A genome-wide association study identifies multiple loci for variation in human ear morphology


Here we report a genome-wide association study for non-pathological pinna morphology in over 5,000 Latin Americans. We find genome-wide significant association at seven genomic regions affecting: lobe size and attachment, folding of antihelix, helix rolling, ear protrusion and antitragus size (linear regression \( P \) values \( 2 \times 10^{-8} \) to \( 3 \times 10^{-14} \)). Four traits are associated with a functional variant in the Ectodysplasin A receptor (EDAR) gene, a key regulator of embryonic skin appendage development. We confirm expression of \( Edar \) in the developing mouse ear and that \( Edar \)-deficient mice have an abnormally shaped pinna. Two traits are associated with SNPs in a region overlapping the T-Box Protein 15 (TBX15) gene, a major determinant of mouse skeletal development. Strongest association in this region is observed for SNP rs17023457 located in an evolutionarily conserved binding site for the transcription factor Cartilage paired-class homeoprotein 1 (CART1), and we confirm that rs17023457 alters \( \text{in vitro} \) binding of CART1.

1 Department of Genetics, Evolution and Environment, UCL Genetics Institute, University College London, London WC1E 6BT, UK. 2 Centre for Cardiovascular Genetics, BHF Laboratories, Institute Cardiovascular Sciences, University College London, Rayne Building, London WC1E 6JF, UK. 3 Centro Nacional Patagónico, CONICET, Puerto Madryn U9129ACD, Argentina. 4 National Institute of Anthropology and History, Mexico City 4510, Mexico. 5 GEMMOL (Genética Molecular), Universidad de Antioquia, Medellín 5001000, Colombia. 6 Instituto de Alta Investigación, Universidad de Tarapacá, Programa de Genética Humana ICBM Facultad de Medicina Universidad de Chile and Centro de Investigaciones del Hombre en el Desierto, Arica 1000000, Chile. 7 Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM-Instituto Nacional de Medicina Genómica, Mexico City 4510, Mexico. 8 Facultad de Medicina, UNAM, Mexico City 4510, Mexico. 9 Departamento de Genética, Universidad Federal do Rio Grande do Sul, Porto Alegre 91501-970, Brazil. 10 Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima 31, Peru. 11 Departamento de Zoología y Antropología Física, Universidad Complutense de Madrid, Madrid 28040, Spain. 12 Departamento de Antropología, Facultad de Ciencias Sociales y Humanas, Universidad de Antioquia, Medellín 5001000, Colombia. 13 The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, UK. 14 UCL Ear Institute, University College London, London W1X 8EE, UK. 15 Royal National Throat Nose and Ear Hospital, London W1X 8EE, UK. 16 Schools of Biosciences and Mathematics & Statistics, University of Melbourne, Melbourne, Victoria 3010, Australia. Correspondence and requests for materials should be addressed to A.R.L. (e-mail: a.ruizlin@ucl.ac.uk).
The human pinna is made up of a piece of cartilage covered with skin and attached to the skull by ligaments, muscles and fibrous tissue. This cartilage does not extend into the ear lobe, which consists mostly of areolar and adipose tissue. A range of disorders affecting human pinna development have been described, occurring in isolation or as part of complex syndromes with multiple affected organs.3,4 There is also great non-pathological variation between humans in pinna shape and size, and this variation has been reported to be influenced by age, sex and ethnicity.3–5 The study of rare, familial microtia cases (a disorder characterized by a small, abnormally shaped pinna) has established that mutations in the HOXA2 gene can severely impact pinna development.6 However, no genetic variants influencing normal pinna morphology have yet been reported. Here we aimed at identifying such variants by performing a genome-wide association study (GWAS) in a large sample of Latin American individuals with no pinna abnormalities. We identified seven loci with genome-wide significant association to variation in various pinna features, including several strong candidate genes with known developmental effects. We provide follow-up experimental evidence supporting the pinna morphology associations for two gene regions (which include the Ectodysplasin A receptor (EDAR) and the T-Box Protein 15 (TBX15) genes).

Results

Our study sample is part of the CANDELA (Consortium for the Analysis of the Diversity and Evolution of Latin America; http://www.ucl.ac.uk/silva/candela) cohort collected in five Latin American countries (Brazil, Colombia, Chile, Mexico and Peru) for the study of the genetics of physical appearance. Our study sample consists of individuals of both sexes (median age 22 years), with a mixed African, European and Native American genetic background (Supplementary Table 1). Using facial photographs, we performed a qualitative assessment (on a three-point-ordered categorical scale; Fig. 1; Supplementary Figs 1 and 2) of 10 pinna traits in 5,062 individuals: ear protrusion, lobe size, lobe attachment, tragus size, antitragus size, helix rolling, folding of antihelix, crus helix expression, superior crus of antihelix expression and Darwin’s tubercle.

Variation in the pinna traits examined.

The trait scores show a weak-to-moderate correlation between them, and with age, sex, body mass index (BMI) and genetic ancestry (Supplementary Table 2). Most noticeably, lobe attachment shows a moderate and significant (permutation P value < 0.001) negative correlation with lobe size (r = −0.49). Significant but weaker positive correlations were observed for folding of antihelix with helix rolling (r = 0.25) and with superior crus of antihelix expression (r = 0.23), as well as between ear protrusion and helix rolling (r = 0.16).

Individuals were genotyped on Illumina’s Omni Express BeadChip. After quality control, 671,038 single-nucleotide polymorphisms (SNPs) and 4,919 individuals were retained for further analyses. Autosomal admixture proportions for the full sample were estimated as: 53% European, 43% Native American and 4% African (Supplementary Fig. 3). On the basis of a kinship matrix derived from the genome-wide SNP data, we estimated narrow-sense heritability using GCTA7 and found moderate and significant values for all traits, with the highest heritability observed for ear protrusion (61%) and the lowest for tragus size (25%) (Supplementary Table 3). Similar heritabilities have been calculated for a range of facial traits using family data.8,9

Primary association analyses.

For the primary genome-wide association tests, we applied multivariate linear regression, as implemented in PLINK, using an additive genetic model adjusting for: age, sex, height, BMI and the first five principal components (PCs) calculated from SNP data. The resulting statistics showed no evidence of residual population stratification for any of the traits (Supplementary Fig. 4). To account for the possibility of cryptic relatedness between individuals, we also performed genome-wide association tests using random-effects mixed linear regression models (using Fast.MRML10) and obtained similar results as in the PLINK analyses (Supplementary Table 4).

Secondary analyses. Since correlations between traits could have a shared underlying genetic basis, we performed a multivariate analysis combining all phenotypes in a single regression model (using MULTIPHEN11). This analysis identified the same set of regions as in the independent regression tests, but no additional associated regions (Supplementary Fig. 5). We also examined the association signals for all index SNPs (Table 1) in each country separately and combined results as a meta-analysis using METAL12. For each association, the effects were in the same direction in all countries, the variability of effect size across countries reflecting sample size (Fig. 2; Supplementary Table 6; Supplementary Fig. 6a). There was no significant evidence of effect size heterogeneity across countries for any of the associations. A full genome-wide meta-analysis did not reveal additional regions showing significant association with pinna morphology (Supplementary Fig. 6b). The seven index SNPs of Table 1 provide modest phenotypic prediction accuracy (Supplementary Table 7). The fraction of the phenotypic variance explained by these SNPs is small relative to the heritability estimates (Supplementary Tables 3 and 7), suggesting a polygenic architecture for these traits beyond that captured by the genome-wide significance threshold used.

Features of associated regions.

The genomic regions showing genome-wide significant association have features with independent evidence suggestive of an involvement in pinna development. This evidence is particularly strong for the regions in 2q12.3 and 1p12, and we followed-up these two regions with additional experiments.

2q12.3. Includes SNPs associated with four traits (lobe size, lobe attachment, helix rolling and ear protrusion; Table 1). These SNPs extend over ~500 kb and show substantial linkage disequilibrium (LD; Fig. 3b). Strongest association with all four traits was found for SNP rs3827760 (Table 1), and conditioning on this SNP abolishes the signal of association at other SNPs in the region (Supplementary Fig. 8a). Marker rs3827760 leads to a functional p.Val370.Ala substitution in the intracellular death
domain of EDAR. This residue affects the interaction with the EDAR-binding death domain adapter protein EDARADD\(^{16}\); the derived EDAR allele encodes a protein with higher activity than the ancestral one\(^{17,18}\). EDAR signalling acts during prenatal development to specify the location, size and shape of ectodermal appendages, such as hair follicles, teeth and glands\(^{19}\). The p.Val370Ala variant has been associated with characteristic tooth morphologies, hair type and sweat gland density in East Asians\(^{20–25}\), where this allele is present at high frequency while being nearly absent in European or African populations (Supplementary Table 8). Consistent with these effects in humans, mice expressing EDAR370A, or with increased EDAR function, show thickened and straightened hair fibres\(^{17,20,24,25}\).

We examined Edar expression in the developing mouse embryo (Fig. 4a). The structure of the pinna is defined prenatally in both mouse and human, in mice primarily between gestation days 13 and 16 (ref. 2). Using whole-mount in situ hybridization, we confirmed Edar expression along the distal margin of the developing pinna (Fig. 4a), in addition to the well-characterized expression in the developing hair follicles\(^{16}\). Edar expression at the distal margin of the embryonic pinna may aid in determining its growth and expansion, thus influencing the ultimate form adopted by the ear. To assess the functional significance of this Edar expression during pinna development, we examined postnatal pinna morphology in Edar mouse mutants. The Edar\(^{dlJ}\) (ref. 26) and EdarTg951 (ref. 16) mouse lines have a loss and a gain of Edar function, respectively (see Methods). At 2 weeks of age, the pinna of homozygous Edar\(^{dlJ}/dlJ\) mutant and heterozygous Edar\(^{dlJ}/+\) littermates, wild-type mice and EdarTg951 mice. Landmark coordinate PC1, capturing 69% of the variation in shape, reflects mainly a change in the extent of helix rolling of the mouse pinna (Fig. 4c). Quantitative assessment of mouse pinna protrusion (assessed by measuring protrusion angle) and shape (assessed by PC analysis of two-dimensional landmark coordinates) revealed significant differences (linear regression P value < 0.0007; Supplementary Note 1; Supplementary Figs 9 and 10) between the homozygous Edar\(^{dlJ}/dlJ\) mutant and heterozygous Edar\(^{dlJ}/+\) littermates, wild-type mice and EdarTg951 mice. Landmark coordinate PC1, capturing 69% of the variation in shape, reflects mainly a change in the extent of helix rolling of the mouse pinna (Fig. 4e; Supplementary Fig. 10d), consistent with one of the effects we observed of EDAR variants on human pinna shape. The Edar high copy-number transgenic (EdarTg951) does not display a detectable change in helix rolling likely due to the fact that the wild-type mouse pinna does not have a prominent helix roll, thus hampering the detection of any further pinna flattening that might be caused by increased Edar function.

---

**Figure 1** | Genome-wide associations of pinna traits. Variation in 10 pinna traits was evaluated in 5,062 individuals. The photographs at the top indicate the location of the traits examined. At the bottom is shown a composite Manhattan plot for the seven traits showing genome-wide significant association with at least one genome region. The rs numbers for the most significantly associated (index) SNP in each region are provided (Table 1). Each of the seven regions on the Manhattan plot is connected with the associated trait on the photos via a line of different colour (composite panels in this and subsequent figures were made using Photoshop\(^{39}\)).
1p12. SNPs in this region show genome-wide association with antihelix folding and antitragus size. This region extends over ~800 kb and overlaps the gene encoding transcription factor *TBX15* (Fig. 3a), a key regulator of cartilaginous and skeletal development in the mouse 27. A spontaneous *Tbx15* mouse mutant (called ‘droopy ears’), is characterized by craniofacial and cartilage abnormalities in the mouse 31. The location of rs17023457 in a paired-class homeoprotein 1) (Supplementary Fig. 11a), suggests that this SNP could directly influence the expression of neighbouring genes involved in cartilaginous skeletal development in the mouse 27. A spontaneous *Tbx15* mouse mutant (called ‘droopy ears’), is characterized by craniofacial and cartilage abnormalities in the mouse 31. An intergenic SNP rs17023457 in the forward and reverse orientation, respectively constructs with the C allele when the genomic sequence was used to drive the expression of the luciferase gene, showed a decreased expression of 36 and 22% in the rs17023457 T or C alleles driving the expression of the CART1 in vitro, confirming the specificity of the assay and supporting the binding and full transactivation occurs or not. The implications of these observations for the in vivo regulation of *TBX15* expression remain to be established.

Other regions. For the other regions showing genome-wide association, there is currently less compelling information suggestive of a mechanism explaining their effect on pinna morphology, and we did not attempt their experimental follow-up. SNPs in 3q23 are associated with lobe size, with strongest association being seen for intergenic SNP rs10212419 (P value 3 × 10^{−14}), in a region with substantial LD over ~500 kb (Fig. 3d). Intriguingly, considering that the ear lobe is made up mainly of loose connective tissue, intergenic SNPs in this region have been strongly associated with keloid formation 32, an exaggerated skin wound-healing reaction characterized by excessive deposition of extracellular matrix and collagen fibres. Some highly penetrant mutations across this genomic region have also been found to result in alterations of craniofacial development involving pinna morphology. The gene nearest to SNP rs10212419 (~59 kb away) encodes mitochondrial ribosomal protein S22 (MRPS22), mutations of which can lead to combined oxidative phosphorylation deficiency 5 (COXPD5), a disorder whose phenotype can include low-set, posteriorly rotated ears (OMIM #611719). A report of a patient with a similar ear phenotype found a non-synonymous substitution at an evolutionarily conserved site within MRPS22 (ref. 33). Another interesting candidate gene in the vicinity is that encoding the forkhead box L2 transcription factor (FOXL2). Coding and regulatory mutations of this gene cause BPES (blepharophimosis, ptosis, epicanthus inversus syndrome) 34,35, a disorder characterized by a range of craniofacial abnormalities, including alterations in ear lobe morphology 36.

The 2q31.1 region shows considerable LD over ~100 kb and is associated with lobe attachment size, with strongest association seen for intergenic SNP rs2080401 (P values of 9 × 10^{−12} and 1 × 10^{−10} for these two phenotypes, respectively; Fig. 3c). This SNP is located ~31 kb upstream of the gene encoding Specificity Protein 5 (SP5), a Sp1-related transcription factor that mediates responses to Wnt/beta-catenin signalling. Of potential interest, there is evidence of an involvement of this pathway in musculo-skeletal development 37,38. The region in 6q24.2, associated with lobe size, shows substantial LD over ~500 kb overlapping the hypothetical protein-coding gene *LOC153910*. Strongest association was seen for intronic rs263156 (P value 2 × 10^{−13}; Fig. 3f). Potentially of relevance to ear development, about 100 kb from rs263156 is GPR126 (G protein-coupled receptor 126), a gene strongly associated with human height 39,41. Borderline genome-wide significance was found for rs1960918 in 4q31.3 (P value 2 × 10^{−8} for helix expression of the neighbour genes in this genomic region have also been found to result in alterations of craniofacial development involving pinna morphology.

### Table 1: Chromosomal location and −log_{10} (P) for index SNPs showing strongest genome-wide association to pinna traits.

<table>
<thead>
<tr>
<th>Region</th>
<th>SNP</th>
<th>Nearest gene</th>
<th>EP</th>
<th>AS</th>
<th>FoA</th>
<th>HR</th>
<th>LA</th>
<th>LS</th>
<th>TS</th>
<th>CHE</th>
<th>DT</th>
<th>SCoAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p12</td>
<td>rs17023457</td>
<td>CART1 binding site</td>
<td>2.37</td>
<td>7.76</td>
<td>10.94</td>
<td>1.12</td>
<td>0.97</td>
<td>0.16</td>
<td>0.27</td>
<td>1.62</td>
<td>0.31</td>
<td>1.26</td>
</tr>
<tr>
<td>2q12.3</td>
<td>rs3827760</td>
<td>EDAR</td>
<td>10.00</td>
<td>0.13</td>
<td>1.92</td>
<td>12.00</td>
<td>8.66</td>
<td>11.85</td>
<td>4.33</td>
<td>0.63</td>
<td>0.59</td>
<td>0.13</td>
</tr>
<tr>
<td>2q31.1</td>
<td>rs2080401</td>
<td>SP5</td>
<td>0.16</td>
<td>0.40</td>
<td>0.41</td>
<td>0.49</td>
<td>11.03</td>
<td>9.90</td>
<td>1.13</td>
<td>0.50</td>
<td>0.25</td>
<td>0.03</td>
</tr>
<tr>
<td>3q23</td>
<td>rs10212419</td>
<td>MPRS22</td>
<td>0.08</td>
<td>0.27</td>
<td>0.06</td>
<td>0.12</td>
<td>1.41</td>
<td>13.49</td>
<td>1.98</td>
<td>0.08</td>
<td>2.59</td>
<td>0.15</td>
</tr>
<tr>
<td>4q13.3</td>
<td>rs1960918</td>
<td>LRBA</td>
<td>0.42</td>
<td>3.16</td>
<td>1.13</td>
<td>7.64</td>
<td>0.88</td>
<td>1.81</td>
<td>2.96</td>
<td>0.46</td>
<td>0.70</td>
<td>0.15</td>
</tr>
<tr>
<td>6q24.2</td>
<td>rs263156</td>
<td>LOC153910</td>
<td>0.31</td>
<td>0.13</td>
<td>0.99</td>
<td>1.75</td>
<td>5.66</td>
<td>12.61</td>
<td>0.57</td>
<td>0.04</td>
<td>2.18</td>
<td>0.78</td>
</tr>
<tr>
<td>18q21.2</td>
<td>rs1619249</td>
<td>LOC100287225</td>
<td>1.22</td>
<td>0.27</td>
<td>7.99</td>
<td>1.22</td>
<td>0.01</td>
<td>0.92</td>
<td>0.92</td>
<td>1.48</td>
<td>0.33</td>
<td>0.09</td>
</tr>
</tbody>
</table>

SNPs: single-nucleotide polymorphism. Trait acronyms (Fig. 1): AS, antitragus size; CHE, crus helix expression; DT, Darwin’s tubercle; EP, ear protrusion; FoA, folding of antihelix; HR, helix rolling; LA, lobe attachment; LS, lobe size; SCoAE, superior crus of antihelix expression; TS, tragus size.

Genome-wide significant values (−log_{10} (P) > 7.3) are highlighted in the darkest shade of red. Below this significance threshold, intensity of colour background is proportional to −log_{10} (P) value. Effect sizes for associated alleles are shown in Fig. 2 and Supplementary Table 5. Intragenic SNPs are shown in bold. In 1p12, SNP rs17023457 is in an evolutionarily conserved binding site for the CART1 transcription factor. GWAS P values were obtained using an additive multivariate regression model as described in Methods.
rolling; Fig. 3e) and rs1619249 in 18q21.2 (P value $1 \times 10^{-8}$ for antihelix folding; Fig. 3g). Intrinsic marker rs1960918 is in an LD region of ~400 kb overlapping the LRBA gene, whose product (LPS-responsive vesicle trafficking, beach and anchor containing) is known to be involved in coupling signal transduction, vesicle trafficking and immunodeficiency,
Figure 3 | LocusZoom and linkage disequilibrium plots of significantly associated genetic regions. Plots of the seven genomic regions showing genome-wide significant association to pinna traits (Table 1). For regions showing association with several pinna traits, we present here only results for the trait with strongest association (plots for the other associated traits are presented in Supplementary Fig. 7). Association results from a multivariate linear regression model (on a $-\log_{10} P$ scale; left $y$ axis) are shown for SNPs $\sim 500$ kb on either side of the index SNP (that is, the SNP with the smallest $P$ value, purple diamond; Table 1) with the marker (dot) colour indicating the strength of LD ($r^2$) between the index SNP and that SNP in the 1000 Genomes AMR data set. Local recombination rate in the AMR data is shown as a continuous blue line (scale on the right $y$ axis). Genes in each region, their intron–exon structure, direction of transcription and genomic coordinates (in Mb, using the NCBI human genome sequence, Build 37, as reference) are shown at the bottom. Plots were produced with LocusZoom\textsuperscript{56}. Below each region we also show an LD heatmap (using $D'$, ranging from red indicating $D' = 1$ to white indicating $D' = 0$) produced using Haploview\textsuperscript{57}. Note that the location of SNPs on the LD heatmap can be shifted relative to the regional display on top of it.
with no obvious functional connection to pinna development. Intergenic SNP rs1619249 is in an LD region of about 300 kb, the closest candidate being LOC100287225 (about 100 kb from rs1619249), a hypothetical gene of unknown function.

Discussion
In conclusion, we have identified the first genetic variants influencing normal variation in human pinna morphology. It will be important to evaluate further the role of these regions in ear development and its disorders. Since pinna morphology in

Figure 4 | Effect of Edar genetic variation on mouse pinna shape. (a) Whole-mount in situ hybridization detecting Edar expression in the developing mouse embryo at 15 days’ gestation. (b–e) Impact of Edar genotype on mutant mouse ear shape. (b,c) Photographs of wild-type (b) and homozygous EdardIJ/EdardIJ (c) mutant mice from top and side views (respectively, on the upper and lower panels). (d,e) Boxplots, respectively, of ear protrusion angle and of landmark coordinate PC1 (y axis) for different mouse genotypes (x axis) (Supplementary Note 1; Supplementary Fig. 10 shows additional analyses for ear protrusion). Boxplot whiskers extend to data points within 1.5 times the interquartile range on both sides. In d,e numbers in parenthesis below genotypic categories refer to the number of mice examined for each. On the right of e are shown average PC1 wireframes for EdardIJ homozygous mice (bottom) or for mice with other genotypes (top).
mammals shows extensive evidence of evolutionary adaptation, particularly in relation to thermoregulation, acoustic perception and sound localization[2–3], it will be interesting to examine whether variation in these genomic regions relates to adaptive changes in the pinna across species. Finally, there is increasing interest in the use of pinna variation in forensics[15–16], and development of such applications may benefit from a refined knowledge of the genetic determinants of pinna morphology.

Methods

Study subjects. In total, 5,062 volunteers from 5 Latin American countries (Brazil, Chile, Colombia, Mexico and Peru), part of the CANDELA consortium sample (http://www.uc.lac.uk/silva/candela), were included in this study. Ethics approval was obtained from: Universidad Nacional Autónoma de México (Mexico), Universidad de Antioquia (Colombia), Universidad Perúana Cayetano Heredia (Peru), Universidad de Tarapacá (Chile), Universidade Federal do Rio Grande do Sul (Brazil) and University College London (UK). All participants provided written informed consent. Blood samples were collected by a certified phlebotomist and DNA extracted following standard laboratory procedures. Five digital photographs of the face: left side (−90°), left angle (−45°), frontal (0°), right angle (45°) and right side (90°) were taken from ~1.5 m at eye level using a Nikon D90 camera fitted with a Nikkor 50 mm fixed focal length lens. Other phenotypes including height, weight, BMI, age and sex were also recorded for each participant.

Pinna phenotyping. Right side, right angle and frontal photographs were used to score 10 pinna traits. These were (Fig. 1): ear protrusion, lobe size, lobe attachment, tragus size, antitragus size, helix rolling, folding of antihelix, crus helix expression, superior crus of antihelix expression and Darwin’s tubercle. Each trait was scored as an ordered categorical variable, with 0 being the lowest level of expression of the trait and 2 the highest (Supplementary Table 1). Software to assist scoring and all photographs was developed in MATLAB[47]. Intraclass correlation coefficients[48] calculated by double-scoring photographs of 100 subjects indicate a moderate-to-high intra- and inter-rater reliability of the trait scores (Supplementary Table 9). Photographs for all the volunteers were scored by the same rater (G.R.).

DNA genotyping and quality control. DNA samples from participants were genotyped using the HumanOmniExpress BeadChip (Illumina). PLINK[12] was used to exclude SNPs and individuals with >5% missing data, markers with minor allele frequency <1% and related individuals. After applying these filters, 680,634 SNPs were retained for further analysis. Owing to the admixed nature of the study sample (Supplementary Fig. 3), there is an inflation in Hardy-Weinberg P values. We therefore did not exclude markers based on the Hardy-Weinberg deviation.

Statistical analyses. P values for Pearson correlation coefficients were obtained by permutation. All regressions were performed using an additive multivariate linear or logistic model, providing P values that are obtained from the standard f-statistic derived from their regression coefficients. Narrow-sense heritability (defined as the phenotypic variance explained by a genetic relatedness matrix, GRM, computed from the 500,000 SNPs and the pinna traits estimated using GCTA[17]) by fitting an additive linear model with a random-effect term whose variation is given by the GRM, with age, sex, height and BMI as covariates. The GRM was obtained using the LDAD approach[18], which accounts for LD between SNPs. An LD-pruned set of 93,328 autosomal SNPs was used to estimate continental ancestry using ADMIXTURE[49] (Supplementary Fig. 3). PLINK[12] was used to perform the primary genome-wide association tests for each phenotype using multiple linear regression with an additive genetic model incorporating age, sex, height, BMI and 3 (or 10) genetic PCs as covariates. These genetic PCs were obtained using (MATLAB) from an LD-pruned set of 93,328 SNPs. Inspection of the scree plot (Supplementary Fig. 4) and the PC scatterplot (not shown) indicates that only the first 5 PCs show evidence of structure. Consistent with this, PLINK results with 5 or 10 PCs were nearly identical. The Q–Q plots for all PLINK analyses showed no sign of inflation, the genomic control factor lambda being 1.03 in all cases. To account for multiple testing, we also applied (using MATLAB) a global FDR test across all traits and SNPs (Supplementary Table 5). To account for the possibility that cryptic relatedness among the individuals studied could affect the results, we also performed a GWAS combining all traits in a single analysis using a multivariate regression model as implemented in MULTIPHEX[30] on an LD-pruned data set (Supplementary Fig. 5). This approach uses a SNP genotype as the dependent variable and all other phenotypes are jointly taken as covariates. Due to this increased complexity the runtime per SNP is considerably longer, so an LD-pruned subset of 189,707 SNPs was used for this analysis (ensuring that all genome-wide or suggestive SNPs from the primary analyses are included). A meta-analysis (Supplementary Table 6, Supporting Information) was carried out for each of the 7 SNPs separately in each country sample (using PLINK as above) and combining the results in METAL[12]. Forest plots were produced with METAL combining all regression coefficients and s.e. Cochran’s Q-statistic was computed for each trait to test for effect size heterogeneity across country samples. Prediction of trait phenotypes from genotypes at the seven index SNPs identified in the primary GWAS analysis (Table 1) was performed in MATLAB[42] using linear regression and nonlinear neural network models[31] (Supplementary Table 7).

Whole-mount in situ hybridization of mouse embryos. Embryos were fixed in 4% paraformaldehyde in PBS at 4°C overnight, dehydrated to methanol, rehydrated, treated with proteinase K, refixed and hybridized to a digoxigenin-labelled cRNA covering the entire Edar transcript (AF160502). Hybridization was performed in 50% formamide, 750 μM NaCl, 75 mM Na3C6H5O7, 1% SDS containing 50 μg ml−1 yeast RNA and 50 μg ml−1 heparin at 65°C overnight. After washing and blocking of embryos in 10% sheep serum, the signal was detected using 1/2,000 sheep anti-digoxigenin conjugated to alkaline phosphatase (sheep anti-digoxigenin alkaline phosphatase antibody; 11093274910, Roche Applied Science, Mannheim, Germany) followed by NBT (nitro-blue tetrazolium chloride) and BCIP (3-bromo-4-chloro-3’-indolylphosphate p-toluidine) staining in 100 mM NaCl, 100 mM Tris pH 9.5, 50 mM MgCl2 and 0.1% Tween 20.

Animals and sample collection. Animal studies were reviewed and approved by The Roslin Institute Animal Welfare and Ethical Review Body. The humane care and use of mice (Mus musculus) in this study was carried out under the authority of the appropriate UK Home Office Project License. Forty-two mouse head samples (male and female) were included in the study. We obtained four strains including C57BL/6J, C3H/HeN, C57BL/6J × 129/SvEv (C57/129F2), and C57BL/6J × 129/SvEv (C57/129F2) × 129/SvEv (F2) from the Jackson Laboratory (USA), the University of Antioquia (Colombia), Universidad Peru´ana Cayetano Heredia (Peru), Universidad de Tarapacá (Chile), Universidade Federal do Rio Grande do Sul (Brazil) and University College London (UK). All participants provided written informed consent. DNA extracted following standard laboratory procedures. Five digital photographs was developed in MATLAB47. Intraclass correlation coefficients48 calculated by double-scoring photographs of 100 subjects indicate a moderate-to-high intra- and inter-rater reliability of the traits scores (Supplementary Table 9). Photographs for all the volunteers were scored by the same rater (G.R.).

Mouse pinna phenotyping. Overall, pinna shape variation was examined using geometric morphometric techniques[52]. Twenty-seven landmarks and semilandmarks (depicted in Supplementary Fig. 8) were digitized, scaled and superimposed using TPS dịp and TPSRelW (http://life.bio.sunysb.edu/morph/) routine was used to allow semilandmarks to slide so as to minimize bending energy. Reliability of pinna landmarking was evaluated by scoring the same landmarks by a second rater and examining the variation (Supplementary Table 10). Generalized procrustes analysis was used to remove the effects of translation, rotation and scaling[52]. After alignment the superimposition of the generalized procrustes analysis-adjusted landmark coordinates, only the shape component remained in the aligned specimens. The superimposed coordinates were exported to the software MorphoJ[50] to obtain models of shape. Elliptical Fourier analysis[51] was used to reduce a series of 2N coefficients computed for each of the shapes included in the analysis (Supplementary Note 1): one with four distinct genotype categories and one considering genotypes as a binary variable, based on the notion that the Edar loss-of-function homozygous genotype (Edar−/−Edar−/−) is recessive.

Human tissue culture for the CART1-binding assay. The human hepatoma Huh7 (ECACC, UK) cell line was grown in high-glucose DMEM (PAA) supplemented with 2 mM L-glutamine and 10% fetal bovine serum and maintained at 5% CO2 at 37°C. Nuclear extracts were obtained from Huh7 cells using the NE-PER Nuclear and Cytoplasmic Extraction Reagents Kit (Thermo Scientific, USA) as described in the manual, with the addition of Complete Protease Inhibitor (Roche, UK) to buffers CER I and NER I.
Electrophoretic mobility shift assay. Probe sequences: rs17023457 T>C
rs17023457 T forward 5'-ACTAATCTACACATCCTCTTTCCGAAAATC-3’
rs17023457 T reverse 5’-TGATGTCGCGCGGGAGTGATGT AGT-3’
rs17023457 C forward 5’-ACTAATCTACACATCCTCTTTCCGAAAATC-3’
rs17023457 C reverse 5’-TGATGTCGCGCGGGAGTGATGT AGT-3’
CARTI consensus 5’-CCACATATAATTGTTATCTGG-3’
CARTI consensus reverse 5’-CAAGATATAATTGTTATCTGG-3’
Probes were labelled using the Biotin 3’-End DNA Labelling Kit (Pierce) as described in the manufacturer’s manual. Binding reactions consisted of 2 μl 10 × binding buffer (100 mM Tris, 500 mM KCl, pH 7.5), 1 μg p(dI-dC)2, 200 fmol biotin-labelled DNA, made to a total of 20 μl with H2O and incubated at 25 °C for 30 min, following by the addition of 5 x loading buffer. Competition reactions were performed with 30 min incubation on ice, prior to addition of labelled probes, using 100 × CARTI cold-competitor oligonucleotides. Samples were run on 6% polyacrylamide gel, electrophoresed for 150 min at 120 V.

Luciferase reporter assay. Luciferase reporter vectors were based on the pGL3 Promoter vector (Promega), with a 872-bp genomic sequence surrounding rs17023457 (Supplementary Fig. 11c) inserted upstream of the SV40 promoter in both orientations using the following primers designed for the InFusion PCR Cloning kit (Clontech):

Forward EarIFF: 5’-TCTATGAGTTACACATACATTACGGTTGAGGG-3’
EarIFR: 5’-GATCGGATCCGCGGACTGATCTCAAGGTCGCAA-3’
Reverse EarRF: 5’-GATCGGATCCGCGGACTGATCTCAAGGTCGCAA-3’
EarRFR: 5’-TCTATGAGTTACACATACATTACGGTTGAGGG-3’
Site-directed mutagenesis was performed using the QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) as detailed in the manual, using the following oligonucleotide sequences (designed for reverse strands):

Primer name; primer sequence (5’-3’)
SDMFFC 5’-GATCGGATCCGCGGACTGATCTCAAGGTCGCAA-3’
SDMERFC 5’-ATGATCGGCACACTAATACACTACACCTCTTGTGGCCAA-3’
TAC-3’
Primer name; primer sequence (5’-3’)
SDMRF5 5’-GACGCTACTAACATACACACCTTCCTTGACGATTGCATC-3’
CAG-3’
SDMBRG 5’-CTGAAAGCTGTTAGTCCGAAAGGATGTTAGTATGTTG-3’
GTC-3’
All constructs were verified by direct sequencing.

Huh7 cells were seeded at a density of 2 x 10^3 per well in a 96-well plate and grown to confluence overnight in the appropriate media (described above). Cells were transfected with 250 ng pGL3 reporter construct, with 10 ng pRLTK as a transfection control. Transfection was carried out in Opti-Mem serum-free media (Promega), and measured in the Luciferase reporter assay using Passive Lysis Buffer (Promega) and luciferase expression was determined of scapula and pelvis, and short stature in Cousin syndrome.


Acknowledgements

We thank the volunteers for their enthusiastic support. We also thank Alvaro Alvarado, Mónica Ballesteros Romero, Mari-Wyn Burley, Ricardo Cebrecos, Debayoti Choudhury, Miguel Ángel Contreras Sieck, Francisco de Ávila Becerril, Joyce De la Piedra, María Teresa Del Solar, Paola Everardo Martínez, William Flores, Martha Granados Riveros, Wendy Hart, Ilich Jalet Moreno, Jodie Lampert, Mowmita Basak Mow, Paola León-Mimila, Javier Mendoza, Francisco Quispalayla, Diana Rogel Diaz, Ruth Rojas, Norman Russell, Vanessa Sarabia and Fabienne Tessiot for assistance with volunteer recruitment, sample processing and data entry. We are very grateful to the institutions that kindly provided facilities for the assessment of volunteers, including: Escuela Nacional de Antropología e Historia and Universidad Nacional Autónoma de México (Mexico); Pontificia Universidad Católica del Perú, Universidad de Lima and Universidad Nacional Mayor de San Marcos (Peru). This work was funded by grants from the Leverhulme Trust (F/R/07 134/DF to A.R.-L.), BBSRC (BB/J021231/1 to A.R.-L. and Institute Strategic Programme grant to The Roslin Institute), Universidad de Antioquia (GENMOL sostenibilidad 2015–2016 and MASO 2013–2014). S.E.H. holds a British Heart Foundation Chair and acknowledges funding from grant RG008/08 and is also supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. A.J.P.S. is funded by a BHF Fellowship (FS/13/6/29977).

Author contributions


Additional information

Supplementary Information accompanies this paper at http://www.nature.com/ncemms.

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/


This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/