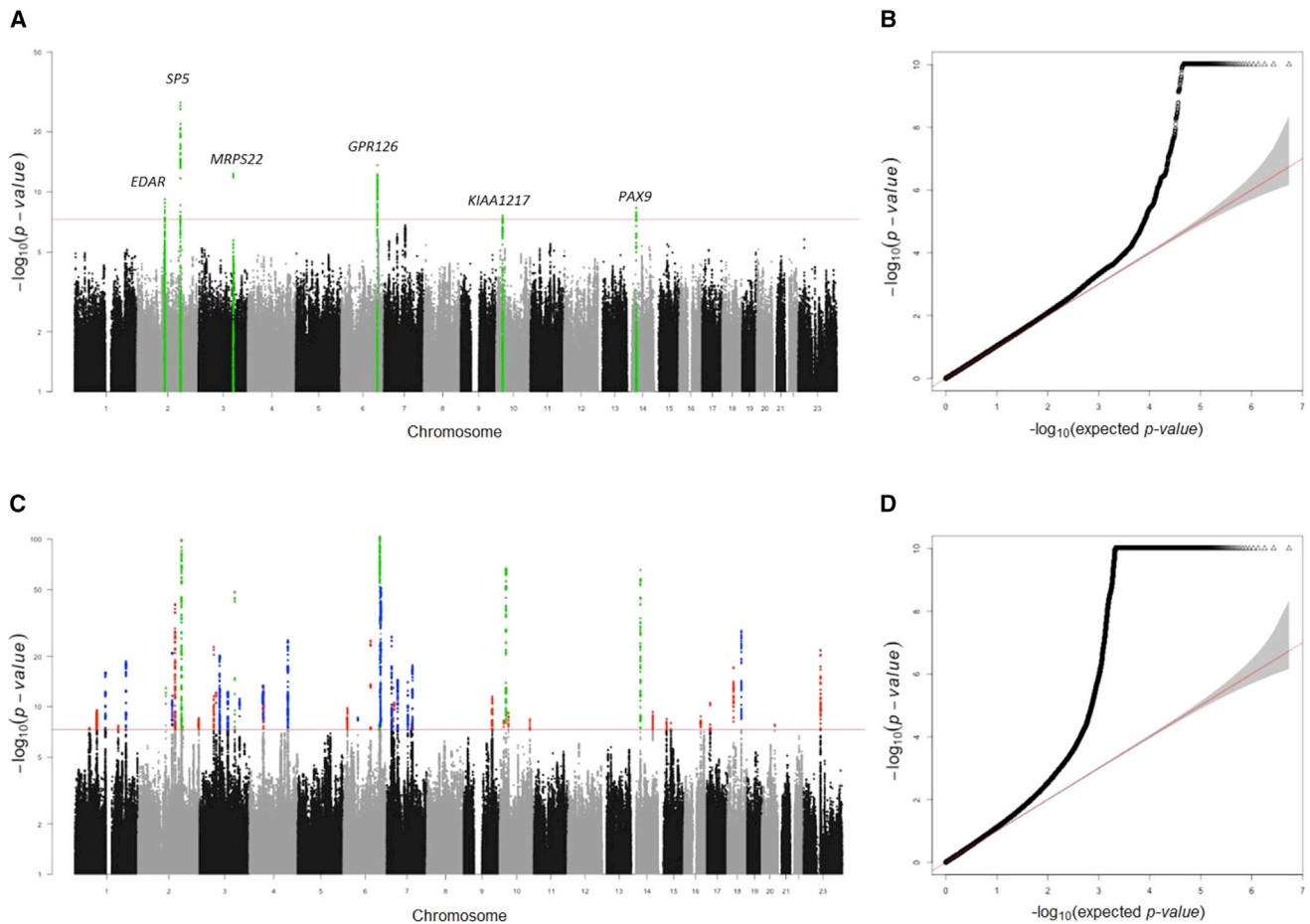


Figure 1. Genome-wide Scans

(Left) Manhattan plots showing the $-\log_{10}$ -transformed p values (y axis) by physical genomic position (x axis) for each SNP in (A) the meta-analysis of the three rater-scored cohorts and (C) the meta-analysis of all four cohorts. The horizontal line represents the threshold for genome-wide significance ($p < 5 \times 10^{-8}$). (A) Six significant loci (green) were observed, and genes near the lead SNP in each locus are annotated. (C) 49 associated loci were observed: the same six loci in (A) are shown in green, and these reached genome-wide significance in more than one cohort; the 15 loci in blue showed genome-wide significance in one cohort and replication-level significance ($p < 0.001$) in at least one additional cohort or the meta-analysis of expert-rater-scored cohorts; and the 28 loci in red were observed via meta-analysis.

(Right) Quantile-quantile plots showing the observed distribution of $-\log_{10}$ -transformed p values (y axis) against the expected distribution (x axis) under the null hypothesis of no association (diagonal line) for (B) the meta-analysis of the three rater-scored cohorts (genomic inflation factor = 1.066) and (D) the meta-analysis of all four cohorts (genomic inflation factor = 1.563). The presence of significantly associated loci is indicated by the deviation of observed p values from the tail of the null distribution, as shown by points above the diagonal in the upper right of the plots.

Expression Experiments

To confirm that genes located among associated loci are expressed in relevant tissues during development, we used RNA-seq to measure the expression of 174 genes located within 250 kb of one of the 49 lead SNPs. Gene expression was measured in two human fetal ears (at days 57 and 59 of development, when external ear structures are present but still developing⁷) and in mouse embryonic day 10.5 branchial arch tissue isolated from two mutants, *short body-short ear (sbse)* and *dumbo (dmbo)*, and sex- and background-matched wild-type (C57BL/6) controls. These mutants present with microtia (*sbse*) or low-set ears with “lobe duplication” (*dmbo*).³⁰ For 4 of the 49 loci, no genes were located within 250 kb of the lead SNP. The majority of the remaining 45 loci

had one or more genes that were robustly expressed in these relevant embryonic tissues (Figure S4). Human fetal ear tissue showed similar expression levels at days 57 and 59, and the greatest expression was observed on both days for *PRRX1* (MIM: 167420; 1q24.2), a homeobox gene with relevant biology (see Discussion). Some of the genes located at the six loci (especially 2q31.1 and 2q13; Figure 3) identified in the GWAS of rater-scored cohorts and recapitulated in the 23andMe cohort were among the top ranking genes (of the 174 genes tested) in terms of expression in humans or differential expression between mutant and wild-type mice. For example, mouse orthologs of both *SP5* (2q31.1) and *EDAR* (2q13) exhibited higher expression in *dmbo* mutant mice than in wild-type controls (\log_2 fold change > 0.5), and orthologs of *GADI*

Table 1. Evidence of Association for the Lead SNP in Each Significant ($p < 5 \times 10^{-8}$) Locus Nominated in the Meta-analysis across the Rater-Scored Cohorts

	2q13	2q31.1	3q23	6q24.2	10p12.2	14q13.1
Gene candidate(s)	<i>EDAR</i>	<i>SP5</i>	<i>MRPS22, FOXL2</i>	<i>ADGRG6 (GPR126)</i>	<i>KIAA1217, ARHGAP21</i>	<i>PAX9, NKX2-8</i>
Lead SNP	rs3827760	rs6756973	rs9866054	rs58122955	rs7096127	rs1950357
Base position	109,513,601	171,542,573	138,997,688	142,921,276	24,506,439	37,209,698
Functional position	missense	intronic	intergenic	intronic	intronic	intronic
Data source	genotyped	imputed	genotyped	imputed	imputed	imputed
Minor/major alleles	G/A	C/G	A/G	A/G	C/T	C/A
European American						
MAF	0.015	0.413	0.257	0.223	0.451	0.386
Beta	0.028	0.073	-0.041	0.119	-0.034	0.086
SE	0.081	0.020	0.022	0.023	0.019	0.020
p value	0.733	1.84×10^{-4}	0.059	2.78×10^{-7}	0.077	2.40×10^{-5}
Latin American						
MAF	0.404	0.661	0.550	0.185	0.490	0.359
Beta	0.062	0.088	-0.025	0.067	-0.043	0.027
SE	0.012	0.011	0.011	0.013	0.010	0.010
p value	7.67×10^{-8}	7.29×10^{-16}	0.019	2.34×10^{-7}	2.06×10^{-5}	9.41×10^{-3}
Chinese						
MAF	0.946	0.419	0.429	0.267	0.303	0.295
Beta	0.129	0.101	-0.128	0.051	-0.052	0.065
SE	0.032	0.015	0.015	0.017	0.016	0.016
p value	5.66×10^{-5}	9.84×10^{-12}	2.60×10^{-18}	0.002	0.001	8.04×10^{-5}
Meta-analysis for Rater-Scored Cohorts						
p value	6.65×10^{-10}	1.13×10^{-28}	5.27×10^{-13}	2.49×10^{-14}	2.44×10^{-8}	4.82×10^{-9}
23andMe						
MAF	0.007	0.410	0.249	0.238	0.451	0.389
OR	1.490	1.270	0.837	1.335	0.813	1.237
CI	(1.299, 1.709)	(1.238, 1.303)	(0.813, 0.862)	(1.296, 1.375)	(0.793, 0.834)	(1.206, 1.269)
p value	2.04×10^{-8}	4.90×10^{-76}	4.42×10^{-33}	3.31×10^{-87}	8.48×10^{-59}	8.26×10^{-59}
Meta-analysis for All Cohorts						
p value	1.16×10^{-13}	1.65×10^{-99}	4.78×10^{-43}	7.64×10^{-100}	1.56×10^{-65}	2.85×10^{-66}

Abbreviations are as follows: MAF, minor allele frequency; SE, standard error of the beta-coefficient; OR, odds ratio; and CI, 95% confidence interval of the odds ratio.

(MIM: 605363; 2q31.1), *MYO3B* (MIM: 610040; 2q31.1), and *SH3RF3* (2q13) showed reduced expression in the mutant mice (\log_2 fold change < -0.5). Thus, separate genes at the 2q13 and 2q31.1 loci were both up- and down-regulated (see [Figures 2A](#) and [2B](#) for genetic association results for these loci). In a comparison of *sbse* mutants and wild-type mice, orthologs of *ERICH2* (2q31.1) and *NKX2-1* (2q13.1; MIM: 600635) were among the top over-expressed genes in *sbse* mutants among the 174 tested, and orthologs of *GADI* (2q31.1) and *SH3RF3* (2q13) again exhibited reduced expression in the mutants (see [Figures](#)

[2A](#), [2B](#), and [2F](#) for genetic association results for these genes). Together, these results confirm the expression of genes located at associated loci in relevant human tissue and suggest that expression differs by ear phenotype for some genes at associated loci in a mouse model.

Discussion

We have performed the largest genome-wide study to date of earlobe attachment in 74,660 individuals from three

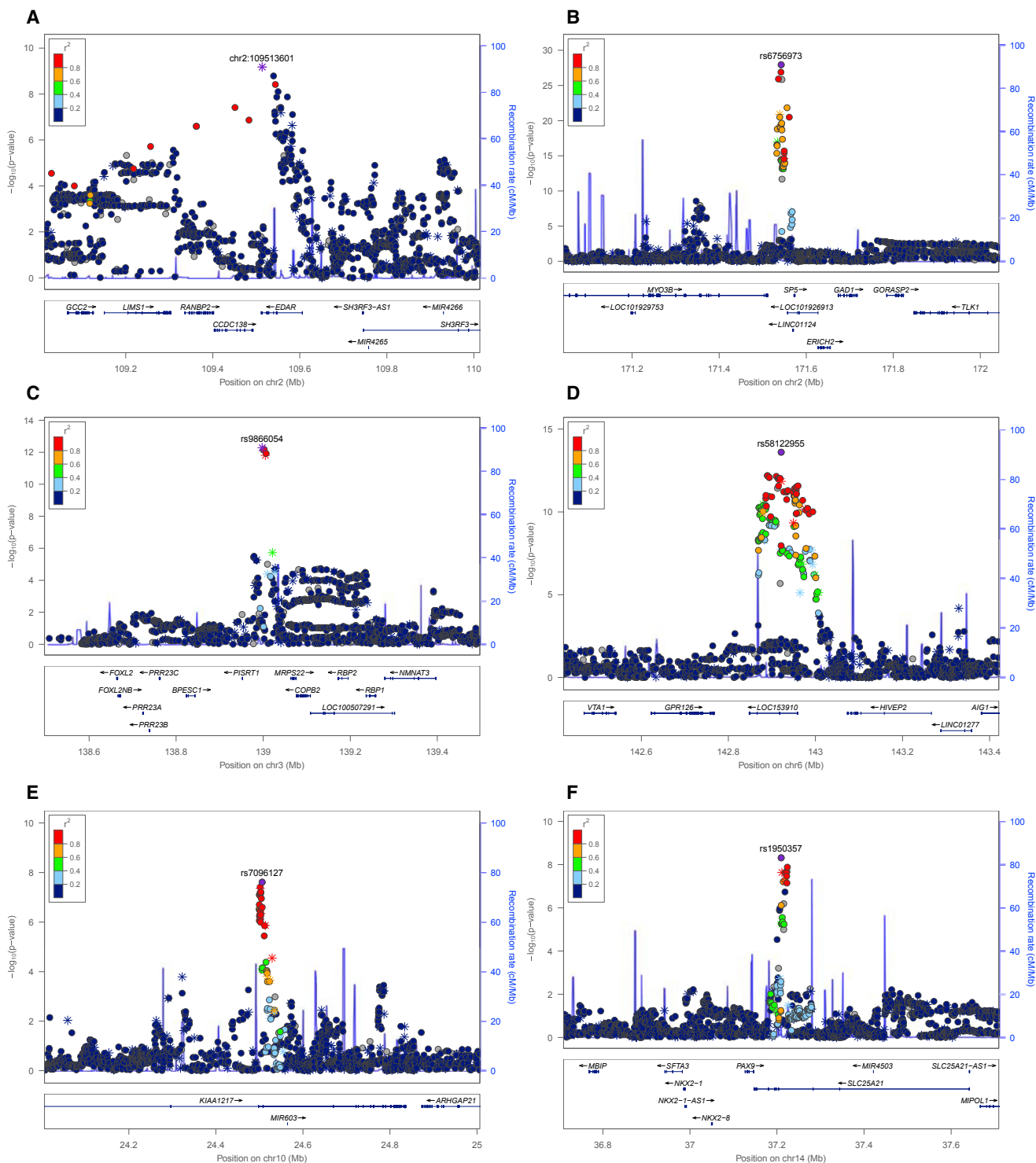


Figure 2. Regional Association Plots Showing Significant Associations Observed in the Meta-analysis of the Three Rater-Scored Cohorts

Regional plots near (A) *EDAR*, (B) *SP5*, (C) *MRPS22*, (D) *ADGRG6* (*GPR126*), (E) *KIAA1217*, and (F) *PAX9* show $-\log_{10}$ -transformed p values (left y axis) by physical position (x axis). Shading denotes the LD (r^2) between each SNP and the lead SNP (purple). The blue overlay represents the recombination rate (right y axis). Gene positions are indicated under each plot.

ancestry groups. All six significant loci observed in the meta-analysis using the tripartite rater-scored phenotype were also significantly associated with earlobe attachment in the 23andMe cohort using the self-reported dichot-

omous phenotype. Furthermore, four of the six loci were the top ranking (by p value) associations observed in the 23andMe cohort (the other two loci ranked 7th and 42nd in the 23andMe GWAS). Thus, the GWAS of the 23andMe

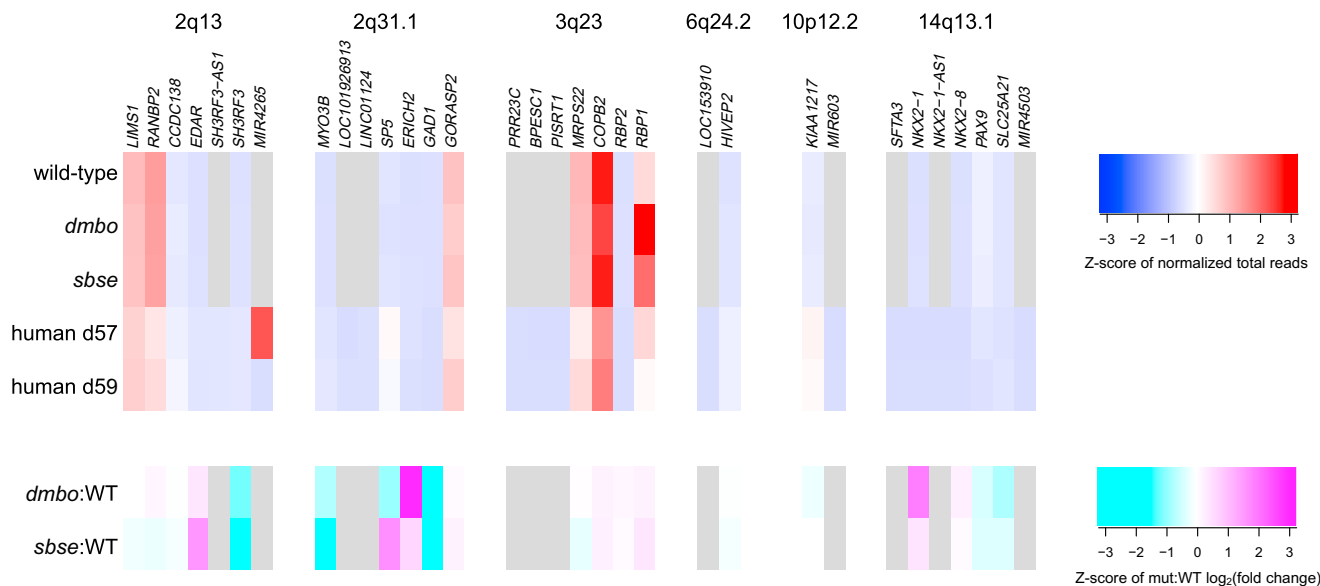


Figure 3. Heatmaps of Gene Expression

Heatmaps of gene expression in second branchial arch tissue from embryonic mice (wild-type mice and *dmbo* and *sbse* mutants) and fetal human pinna (at days 57 and 59) and for genes at the six loci observed in the meta-analysis of rater-scored cohorts; normalized total read counts are on the top, and fold changes between mutant mice and wild-type mice are on the bottom. The shading scale is shown for Z scores of expression and fold change. Genes for which expression data were not measured are shown in gray.

cohort recapitulated all of the major findings from the expert-rater-scored cohorts. Furthermore, four of the six loci contained genes showing notable expression (e.g., a high read count in humans and/or differential expression in mice) in relevant embryonic tissues.

Among the six associated loci from the rater-scored meta-analysis, two (2q13 at *EDAR* and 2q31.1 at *SP5*) have been previously identified in the Latin American cohort,⁴ two (6q24.2 near *ADGRG6* [*GPR126*] and 10p12.2 at *KIAA1217*) have previously shown suggestive evidence of association,^{4,5} and one (3q23 near *MRPS22*) has been previously implicated in earlobe size. In contrast, the association at 14q13.1 near *PAX9* did not show evidence of genetic association in previous studies.^{4,5} These or nearby genes have known biological functions that indicate plausible roles in determining ear morphology. For example, *EDAR* encodes a cell-surface receptor important for the development of ectodermal tissues, including skin. The missense SNP rs3827760 (c.1109T>C [p.Val370Ala]) affects protein activity^{31,32} and is associated with variation in tooth morphology, hair, sweat gland density, and facial morphology in Asians.^{33–41} This variant was the top SNP in the *EDAR* region in the meta-analyses, and its frequency differs dramatically across populations (e.g., G allele frequency of <1% in Europeans, 39% in Latin Americans, and >90% in Han Chinese in the 1000 Genomes Project). Adhikari et al. have shown that *Edar* is expressed in the mouse pinna and that, compared with wild-type mice, mouse mutants with loss of *Edar* function exhibit reduced ear protrusion and length, as well as a different shape.⁴

SP5 encodes a transcription factor involved in the regulation of Wnt-mediated beta catenin signaling, which in turn is critical for multiple aspects of development, including that of the inner ear.⁴² *ADGRG6* (*GPR126*) encodes a G-protein-coupled receptor whose disruption, via either mutation⁴³ or morpholino,⁴⁴ causes a swollen inner-ear phenotype in zebrafish. This locus was shown to be associated with earlobe size, a related phenotype, in the Latin American cohort.⁴ *KIAA1217* is not known to be involved in ear development, but the associated variants are downstream of *ARHGAP21* (MIM: 609870), variants in which have been associated with mandibular prognathism,⁴⁵ a branchial arch defect.

The association at 3q23 occurs nearest *MRPS22*, which is implicated in a Cornelia de Lange-like phenotype including ear and skin dysmorphic features,⁴⁶ and is upstream of *FOXL2* (MIM: 605597), encoding a craniofacial transcription factor. This locus was previously implicated in earlobe size in the Latin American cohort.⁴ The 14q13.3 association is near *PAX9*, encoding a transcription factor involved in mouth and tooth development, as well as *NKX2-8* (MIM: 603245), a homeobox candidate gene for microtia.⁴⁷

In addition to the six loci identified via meta-analysis of rater-scored cohorts and recapitulated in 23andMe, meta-analysis across all four cohorts yielded 43 additional significant associations, driven primarily by the large 23andMe cohort. Among the more promising candidates were genes implicated in both human and mouse ear phenotypes: *TBX15* (MIM: 604127), *PRRX1*, and *ZEB2* (MIM: 605802). *TBX15* (the gene containing the lead SNP at this locus) is a transcription factor responsible for Cousin syndrome

(MIM: 260660), in which ears are low-set and posteriorly rotated.⁴⁸ In mouse, mutations in *TBX15* cause abnormal ear position and the “droopy ear” phenotype,⁴⁹ which resembles that seen in the *dmbo* mutants used in the expression analysis in this study. This locus was associated with ear phenotypes including antitragus size and folding of antihelix in the Latin American sample.⁴ *PRRX1* (40 kb upstream of the lead SNP, which also showed the greatest expression in fetal human ear at day 57) is a homeobox gene implicated in Agnathia-otocephaly (characterized by severe malformations of the mouth, jaw, and ear; MIM: 202650),⁵⁰ and copy number variants affecting *PRRX1* have been observed in several patients with dysmorphic ear phenotypes. Moreover, *PRRX1* is expressed in the inner, middle, and outer ear and first and second branchial arches in mouse, and mutations cause lower ear position and abnormal Meckel’s cartilage.⁵¹ *ZEB2* (500kb downstream of lead SNP) is a homeobox gene implicated in Mowat-Wilson syndrome (MIM: 235730),⁵² in which the ears are cupped and the earlobes are upturned with central depression. In mouse, *ZEB2* is expressed in the inner and middle ear, and first brachial arch, and knock-outs lead to Mowat-Wilson-like features⁵³ or lack of first branchial arch during embryogenesis.⁵⁴ Other notable candidates include the growth factor *BMP5* (MIM: 112265), and homeobox transcription factors *DLX5* (MIM: 600028) and *DLX6* (MIM: 600030), which are all expressed in ear and related tissues and are all implicated in ear phenotypes in mice.^{55,56}

The myriad associations overwhelmingly demonstrate the polygenic nature of earlobe attachment, standing in contrast to previous notions regarding its Mendelian nature, which have been perpetuated through the primary literature and educational materials for nearly a century. In fact, the large number of confidently identified loci is on par with many continuous anthropometric traits such as height and body composition. Moreover, earlobe attachment is correlated with other aspects of lobe morphology, including earlobe size, so the overlap in associated loci with the previous study of lobe size⁴ is unsurprising. The effect sizes of variants observed in this study are fairly large for individual SNPs (resulting in a difference of up to 0.2 phenotype standard deviations per allele in the rater-scored cohorts and up to an odds ratio of 1.5 in the 23andMe cohort), although they are not large enough to cause Mendelian segregation. Moreover, differences in allele frequencies across ancestry groups were observed for some associated SNPs, which could explain part of the ethnic heterogeneity observed for earlobe attachment. Specifically, 16 of the lead SNPs of the 49 associated loci, notably the *EDAR* variant rs3827760 (which had a MAF difference of 0.93 between the European American and Chinese cohorts), showed MAF differences greater than 0.2 across the ancestry groups.

Consistent with a polygenic trait in a well-powered GWAS, we observed evidence of genomic inflation (e.g., genomic inflation factors greater than 1.0) separately in

the meta-analyses and the 23andMe cohort. This occurs because the genomic inflation factor, although designed to be calculated from a set of independent null markers, is instead calculated in GWASs from the set of all of the SNPs tested, including truly associated SNPs in LD with causal variants. For polygenic traits (for which there are multiple truly associated loci) in well-powered studies (where small p values are obtained even for SNPs weakly correlated with true causal alleles), the lambda is expected to be greater than 1.⁵⁷ We argue that this inflation is not due to population stratification because association models were adjusted for genetic ancestry estimated from the genetic data in each cohort. Moreover, inflation was not observed in the individual expert-rater-scored cohorts or in a subset of 2,000 participants from the 23andMe cohort, as would be expected if population stratification had caused epidemiological confounding. Instead, the inflation observed for earlobe attachment, which is similar to that observed for highly polygenic traits such as height, is expected given the contributions of numerous associated loci each tagged by many correlated SNPs.

Strengths of this work include the high-quality phenotyping based on digital imagery in the rater-scored cohorts, the inclusion of cohorts from different ancestry groups, the large sample size, and the method of meta-analysis, which was chosen to be robust to phenotype differences across the cohorts. Although phenotype data were collected via self-report in the 23andMe cohort, the fact that associations observed in the rater-scored cohorts were also identified in the 23andMe cohort suggests that the large sample size of the 23andMe cohort counterbalances noise (if any) as a result of the method of data collection. Despite these strengths, and because of the differences in phenotype assessment across cohorts, this study was limited by the fact that, within our testing framework, we were not able to directly test the heterogeneity of effects among cohorts. In addition, the large sample size of the 23andMe cohort, which benefited the statistical power of study, most likely had an outsized effect on the meta-analysis across all four cohorts. For this reason, we also reported results for meta-analysis across the three expert-rater-scored cohorts.

In conclusion, we have identified 49 associations with earlobe attachment, including 21 loci meeting the standard of genome-wide discovery ($p < 5 \times 10^{-8}$) plus independent replication ($p < 0.001$) and 28 loci showing evidence of discovery only (i.e., without independent replication), via meta-analysis. These genes provide insight into the complex biology of ear development. The fact that we observed several associated genes in which pathogenic variants are known to cause human syndromes with ear phenotypes is consistent with our hypothesis that whereas deleterious variants in genes can cause congenital defects and Mendelian conditions, regulatory variants in the same genes can influence normal phenotypic variation. Ultimately, understanding

the genetics of normal human morphological traits can provide insights into the genes and pathways involved in developmental malformations.

Supplemental Data

Supplemental Data include four figures and six tables and can be found with this article online at <https://doi.org/10.1016/j.ajhg.2017.10.001>.

Consortia

Members of the 23andMe Research Team include Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson.

Conflicts of Interest

D.H. is employed by and owns stock or stock options in 23andMe.

Acknowledgments

We thank the participants of the 3D Facial Norms Project, CANDELA, and the Taizhou Longitudinal Study for their contributions toward this effort. We thank the research participants and employees of 23andMe for making this work possible. We thank Daniela Luquetti, Esra Camci, and Jessica Rosin (Seattle Children's Research Institute [SCRI]) and Mei Deng and Ian Glass (Birth Defects Research Laboratory, University of Washington) for contributions to tissue collection and RNA preparation for sequencing and Andrew Timms (SCRI) for initial bioinformatics processing. This work was funded by the following grants and contracts: Chinese Academy of Sciences Strategic Priority Research Program grant XDB13041000 (S.W.); National Natural Science Foundation of China grant 91631307 (S.W.), National Institute of Dental and Craniofacial Research grants and contracts U01-DE020078 (S.M.W. and M.L.M.), U01-DE020057 (M.L.M. and Jeffery C. Murray), R01-DE016148 (M.L.M. and S.M.W.), R00-DE02560 (E.J.L.), R01-DE027023 (S.M.W. and J.R.S.), and HHSN2682012000081 (Center for Inherited Disease Research, Johns Hopkins University); National Human Genome Research Institute grant X01-HG007821 (M.L.M., S.M.W., and E.F.); Centers for Disease Control grant R01-DD000295 (G.L.W.); Leverhulme Trust grant F/07 134/DF (A.R.-L.); Biotechnology and Biological Sciences Research Council grant BB/I021213/1 (A.R.-L.); and the Laurel Foundation Endowment for Craniofacial Research (T.C.C.).

Received: July 24, 2017

Accepted: October 4, 2017

Published: November 30, 2017

Web Resources

1000 Genomes Project, <http://www.internationalgenome.org/>
23andMe, <https://www.23andme.com/>

Beagle, <http://faculty.washington.edu/browning/beagle/beagle.html>

ClinVar, <https://www.ncbi.nlm.nih.gov/clinvar/>

dbGaP, <https://www.ncbi.nlm.nih.gov/gap>

DECIPHER, <http://decipher.sanger.ac.uk/>

GREAT, <http://great.stanford.edu/public/html/>

HaploReg, <http://archive.broadinstitute.org/mammals/haploreg/>

IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html

Integrative Genomics Viewer, <http://software.broadinstitute.org/software/igv/>

Minimac2, <https://genome.sph.umich.edu/wiki/Minimac2>

Mouse Genome Informatics, <http://www.informatics.jax.org/>

OMIM, <http://www.omim.org/>

PubMed, <https://www.ncbi.nlm.nih.gov/pubmed/>

SHAPEIT, http://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html

VISTA Enhancer Browser, <https://enhancer.lbl.gov/>

References

1. Barnes, P., and Mertens, T.R. (1976). A survey and evaluation of human genetic traits used in classroom laboratory studies. *J. Hered.* *67*, 347–352.
2. Ordu, K.S., Didia, B.C., and Egbunefu, N. (2014). Inheritance Pattern of Earlobe Attachment amongst Nigerians. *Green. J. Hum. Phys. Anat.* *2*, 1–7.
3. Wiener, A.S. (1937). Complications in ear genetics. *J. Hered.* *28*, 425–426.
4. Adhikari, K., Reales, G., Smith, A.J., Konka, E., Palmen, J., Quinto-Sanchez, M., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Fuentes, M., et al. (2015). A genome-wide association study identifies multiple loci for variation in human ear morphology. *Nat. Commun.* *6*, 7500.
5. Eriksson, N., Macpherson, J.M., Tung, J.Y., Hon, L.S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I., and Mountain, J. (2010). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* *6*, e1000993.
6. Lai, L.Y., and Walsh, R.J. (1966). Observations on ear lobe types. *Acta Genet. Stat. Med.* *16*, 250–257.
7. Cox, T.C., Camci, E.D., Vora, S., Luquetti, D.V., and Turner, E.E. (2014). The genetics of auricular development and malformation: new findings in model systems driving future directions for microtia research. *Eur. J. Med. Genet.* *57*, 394–401.
8. Luquetti, D.V., Leoncini, E., and Mastroiacovo, P. (2011). Microtia-anotia: a global review of prevalence rates. *Birth Defects Res. A Clin. Mol. Teratol.* *91*, 813–822.
9. Belezza-Meireles, A., Clayton-Smith, J., Saraiva, J.M., and Tassabehji, M. (2014). Oculo-auriculo-vertebral spectrum: a review of the literature and genetic update. *J. Med. Genet.* *51*, 635–645.
10. Vincent, M., Geneviève, D., Ostertag, A., Marlin, S., Lacombe, D., Martin-Coignard, D., Coubes, C., David, A., Lyonnet, S., Vilain, C., et al. (2016). Treacher Collins syndrome: a clinical and molecular study based on a large series of patients. *Genet. Med.* *18*, 49–56.
11. Shaffer, J.R., Orlova, E., Lee, M.K., Leslie, E.J., Raffensperger, Z.D., Heike, C.L., Cunningham, M.L., Hecht, J.T., Kau, C.H., Nidey, N.L., et al. (2016). Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet.* *12*, e1006149.

12. Weinberg, S.M., Raffensperger, Z.D., Kesterke, M.J., Heike, C.L., Cunningham, M.L., Hecht, J.T., Kau, C.H., Murray, J.C., Wehby, G.L., Moreno, L.M., and Marazita, M.L. (2016). The 3D Facial Norms Database: Part 1. A Web-Based Craniofacial Anthropometric and Image Repository for the Clinical and Research Community. *Cleft Palate Craniofac. J.* 53, e185–e197.
13. Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., et al. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* 10, e1004572.
14. Wang, X., Lu, M., Qian, J., Yang, Y., Li, S., Lu, D., Yu, S., Meng, W., Ye, W., and Jin, L. (2009). Rationales, design and recruitment of the Taizhou Longitudinal Study. *BMC Public Health* 9, 223.
15. Tung, J.Y., Do, C.B., Hinds, D.A., Kiefer, A.K., Macpherson, J.M., Chowdry, A.B., Francke, U., Naughton, B.T., Mountain, J.L., Wojcicki, A., and Eriksson, N. (2011). Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS ONE* 6, e23473.
16. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* 10, 5–6.
17. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097.
18. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5, e1000529.
19. Fuchsberger, C., Abecasis, G.R., and Hinds, D.A. (2015). minimac2: faster genotype imputation. *Bioinformatics* 31, 782–784.
20. Durand, E.Y., Do, C.B., Mountain, J.L., and Macpherson, J.M. (2014). Ancestry Composition: A Novel, Efficient Pipeline for Ancestry Deconvolution. *bioRxiv*. <https://doi.org/10.1101/010512>.
21. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191.
22. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934.
23. Eppig, J.T., Smith, C.L., Blake, J.A., Ringwald, M., Kadin, J.A., Richardson, J.E., and Bult, C.J. (2017). Mouse Genome Informatics (MGI): Resources for Mining Mouse Genetic, Genomic, and Biological Data in Support of Primary and Translational Research. *Methods Mol. Biol.* 1488, 47–73.
24. Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* 35, D88–D92.
25. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* 84, 524–533.
26. Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44 (D1), D862–D868.
27. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501.
28. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
29. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
30. Rosin, J.M., Li, W., Cox, L.L., Rolfe, S.M., Latorre, V., Akiyama, J.A., Visel, A., Kuramoto, T., Bobola, N., Turner, E.E., and Cox, T.C. (2016). A distal 594 bp ECR specifies Hmx1 expression in pinna and lateral facial morphogenesis and is regulated by the Hox-Pbx-Meis complex. *Development* 143, 2582–2592.
31. Mou, C., Thomason, H.A., Willan, P.M., Clowes, C., Harris, W.E., Drew, C.F., Dixon, J., Dixon, M.J., and Headon, D.J. (2008). Enhanced ectodysplasin-A receptor (EDAR) signaling alters multiple fiber characteristics to produce the East Asian hair form. *Hum. Mutat.* 29, 1405–1411.
32. Bryk, J., Hardouin, E., Pugach, I., Hughes, D., Strotmann, R., Stoneking, M., and Myles, S. (2008). Positive selection in East Asians for an EDAR allele that enhances NF-kappaB activation. *PLoS ONE* 3, e2209.
33. Fujimoto, A., Ohashi, J., Nishida, N., Miyagawa, T., Morishita, Y., Tsunoda, T., Kimura, R., and Tokunaga, K. (2008). A replication study confirmed the EDAR gene to be a major contributor to population differentiation regarding head hair thickness in Asia. *Hum. Genet.* 124, 179–185.
34. Fujimoto, A., Kimura, R., Ohashi, J., Omi, K., Yuliwulandari, R., Batubara, L., Mustofa, M.S., Samakkarn, U., Settheetham-Ishida, W., Ishida, T., et al. (2008). A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet.* 17, 835–843.
35. Kimura, R., Yamaguchi, T., Takeda, M., Kondo, O., Toma, T., Haneji, K., Hanihara, T., Matsukusa, H., Kawamura, S., Maki, K., et al. (2009). A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am. J. Hum. Genet.* 85, 528–535.
36. Park, J.H., Yamaguchi, T., Watanabe, C., Kawaguchi, A., Haneji, K., Takeda, M., Kim, Y.I., Tomoyasu, Y., Watanabe, M., Oota, H., et al. (2012). Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. *J. Hum. Genet.* 57, 508–514.
37. Kamberov, Y.G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., Yang, Y., Li, S., Tang, K., Chen, H., et al. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* 152, 691–702.
38. Tan, J., Yang, Y., Tang, K., Sabeti, P.C., Jin, L., and Wang, S. (2013). The adaptive variant EDARV370A is associated with straight hair in East Asians. *Hum. Genet.* 132, 1187–1191.
39. Tan, J., Peng, Q., Li, J., Guan, Y., Zhang, L., Jiao, Y., Yang, Y., Wang, S., and Jin, L. (2014). Characteristics of dental morphology in the Xinjiang Uyghurs and correlation with the EDARV370A variant. *Sci. China Life Sci.* 57, 510–518.
40. Peng, Q., Li, J., Tan, J., Yang, Y., Zhang, M., Wu, S., Liu, Y., Zhang, J., Qin, P., Guan, Y., et al. (2016). EDARV370A

- associated facial characteristics in Uyghur population revealing further pleiotropic effects. *Hum. Genet.* *135*, 99–108.
41. Wu, S., Tan, J., Yang, Y., Peng, Q., Zhang, M., Li, J., Lu, D., Liu, Y., Lou, H., Feng, Q., et al. (2016). Genome-wide scans reveal variants at EDAR predominantly affecting hair straightness in Han Chinese and Uyghur populations. *Hum. Genet.* *135*, 1279–1286.
42. Jacques, B.E., Puligilla, C., Weichert, R.M., Ferrer-Vaquer, A., Hadjantonakis, A.K., Kelley, M.W., and Dabdoub, A. (2012). A dual function for canonical Wnt/ β -catenin signaling in the developing mammalian cochlea. *Development* *139*, 4395–4404.
43. Geng, F.S., Abbas, L., Baxendale, S., Holdsworth, C.J., Swanson, A.G., Slanchev, K., Hammerschmidt, M., Topczewski, J., and Whitfield, T.T. (2013). Semicircular canal morphogenesis in the zebrafish inner ear requires the function of gpr126 (lauscher), an adhesion class G protein-coupled receptor gene. *Development* *140*, 4362–4374.
44. Patra, C., van Amerongen, M.J., Ghosh, S., Ricciardi, F., Sajjad, A., Novoyatleva, T., Mogha, A., Monk, K.R., Mühlfeld, C., and Engel, F.B. (2013). Organ-specific function of adhesion G protein-coupled receptor GPR126 is domain-dependent. *Proc. Natl. Acad. Sci. USA* *110*, 16898–16903.
45. Perillo, L., Monsurrò, A., Bonci, E., Torella, A., Mutarelli, M., and Nigro, V. (2015). Genetic association of ARHGAP21 gene variant with mandibular prognathism. *J. Dent. Res.* *94*, 569–576.
46. Smits, P., Saada, A., Wortmann, S.B., Heister, A.J., Brink, M., Pfundt, R., Miller, C., Haas, D., Hantschmann, R., Rodenburg, R.J., et al. (2011). Mutation in mitochondrial ribosomal protein MRPS22 leads to Cornelia de Lange-like phenotype, brain abnormalities and hypertrophic cardiomyopathy. *Eur. J. Hum. Genet.* *19*, 394–399.
47. Qiao, R., He, Y., Pan, B., Xiao, S., Zhang, X., Li, J., Zhang, Z., Hong, Y., Xing, Y., and Ren, J. (2015). Understanding the molecular mechanisms of human microtia via a pig model of HOXA1 syndrome. *Dis. Model. Mech.* *8*, 611–622.
48. Lausch, E., Hermanns, P., Farin, H.F., Alanay, Y., Unger, S., Nikkel, S., Steinwender, C., Scherer, G., Spranger, J., Zabel, B., et al. (2008). TBX15 mutations cause craniofacial dysmorphism, hypoplasia of scapula and pelvis, and short stature in Cousin syndrome. *Am. J. Hum. Genet.* *83*, 649–655.
49. Candille, S.I., Van Raamsdonk, C.D., Chen, C., Kuijper, S., Chen-Tsai, Y., Russ, A., Meijlink, F., and Barsh, G.S. (2004). Dorsal ventral patterning of the mouse coat by Tbx15. *PLoS Biol.* *2*, E3.
50. Dasouki, M., Andrews, B., Parimi, P., and Kamnasaran, D. (2013). Recurrent agnathia-otocephaly caused by DNA replication slippage in PRRX1. *Am. J. Med. Genet. A.* *161A*, 803–808.
51. ten Berge, D., Brouwer, A., Korving, J., Martin, J.F., and Meijlink, F. (1998). Prx1 and Prx2 in skeletogenesis: roles in the craniofacial region, inner ear and limbs. *Development* *125*, 3831–3842.
52. Garavelli, L., and Mainardi, P.C. (2007). Mowat-Wilson syndrome. *Orphanet J. Rare Dis.* *2*, 42.
53. Van de Putte, T., Francis, A., Nelles, L., van Grunsven, L.A., and Huylebroeck, D. (2007). Neural crest-specific removal of Zfhx1b in mouse leads to a wide range of neurocristopathies reminiscent of Mowat-Wilson syndrome. *Hum. Mol. Genet.* *16*, 1423–1436.
54. Van de Putte, T., Maruhashi, M., Francis, A., Nelles, L., Kondoh, H., Huylebroeck, D., and Higashi, Y. (2003). Mice lacking ZFHx1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. *Am. J. Hum. Genet.* *72*, 465–470.
55. Hellingman, C.A., Verwiel, E.T., Slagt, I., Koevoet, W., Poulblon, R.M., Nolst-Trenité, G.J., Baatenburg de Jong, R.J., Jahr, H., and van Osch, G.J. (2011). Differences in cartilage-forming capacity of expanded human chondrocytes from ear and nose and their gene expression profiles. *Cell Transplant.* *20*, 925–940.
56. Birnbaum, S., Reutter, H., Mende, M., de Assis, N.A., Diaz-Lacava, A., Herms, S., Scheer, M., Lauster, C., Braumann, B., Schmidt, G., et al. (2009). Further evidence for the involvement of MYH9 in the etiology of non-syndromic cleft lip with or without cleft palate. *Eur. J. Oral Sci.* *117*, 200–203.
57. Hinrichs, A.L., Larkin, E.K., and Suarez, B.K. (2009). Population stratification and patterns of linkage disequilibrium. *Genet. Epidemiol.* *33* (Suppl 1), S88–S92.