A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation

How to cite:
Adhikari, Kaustubh; Fuentes-Guajardo, Macarena; Quinto-Sánchez, Mirsha; Mendoza-Revilla, Javier; Camilo Chacón-Duque, Juan; Acuña-Alonzo, Victor; Jaramillo, Claudia; Arias, William; Lozano, Rodrigo Barquera; Pérez, Gastón Macín; Gómez-Valdés, Jorge; Villamil-Ramírez, Hugo; Huneemeier, Tábita; Ramallo, Virginia; Silva de Cerqueira, Caio C.; Hurtado, Malena; Villegas, Valeria; Granja, Vanessa; Gallo, Carla; Poletti, Giovanni; Schuler-Faccini, Lavinia; Salzano, Francisco M.; Bortolini, Maria- Cátira; Canizales-Quinteros, Samuel; Cheeseman, Michael; Rosique, Javier; Bedoya, Gabriel; Rothhammer, Francisco; Headon, Denis; González-José, Rolando; Balding, David and Ruiz-Linares, Andrés (2016). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nature Communications, 7, article no. 11616.

For guidance on citations see FAQs.

© [not recorded]

https://creativecommons.org/licenses/by/4.0/

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.1038/ncomms11616

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
We report a genome-wide association scan for facial features in ~6,000 Latin Americans. We evaluated 14 traits on an ordinal scale and found significant association (P values < 5 × 10⁻⁸) at single-nucleotide polymorphisms (SNPs) in four genomic regions for three nose-related traits: columella inclination (4q31), nose bridge breadth (6p21) and nose wing breadth (7p13 and 20p11). In a subsample of ~3,000 individuals we obtained quantitative traits related to 9 of the ordinal phenotypes and, also, a measure of nasion position. Quantitative analyses confirmed the ordinal-based associations, identified SNPs in 2q12 associated to chin protrusion, and replicated the reported association of nasion position with SNPs in PAX3. Strongest association in 2q12, 4q31, 6p21 and 7p13 was observed for SNPs in the EDAR, DCHS2, RUNX2 and GLI3 genes, respectively. Associated SNPs in 20p11 extend to PAX1. Consistent with the effect of EDAR on chin protrusion, we documented alterations of mandible length in mice with modified Edar function.
Humans show extensive variation in facial features, physical anthropologists having long used this variation to examine human population diversification, including the possibility that these features have been influenced by adaptation to the environment\(^1\). It has also been proposed that the diversity of human faces could have evolved partly to facilitate individual recognition, a key aspect of social interaction\(^1\). Other than their considerable evolutionary interest, a range of categorical and quantitative craniofacial traits have been used in forensics for the purpose of human identification and estimation of ancestry\(^5,6\). Consistent with facial variation being under strong genetic control, heritabilities of $\sim 60\text{–}90\%$ have been estimated for facial phenotypes\(^7,8\). The characterization of gene mutations in patients with dysmorphologies and in animal models has enabled the identification of rare genetic variants with major effects on facial development\(^9\). However, our current understanding of the molecular genetic basis of variable face appearance in the general human population is scant. Only two genome-wide association studies (GWAS) for facial features have so far been published\(^9,11\). These were carried out in Europeans and only one gene region (PAX3) was consistently associated with a facial feature in both studies (position of the nasion, the deepest point on the nasal bridge)\(^9,11\).

We recently reported the successful identification of genes influencing outer ear (pinna) morphology using a three-point ordinal phenotyping approach\(^12\). Here, we extend this methodology to other aspects of facial variation. In addition, in a subsample of individuals, we obtained quantitative measures related to the ordinal phenotypes examined. Our analyses allowed us to replicate the reported PAX3/nasion position association and to identify five other gene regions impacting on face (mostly nose) morphology (EDAR, DCHS2, RUNX2, GLI3 and PAX1). These genes are known to play important roles in craniofacial development.

**Results**

**Study sample and ordinal phenotypes.** Our study sample is part of the CANDELA cohort collected in Latin America\(^13\) (Supplementary Table 1). Using facial photographs of 6,275 individuals, we assessed 14 facial features on an ordered categorical scale reflecting the distinctiveness of each trait (Fig. 1, Supplementary Table 2). We included features of the lower face: chin shape, chin protrusion and upper/lower lip thickness; the middle face: cheekbone protrusion, breadth of nasal root, bridge and wing, columella inclination, nose protrusion, nose profile and nose tip shape; and the upper face: brow-ridge protrusion and forehead profile. These features were selected based on their documented variation in Europeans\(^9\). We found them to be reliably scored (Supplementary Table 3) and to also show extensive variation in the CANDELA sample (Supplementary Fig. 1). Individuals were genotyped on Illumina’s OmniExpress BeadChip and imputation performed using 1000 Genomes data. After quality-control filters, final analyses were carried out on 671,038 genotyped single-nucleotide polymorphisms (SNPs) and 9,117,642 imputed SNPs in 5,958 individuals. On the basis of the genome-wide SNP data, average autosomal admixture proportions for the full sample were estimated as: 50% European, 45% Native American and 5% African (Supplementary Fig. 2).

Significant correlations were observed between the ordinal phenotypes (using a Bonferroni-adjusted permutation $P$ value threshold for significance of $6 \times 10^{-4}$, Supplementary Table 4A). Strongest correlation was observed between upper and lower lip thickness ($r = 0.72$), followed by forehead profile and brow ridge protrusion ($r = 0.57$). The three traits related to nose width (root, bridge and wing breadth) show positive correlations among them ($r = 0.16\text{–}0.37$) and negative correlations with nose protrusion ($r = -0.08$ to $-0.25$). Several of the facial traits examined also show moderate (and significant) correlations with age, sex, body mass index (BMI) and genetic ancestry (Supplementary Table 4B). The strongest correlation with sex was seen for brow ridge protrusion and forehead profile ($r = -0.62$ and $r = -0.47$, respectively). Age correlates most strongly with upper and lower lip thickness ($r = -0.19$ and $r = -0.24$, respectively), while the strongest correlation for BMI was seen with brow-ridge protrusion ($r = 0.17$). Genetic ancestry has strongest correlation with lip thickness (European ancestry being negatively correlated with upper and lower lip thickness, $r = -0.25$ and $r = -0.16$, respectively). European ancestry is also significantly correlated with all the nose features examined, particularly with nose protrusion ($r = 0.18$) and nose wing breadth ($r = -0.15$). On the basis of a kinship matrix derived from the SNP data\(^14\), we estimated narrow-sense heritability for the facial traits using GCTA\(^15\). We found moderate (and significant) values for all traits, with the highest heritability being estimated for nose protrusion ($0.47$) and the lowest for columella inclination ($0.20$; Supplementary Table 5). Similar (or higher) heritabilities have been estimated for a range of facial traits using family data\(^7,8,16\).

**GWAS for ordinal phenotypes.** We performed genome-wide association tests using multivariate linear regression, as implemented in PLINK\(^17\), using an additive genetic model adjusting for: age, sex, BMI and the first five principal components (PCs, Supplementary Fig. 3) computed from the SNP data. The resulting statistics showed no evidence of residual population stratification for any of the traits (Supplementary Fig. 3). Three of the nose traits examined (columella inclination, nose bridge and wing breadth) showed genome-wide significant association ($P$ values $< 5 \times 10^{-8}$) with SNPs in at least one genomic region (Fig. 1, Table 1). Columella inclination and nose bridge breadth show association with SNPs in a single region (4q31 and 6p21, respectively), while nose wing breadth shows association with SNPs in two genomic regions (7p13 and 20p11). To account for the multiple phenotypes tested, we performed a global false-discovery rate test across all traits and SNPs and identified the same significantly associated regions (Supplementary Table 6). We examined association for each index SNP (the variant with the lowest $P$ value in a chromosomal region; Table 1) in all countries sampled separately and combined results as a meta-analysis using METAL (Supplementary Table 7) (ref. 18). For all associations, significant effects were in the same direction in all countries, the variability of effect size across countries reflecting sample size (Fig. 2). There was no significant effect size heterogeneity across countries for any of the associations. To exploit the correlations observed between various facial traits, we performed a multivariate GWAS\(^19\), but this approach did not identify any additional associated regions (Supplementary Table 8).

**Follow-up analyses.** Subsequent to the GWAS described above, we obtained data from an additional set of 501 individuals from the same countries as for the GWAS and used this as a replication sample (descriptive features of this sample are presented in Supplementary Fig. 4). These individuals were phenotyped and genotyped as for the GWAS sample. Association tests for the four index SNPs in Table 1 were performed using the same regression model as for the GWAS, with a Bonferroni-adjusted threshold for significance of $0.05/4 = 0.0125$. All tests were found to be significant in this replication sample (Table 1).

We also followed-up the ordinal facial trait GWAS by obtaining facial measurements (distances and angles) related to...
Ordinal traits:

Lower face:
- Upper lip
- Lower lip
- Chin shape
- Chin protrusion

Middle face:
- Cheekbone protrusion
- Nasal root breadth
- Nose bridge breadth
- Nose tip breadth
- Nose tip shape
- Nose profile
- Columella inclination
- Nose protrusion

Upper face:
- Forehead profile
- Brow ridge protrusion

Quantitative traits:

- Lower face:
  - Upper lip
  - Lower lip
  - Chin protrusion
- Middle face:
  - Forehead profile
  - Brow ridge protrusion
- Upper face:
  - Forehead profile
  - Brow ridge protrusion

Figure 1 | Overview of GWAS for facial features in the CANDELA sample. We first carried out a GWAS using data for 14 ordinal facial features from the lower, middle and upper face in 5,958 individuals. For follow-up, we obtained quantitative proxies for 9 of the 14 ordinal traits initially examined (and also obtained a measure of nasion position) in a subset of 2,955 individuals, and performed another GWAS. For convenience, we summarize results across traits on a single ‘composite’ Manhattan plot shown at the bottom of the figure (ordinal traits on the left and quantitative traits on the right). Each Manhattan plot displays all the SNPs with $P$ values exceeding thresholds for genome-wide suggestive ($10^{-5}$, blue line) or genome-wide significance ($5 	imes 10^{-8}$, red line) for any trait. To avoid cluttering the figure, $P$ values not reaching the suggestive threshold (that is, whose significance can be disregarded) are shown only for one trait (upper lip thickness). The names of the candidate genes closest to each association peak are provided (Table 1). These genes are connected with the list of associated facial features via lines of different colour. The location of these features is illustrated on the face drawings shown at the top of the figure. Face drawings were prepared by Emiliano Bellini. PAR, pseudo-autosomal region.

Table 1 | Properties of index SNPs in chromosomal regions showing genome-wide significant association to ordinal facial traits.

| Chromosomal region | Index SNP | Associated trait | $P$-value | Candidate gene | Alleles | Effect size | Percentage of variance explained | Replication $P$-value |
|--------------------|-----------|-----------------|-----------|----------------|---------|-------------|-------------------------------|---------------------|
| 4q31               | rs12644248 | Columella inclination | $7 \times 10^{-9}$ | **DCHS2** | A>G | $-8.40 \times 10^{-2}$ | 0.49 | $4 \times 10^{-3}$ |
| 6p21               | rs1852985  | Nose bridge breadth | $6 \times 10^{-10}$ | **SUPT3H/RUNX2** | C>T | $6.90 \times 10^{-2}$ | 0.71 | $5 \times 10^{-3}$ |
| 7p13               | rs17640804 | Nose wing breadth | $9 \times 10^{-9}$ | **GLI3** | C>T | $-6.50 \times 10^{-2}$ | 0.62 | $6 \times 10^{-3}$ |
| 20p11              | rs927833   | Nose wing breadth | $1 \times 10^{-9}$ | **PAX1** | T>C | $-7.70 \times 10^{-2}$ | 0.66 | $4 \times 10^{-3}$ |

SNP, single-nucleotide polymorphism.
*For intragenic SNPs, gene names are shown in bold.
†Derived alleles are shown after ancestral alleles.
the ordinal traits initially examined and performing a GWAS on these quantitative data. These measurements were obtained mainly using three-dimensional (3D) anatomical landmark coordinates available for 2,955 of the individuals included in the ordinal trait GWAS (Supplementary Fig. 5a). These landmarks allowed us to define quantitative proxies for seven of the ordinal facial traits, the other traits having no appropriate 3D landmarks allowing related measurements to be obtained (Supplementary Table 9). Since the ordinal assessment of nose root and bridge breadth produced genome-wide significant associations (but could not be measured with the 3D landmarks available), we carried out 2D landmarking of the frontal photographs of these 2,955 individuals and also obtained measurements for these two traits (Supplementary Table 10, Supplementary Fig. 5b). In addition, we used the 3D landmark coordinates to obtain a measure of nasion position so as to evaluate in our sample the reported association of this feature with SNPs in the PAX3 gene region (11).

The ordinal variables showed a moderate-to-high (and significant) correlation with the quantitative variables (all permutation P values < 0.0005; Supplementary Table 11 and Supplementary Fig. 6). Correlation between ordinal and quantitative traits was strongest for nose wing breadth and lower lip thickness (both with \( r = 0.70 \)) and lowest for columella inclination (\( r = 0.16 \)). The pattern of correlation among quantitative traits was similar to that observed for the ordinal traits, as was the correlation between quantitative traits and covariates (Supplementary Table 12). As expected for continuous variables, heritability estimates based on the quantitative phenotypes (Supplementary Table 13) are higher than obtained for the ordinal phenotypes and more in line with published estimates (7,8,16).

As before, we performed a GWAS for the quantitative traits using an additive multivariate regression model adjusting for age, sex, BMI and the first five PCs. We replicated the reported association of nasion position with SNPs in 2q35 overlapping the PAX3 gene region, with strongest association seen for rs7559271 (P value of \( 4 \times 10^{-11} \), Fig. 1, Table 2, Supplementary Fig. 7a). This is the same SNP producing strongest association in the Paternoster et al. GWAS. In addition, we observed genome-wide significant association for six of the nine quantitative proxies of the ordinal traits initially examined (Fig. 1, Table 2). As for the ordinal assessments, the quantitative analysis of columella inclination, nose bridge breadth and nose wing breadth produced genome-wide significant associations with SNPs in 4q31, 6p21 and 7p13, respectively (Fig. 1, Tables 1 and 2). In addition, the 4q31 region also showed genome-wide significant association to two other measurements related to nose morphology: nose protrusion and nose tip angle, with strongest P values of \( 4 \times 10^{-8} \) and \( 3 \times 10^{-8} \), respectively. SNPs in 4q31 produced small but not genome-wide significant P values in the ordinal assessment of nose protrusion and nose tip angle (strongest P values of \( 4 \times 10^{-4} \) and \( 3 \times 10^{-4} \), respectively). The 20p11 region, showing genome-wide significant association in the

![Figure 2 | Effect sizes (regression coefficients) for the derived allele at index SNPs in the genome regions associated with ordinal face traits.](image-url)

(a) 4q31 rs12644248, (b) 6p21 rs1852985, (c) 7p13 rs17640804, (d) 20p11 rs927833. Estimates obtained in each country are shown as blue boxes. Red boxes indicate estimates obtained in the meta-analysis. Box size is proportional to sample size. Horizontal bars indicate confidence intervals representing 2 \( \times \) standard errors. Intervals that include zero (that is, non-significant effects) are shown in light blue.
Table 2 | Properties of index SNPs in regions showing genome-wide significant association to quantitative facial traits.

| Chromosomal region | Index SNP | Associated trait | P-value | Candidate gene* | Alleles† | Effect size | Percentage of variance explained |
|---------------------|-----------|------------------|---------|-----------------|----------|------------|----------------------------------|
| 2q12                | rs3827760 | Chin protrusion  | $4 \times 10^{-10}$ | EDAR    | A>G      | $-7.60 \times 10^{-3}$ | 1.32 |
| 2q35                | rs7559271 | Nasion position  | $4 \times 10^{-11}$ | PAIX3   | A>G      | $8.20 \times 10^{-2}$ | 1.33 |
| 4q31                | rs2045323 | Columella inclination† | $3 \times 10^{-9}$ | DCHS2   | G>A      | $1.80 \times 10^{-2}$ | 0.63 |
| 7p13                | rs17640804 | Nose wing breadth | $5 \times 10^{-10}$ | GLI3    | C>T      | $-4.90 \times 10^{-4}$ | 1.15 |

SNP, single-nucleotide polymorphism.
*For intragenic SNPs, gene names are shown in bold.
†Derived alleles are shown after ancestral alleles.
‡Columella inclination was measured as an angle which decreases at greater ordinal columella inclination (Supplementary Table 9, Supplementary Fig. 6f). Therefore, the allelic effects for the quantitative and ordinal assessments of this trait (Table 1) are of opposite sign. rs12644248, the index SNP associated with categorical columella inclination has a P-value of $4 \times 10^{-15}$ for association with the quantitative assessment of columella inclination.

and overlap the 3'-half of the EctodysplasinA (EDA) receptor gene (EDA; Fig. 3a). The derived G allele at the index SNP in this region (rs3827760) encodes a functional substitution in the intracellular death domain of EDA (370A) and is associated with reduced chin protrusion (Table 2). EDA is part of the EDA signaling pathway (comprising EDA, EDAR and EDARADD (the EDAR-binding death domain adaptor protein)) which specifies prenatally the location, size and shape of ectodermal appendages (such as hair follicles, teeth and glands).23 The death domain has been shown to be involved in the interaction of EDAR with EDARADD, the 370A form having higher activity than the ancestral variant.24 The G allele at rs3827760 is not present in Europeans and Africans but is seen at high frequency in East Asians and is essentially fixed in Native Americans (Table 3). This SNP has been associated in East Asians with characteristic tooth morphologies, hair type and sweat gland density.25–27 Recently, we showed, in the same study sample examined here, that rs3827760 impacts on aspects of pinna morphology, including lobe size and attachment, ear protrusion and helix rolling.12 Mutations in the EDA pathway cause hypohidrotic ectodermal dysplasia.28 This disorder is characterized by a reduced number of sweat glands, oligodontia, decrease in the amount of hair and facial dysmorphism, including a markedly protrusive chin.29

Mouse Edar mutant and transgenic lines with either abolished or increased expression of Edar have been described and these mice show features related to several of the phenotypes associated with Edar in humans.12,30,31 Of particular interest, we recently documented that these mice show changes in ear morphology consistent with the effects of Edar on human ear shape variation.12 We therefore compared mandible length in Edar wild-type mice with Edarnull and Edar19951 mutant mice (Supplementary Figs 8 and 9), which have a loss and a gain of Edar function, respectively.31,32 We found a significant association of mandible length with genotype, with the length decreasing at greater Edar function, consistent with the association of the 370A variant with decreased chin protrusion detected in the CANDELA sample (Fig. 4, Supplementary Table 16). Consistent with the mandible length changes we detect in Edar mutant lines, it has been reported that Eda mouse mutants also show mandibular morphology alterations.33 The impact of the Eda pathway on mandibular morphology has been interpreted as resulting from epithelial–mesenchymal interactions during mouse craniofacial development.33

SNPs in the 4q31 region with P values above the suggestive association threshold in the ordinal trait assessment of columella inclination extend over ~400 kb from the 3'-half of the Dachshous Cadherin-Related 2 gene (DCHS2) into the DCHS2–SFRP2
Figure 3 | Genomic regions showing genome-wide significant association to face traits. For each facial feature we show the results that achieved strongest statistical significance regardless of the type of variable analysed (ordinal, O; or quantitative, Q). (a) 2q12 (Q), (b) 4q31 (O), (c) 4q31 (Q), (d) 6p21 (O), (e) 7p13 (Q), (f) 20p11 (O). Plots not shown here are shown in Supplementary Fig. 7. Association results (on a $-\log_{10} P$ scale; left y-axis) are shown for SNPs ~500 kb on either side of the index SNP (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. Below each region we also show an LD heatmap (using $r^2$, ranging from red indicating $r^2 = 1$ to white indicating $r^2 = 0$) produced using a MATLAB implementation similar to Haploview.

Table 3 | Population frequency of derived alleles at index SNPs associated with facial features in the CANDELA sample.

| Region | SNP     | Allele | CEU (%) | YRI (%) | CHB (%) | NAM (%) | CAN (%) |
|--------|---------|--------|---------|---------|---------|---------|---------|
| 2q12   | rs3827760 | G      | 0       | 0       | 94      | 98      | 42      |
| 2q35   | rs7559271 | G      | 39      | 57      | 62      | 90      | 62      |
| 4q31   | rs12644248| G      | 0       | 2       | 15      | 46      | 26      |
| 4q31   | rs2045323 | A      | 9       | 3       | 18      | 66      | 34      |
| 6p21   | rs1852985 | T      | 13      | 20      | 24      | 55      | 30      |
| 7p13   | rs17640804| T      | 72      | 82      | 96      | 36      | 61      |
| 20p11  | rs927833  | C      | 92      | 36      | 95      | 59      | 77      |

SNP, single-nucleotide polymorphism.
*CEU, YRI, CHB are Europeans, Yoruba and Chinese from the 1000 genomes project. NAM are Native Americans and CAN is the CANDELA sample examined here. NAM data are from populations included in Reich et al. [20].
The 6p21.1 region associated with nose bridge breadth extends across ~500 kb overlapping the suppressor of Ty 3 homologue (S. cerevisiae, SUPT3H) gene and the 5′-half of the Runx-related transcription factor 2 (RUNX2) gene (Fig. 3d). Strongest association is seen for SNPs in the region of SUPT3H/RUNX2 overlap, peaking at SNP rs1852985 for both the ordinal and the quantitative assessment of nose bridge breadth (Fig. 3d, Supplementary Fig. 7). This region is known to contain key RUNX2 regulatory elements.27 (Supplementary Fig. 12). Rare mutations in RUNX2 cause Cleidocranial dysplasia, an autosomal dominant disorder involving alterations of cranial ossification (OMIM #119600). Runx2 has been shown to participate in the differentiation of mouse osteoblasts, chondrocyte and mesenchymal stem cells and bone development38, null Runx2 mutants showing a range of chondrocyte proliferation and maturation defects39. Interestingly, the length of a functional glutamine/alanine repeat in Runx2 has been shown to correlate strongly with the evolutionary facial length in dog breeds and, more broadly, in Carnivora40.

SNPs in the 7p13 region associated with nose wing breadth extend over ~80 kb within the third intron of the GLI Family Zinc-Finger 3 gene (GLI3; Fig. 3e), a DNA-binding transcription factor. Strongest association for both the ordinal and quantitative assessments of nose wing breadth is observed for SNP rs17640804 (Tables 1 and 2, Fig. 3e, Supplementary Fig. 7), located in a genomic region with strong evolutionary conservation (Supplementary Fig. 13). Chromatin immunoprecipitation experiments have shown that rs17640804 can affect the binding of regulatory proteins.41 GLI3 is known to act both as activator and repressor in the sonic hedgehog signalling pathway, a key regulator of chondrocyte differentiation42. Interestingly, it has been shown experimentally that GlI3 interacts with Runx2 in the regulation of mouse osteoblast differentiation43. We therefore tested for statistical interaction between the GLI3 and RUNX2 index SNPs on nose bridge breadth and found it to be significant (P value = 0.004, Supplementary Table 17), even though the GLI3 index SNP by itself does not have a significant effect on nose bridge breadth. Mutations in GLI3 have been shown to cause several Mendelian disorders associated with craniofacial and limb abnormalities, including GCPS (Greig cephalopolysyndactyly syndrome). GCPS is characterized by a range of craniofacial abnormalities including a broad nose44. A mouse null GlI3 mutant has been reported to show a range of craniofacial abnormalities, including a wider nose45.

Strongest association in 20p11 with the ordinal assessment of nose wing breadth was observed for SNP rs927833 located in LOC100270679, a long intergenic non-protein coding RNA (LINC01432). There is substantial LD around this SNP and suggestive evidence of association (that is, P values < 10^-5), for SNPs over a region of ~400 kb extending to the Paired-box gene 1 (PAX1; Fig. 3f), a strong candidate gene in this region. PAX1 is a key developmental transcription factor which has been shown experimentally to affect chondrocyte differentiation through its participation in a regulatory pathway that also includes RUNX2 and SOX9 (ref. 46). More broadly, a Pax-Six-Eya-Dach (Dachshund) network, involving protein−protein and protein−DNA interactions impacting on a range of basic developmental processes has been described.47 As indicated above, another PAX gene (PAX3) has been twice reported to impact on nasion position48,49, and we replicate that association here. A missense mutation in PAX1 has been shown to cause autosomal recessive oto-facio-cervical syndrome, a disorder characterized by various skeletal and facial abnormalities48. It has also been reported that mouse embryos with GlI3-null mutations display drastically...
reduced Pax1 expression, possibly mediated through Gli3’s involvement in the sonic hedgehog signalling pathway. Consistent with these experimental findings, we observe a significant statistical interaction of the Gli3 and Pax1 index SNPs on nose wing breadth (P value = 0.005, Supplementary Table 17).

**Discussion**

Since quantitative traits are expected to provide higher power for detecting genetic effects than categorical traits, most recent efforts to identify genes for facial features have focused on quantitative assessments from 3D image data. However, thus far the use of these phenotyping tools has not resulted in many robust genetic finds, mainly the Pax3-nasion position association replicated here. Rather surprisingly, because of the hair type and tooth morphology, the use of these phenotyping tools has not resulted in many robust genetic finds, mainly the Gli3 and Pax1 index SNPs on nose wing breadth (P value = 0.005, Supplementary Table 17).

Consistent with these experimental findings, we observe a significant statistical interaction of the Gli3 and Pax1 index SNPs on nose wing breadth (P value = 0.005, Supplementary Table 17).

**Methods**

**Study subjects.** In all, 6,275 volunteers from 5 countries (Colombia, N = 1,400; Brazil, N = 658; Chile, N = 1,760; Mexico, N = 1,200; and Peru, N = 1,255), part of the CANDELA consortium sample (http://www.srl.ac.uk/silva/candela/), aged between 18 and 45 years were included in this study (Supplementary Table 1). Ethics approval was obtained from: Universidad Nacional Autónoma de México (Mexico), Universidad de Antioquia (Colombia), Universidad Peruana Cayetano Heredia (Peru), Universidad de Tarapacá (Chile), and University of Brasilia. All participants provided written informed consent. Individuals with dysmorphologies, a history of facial surgery or trauma, or with BMI over 33 were excluded (due to the effect of obesity on facial features). Blood samples were collected by a certified phlebotomist and DNA extracted following standard laboratory procedures. Subsequent to the GWAS, an additional 501 individuals were recruited to serve as a replication sample (Supplementary Fig. 4). These individuals were recruited following the same procedures as for the sample included in the GWAS.

**Ordinal phenotyping.** This was carried out in the same way for the GWAS and replication samples. Right side and frontal photographs were used to score 14 facial

**Diff**
traits. This included: chin shape and protrusion, cheekbone and brow-ridge protrusion, forehead profile, upper and lower lip thickness and seven nose features (breadth of nasal root, bridge and wing, columella inclination, nose profile and nose tip shape). These features were selected based on their reported variation in European populations4. Software to assist scoring of photographs was developed in MATLAB (ref. 59). Intraclass correlation coefficients (ICCs)60 calculated for repeated scoring of photographs of the subjects by two independent raters (M.F.-G. and I.P.A.) indicate a moderate–to-high intra-rater reliability of the trait scores (Supplementary Table 3), with relatively lower inter-rater reliability for certain traits. Photographs for all the volunteers were scored by the same rater (M.F.-G.).

Quantitative phenotyping. Quantitative phenotypes were obtained using Procrustes-adjusted 3D facial landmark coordinates available for 2,955 of the individuals included in the ordinal trait GWAS. These coordinates were obtained for 34 anatomical landmarks as detailed in ref. 20 (Supplementary Fig. 5). Briefly, landmarks were placed and raw 3D coordinates obtained using Photomodeler software and five facial photographs (taken at 0°, 45°, 90°, 135° and 180°, where 0° is the left side view). The raw 2D landmark coordinates were Procrustes-adjusted using MorphoJ61. Quantitative measurements (distances and angles) were defined corresponding to seven of the ordinal traits initially examined (Supplementary Table 9). Since no 3D landmarks allowing quantitative proxies for nose root and bridge breadth were available we placed 2D landmarks on the frontal photographs of the same individuals with 3D landmarks (Supplementary Fig. 5, Supplementary Table 10): two landmarks were added each for nasal root and for nose bridge width, in addition to the major frontally visible 3D landmarks. Since the 3D coordinates are free of head tilts and rotations (thus allowing more accurate measurements) the 2D coordinates were calibrated with reference to the 3D coordinates using corresponding frontal landmarks (having both 2D and 3D coordinates) (Supplementary Fig. 5a,b).

DNA genotyping and quality control. DNA samples from participants were genotyped on the Illumina HumanOmniExpress chip including 730,525 SNPs. PLINK v1.9 (ref. 62) was used to exclude SNPs and individuals with >5% missing data, markers with minor-allele frequency <1%, related individuals (PLink IBD estimate > 0.1), and those who failed the X-chromosome sex concordance check (sex estimated from X-chromosome heterozygosity not matching recorded sex information). After applying these filters 671,038 SNPs and 5,958 individuals (1,303 from Colombia, 608 from Brasil, 1,651 from Chile, 1,165 from Mexico, 1,231 from Peru) were retained for further analysis. Due to the admixed nature of the study sample (Supplementary Fig. 2) there is an inflation in Hardy–Weinberg P values. We therefore did not exclude markers based on Hardy–Weinberg deviation, but performed stringent quality controls at software and biological levels, and checked the genotyping cluster plots for each index SNP manually (Supplementary Fig. 14). The replication sample was genotyped in the same way and the genotyping data submitted to the same quality controls as for the GWAS sample.

SNP genotype imputation. The chip genotype data was phased using SHAPEIT2 (ref. 63). IMPUTE2 (ref. 64) was then used to impute genotypes at untyped SNPs using various haplotype positions from the 1000 Genomes Phase I data. The 1000 Genomes reference data set includes haplotype information for 1,092 individuals across the world for 36,820,992 variant positions. Positions that are monomorphic in 1000 Genomes Latin American samples (CLM, MEX and PUR) were excluded, leaving to 11,025,002 SNPs being imputed in our data set. 48,695 had quality scores < 0.4 and 4800 (2014). The replication sample was genotyped in the same way and the genotyping data submitted to the same quality controls as for the GWAS sample.

Statistical genetic analyses. Narrow-sense heritability (defined as the additive phenotypic variance explained by a Genetic Relatedness Matrix, GRM, computed from the SNP data) was estimated using GCTA65 by fitting an additive linear model incorporating index SNP, and by all index SNPs altogether, were estimated from linear mixed regression analyses that included age, sex and BMI as covariates. The GRM was obtained using the LDAK approach66, which incorporates population structure and allows for genome-wide association analyses to be properly accounted for population stratification67. Similar analyses were performed on the imputed data set with two approaches: using the best-guess imputed genotypes in PLINK and using the IMPUTE2 genotype probabilities in SNPTEST v2.5 (ref. 66). Both were consistent with each other and with the results from the chip genotype data. For analysis of the X chromosome an inactivation model was used (male genotypes encoded as 0/2 and female genotypes as 0/1/2). The genetic PCs were obtained (using PLINK v1.9 (ref. 62) from an LD-pruned dataset of 93,328 SNPs. They were selected by inspecting the proportion of variance explained by checking scores and PC scatter plots (Supplementary Fig. 3a). Individual outliers were removed and PCs recalculated after each removal. The top PCs appear to be a good proxy for continental ancestry (Supplementary Fig. 3b). Using these PCs the Q–Q plots (Supplementary Fig. 3c) for all association tests showed no sign of inflation, the genomic control lambda being <1.02 in all cases (Supplementary Fig. 3d), thus confirming that we are appropriately accounting for genetic relatedness.

Mouse analyses. Animal studies were reviewed and approved by The Roslin Institute Animal Welfare and Ethical Review Body (AWERB). 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. The mouse samples and head photographs examined are from the same set described fully in Adhikari et al. Briefly, we included fourteen and 15-day-old animals (17 males and 23 female). The mouse genotypes were EdarD4800 (a loss of function EDARp.E379K mutation52) as either homozygote or heterozygote, wild-type (+/+), and the homozygous Edar null (−/−) line (which has ~16 extra copies of Edar per haploid genome49,35). Thirteen 2D anatomical landmarks were placed on lateral photographs of the mouse heads, using TPSdig and TPSutil (http://life.bio.sunysb.edu/morph/; Supplementary Fig. 8). Generalized procrustes analysis was carried out using the software MorphoJ61 to check whether the distribution of landmarks was homogenous in the 35 associated regions reported in the literature (Supplementary Table 15) and performed individual SNP associations, global Kolmogorov–Smirnov tests and Polygenic Risk Score tests using PLINK.
11. Paternoster, L. et al. Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasiion position. Am. J. Hum. Genet. 90, 478–485 (2012).

12. Adhikari, K. et al. A genome-wide association study identifies multiple loci for variation in human ear morphology. Nat. Commun. 6, 7500 (2015).

13. Ruiz-Linares, A. et al. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS Genet. 10, e1004572 (2014).

14. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. 91, 1011–1021 (2012).

15. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 87, 178–187 (2010).

16. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).

17. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).

18. O’Reilly, P. F. et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS ONE 7, e34861 (2012).

19. Quinto-Sanchez, M. et al. Facial asymmetry and genetic ancestry in Latin American admixed populations. Am. J. Phys. Anthropol. 157, 58–70 (2015).

20. Peng, Q. et al. EDARV370A associated facial characteristics in Uygur population revealing further pleiotropic effects. Hum. Genet. 135, 99–108 (2015).

21. Zeng, C. et al. Genetic determination of human facial morphology: links between cleft-lips and normal variation. Eur. J. Hum. Genet. 19, 1192–1197 (2011).

22. Sadler, A., Viriot, L., Pantalacci, S. & Lauter, V. The ectodysplasin pathway: from diseases to adaptations. Trends Genet. 30, 34–31 (2014).

23. Bryk, J. et al. Positive selection in East Asians for an EDAR allele that enhances NF-kappab activation. Proc. Natl Acad. Sci. 106, 2209–2209 (2008).

24. Park, J. H. et al. Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. Hum. Mutat. 32, 70–72 (2011).

25. Goodwin, A. F. et al. Craniofacial morphometric analysis of individuals with X-linked hypohidrotic ectodermal dysplasia. Mol. Genet. Genomic Med. 2, 422–429 (2014).

26. Headon, D. J. et al. Gene defect in ectodermal dysplasia implicates a death domain adapter in development. Nature 414, 913–916 (2001).

27. Mou, C. et al. Enhanced ectodysplasin-A receptor (EDAR) signaling alters multiple fiber characteristics to produce the East Asian hair form. Hum. Mutat. 29, 1405–1411 (2008).

28. Monreal, A. W. et al. Mutations in the human homologue of mouse dll cause autosomal recessive and dominant hypohidrotic ectodermal dysplasia. Nat. Genet. 22, 366–369 (1999).

29. Bornert, F. et al. Subtle morphological changes in the mandible of Tabby mice revealed by micro-CT imaging and ellipsoidal Fourier quantification. Front. Physiol. 2, 15 (2011).

30. Le Pabic, P., Ng, C. & Schilling, T. F. Fat-Dachsous signaling coordinates subchondral bone development. Front. Physiol. 5, 353–357 (2014).

31. Chang, C. C. et al. Second-generation PLINK: reaching the challenge of larger and richer datasets. Gigascience 4, 7 (2015).

32. O’Connell, J. et al. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet. 10, e1004224 (2014).

33. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. 44, 955–959 (2012).

34. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664 (2009).

35. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11, 499–511 (2010).

36. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. PLoS Genet. 2, 2074–2093 (2006).

37. Pruim, R. J. et al. LocasZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).

38. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploviz: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265 (2005).

39. Reich, D. et al. Reconstructing native American population history. Nature 488, 370–374 (2012).

Acknowledgements

We thank the volunteers for their enthusiastic support for this research. We also thank Ezra Konca, Ivan Pulgar-Alarcon, Alvaro Alvarado, Monica Ballesteros Romero, Ricardo Cebedros, Miguel Angel Contreras Sieck, Francisco de Avila Becerril, Joyce De La Piedra, Maria Teresa Del Solar, Paola Everardo Martinez, William Flores, Martha Granados Lemos, Rosaline Paim, Ricardo Ganzi, Sergeant Joao Felisberto Menezes Xavier, Major Eugenio Correa de Souza Junior, Wendy Hart, Ilich Jafet Moreno, Paola Leon-Mimila, Francisco Quispealaya, Diana Rogel Diaz, Ruth Rojas, Norman Russell and Vanessa Sarabia, for assistance with volunteer recruitment, sample processing and data

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms11616 | www.nature.com/naturecommunications
entry. We are indebted to Emiliano Bellini (www.belliniplastica.blogspot.com) for the illustration in Fig. 1, and to Elizabeth Guajardo Celsi for some of the face drawings used in the phenotyping software. We thank three anonymous reviewers for thoughtful comments that significantly improved this paper. 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 (México); Universidade Federal do Rio Grande do Sul (Brasil); 13ª Companhia de Comunicações Mecanizada do Exército Brasileiro (Brasil); Pontifícia Universidad Católica del Perú, Universidad de Lima and Universidad Nacional Mayor de San Marcos (Perú). This work was funded by grants from the Leverhulme Trust (F/07 134/DF to A.R.-L), BBSRC (BB/I021213/1 to A.R.-L), Universidad de Antioquia (CODI sostenibilidad de grupos 2013–2014 and MASO 2013–2014), Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (Apoio a Núcleos de Excelência Program) and Fundação de Aperfeiçoamento de Pessoal de Nível Superior.

Author contributions
Conceived and designed study: K.A., M.F.-G., J.R., R.G.-J., D.H., D.B., A.R.-L. Contributed reagents/material: M.F.-G., J.M.-R., V.A.A., C.J., W.A., R.B.L., G.M.P., J.G.-V., H.V.-R., T.H., Y.R., C.C.d.C., M.H., V.V., V.G., D.H. Performed experiments: M.F.-G., J.M.-R., J.C.C.-D., V.A.A., M.Q.-S., D.H. Analyzed data: K.A., M.F.-G., M.Q.-S., D.H., A.R.-L. Supervised research (PI): J.G.-V., C.G., G.P., L.S.-F., F.M.S., M.-C.B., S.C.-Q., F.R., G.B., R.G.-J., D.H., D.B. Wrote the manuscript, incorporating input from other authors: K.A., M.F.-G., A.R.-L. Critical revision of the manuscript: R.G.-J., D.H., D.B.

Additional information
Accession codes: The MATLAB program used to perform the ordinal scoring of facial features can be downloaded from http://www.ucl.ac.uk/silva/candela. P values for all SNPs tested in the GWAS analyses will be hosted in GWAS Central (http://www.gwascentral.org/), and also made available through http://www.ucl.ac.uk/silva/candela on the next data release of the GWAS Central database, scheduled for June 2016.

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

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

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

How to cite this article: Adhikari, K. et al. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat. Commun. 7:11616 doi: 10.1038/ncomms11616 (2016).

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 under the license holder to reproduce the material.

To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/