nature genetics

ARTICLES

https://doi.org/10.1038/s41588-020-00741-7

Insights into the genetic architecture of the human face

Julie D. White[1,19] ✉, Karlijne Indencleef[2,3,19] ✉, Sahin Naqvi[4,5], Ryan J. Eller[6], Hanne Hoskens[3,7], Jasmien Roosenboom[8], Myoung Keun Lee[8], Jiarui Li[2,3], Jaaved Mohammed[4], Stephen Richmond[9], Ellen E. Quillen[10,11], Heather L. Norton[12], Eleanor Feingold[13], Tomek Swigut[4], Mary L. Marazita[8,13], Hilde Peeters[7], Greet Hens[14], John R. Shaffer[8,13], Joanna Wysocka[4,15,16], Susan Walsh[6], Seth M. Weinberg[8,13,17], Mark D. Shriver[1] and Peter Claes[2,3,7,18] ✉

The human face is complex and multipartite, and characterization of its genetic architecture remains challenging. Using a multivariate genome-wide association study meta-analysis of 8,246 European individuals, we identified 203 genome-wide-significant signals (120 also study-wide significant) associated with normal-range facial variation. Follow-up analyses indicate that the regions surrounding these signals are enriched for enhancer activity in cranial neural crest cells and craniofacial tissues, several regions harbor multiple signals with associations to different facial phenotypes, and there is evidence for potential coordinated actions of variants. In summary, our analyses provide insights into the understanding of how complex morphological traits are shaped by both individual and coordinated genetic actions.

In 1991, Atchley and Hall epitomized one of the major problems in contemporary biology as the need "to understand how complex morphological structures arise during development and how they are altered during evolution" (p. 102)[1]. This problem continues to captivate biologists, geneticists, anthropologists and clinicians almost three decades later. In their review, the authors describe a 'complicated developmental choreography' in which intrinsic genetic factors, epigenetic factors and interactions between the two make up the progeny genotype, which engages with the environment to ultimately produce a complex morphological trait composed of separate component parts[1]. We now understand that the intrinsic genetic factors ultimately contributing to complex morphological traits consist not only of single variants altering protein structure and/or function, but also noncoding variants and interactions among variants, each affecting multiple tissues and developmental timepoints. This realization requires methods capable of describing the genetic architecture of complex morphological traits, which includes identifying the individual genetic variants contributing to morphological variation and interactions among those variants[2,3].

The human face—an exemplar complex morphological structure—is highly multipartite and results from the intricate coordination of genetic, cellular and environmental factors[4–6]. Through prior genome-wide association studies (GWAS), over 100 loci have been implicated in normal-range facial morphology[7–23] (Supplementary Table 1). However, as with all complex morphological traits, our ability to identify and describe the genetic architecture of the face is limited by our ability to accurately characterize its phenotypic variation[4], identify variants of both large and small effect[15] and identify interactions between variants. We previously described a data-driven approach to facial phenotyping, which facilitated the identification and replication of 15 loci involved in global-to-local variation in facial morphology[16]. Here, we apply this phenotyping approach to two larger cohorts from the United States and United Kingdom (n_{Total} = 8,246; Supplementary Table 2) and apply multivariate techniques to uncover new biological insights into the genetic architecture of the human face. We now identify 203 genome-wide-significant (120 also study-wide-significant) signals, located in 138 cytogenetic bands, associated with multivariate normal-range facial morphology. Many of these loci harbor genes that are involved in craniofacial syndromes but that had not yet been observed in GWAS for normal-range facial morphology; however, 53 genome-wide-significant (26 also study-wide-significant) peaks are located in regions with no previously known role in facial development or disease, potentially pointing to previously unknown genes and pathways involved in facial development. We additionally provide evidence that variants at our genome-wide-significant peaks are involved in regulating enhancer activity in cell types controlling facial morphogenesis across the developmental timeline.

[1]Department of Anthropology, Pennsylvania State University, State College, PA, USA. [2]Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium. [3]Medical Imaging Research Center, UZ Leuven, Leuven, Belgium. [4]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA. [5]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [6]Department of Biology, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA. [7]Department of Human Genetics, KU Leuven, Leuven, Belgium. [8]Department of Oral Biology, Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA, USA. [9]Applied Clinical Research and Public Health, School of Dentistry, Cardiff University, Cardiff, UK. [10]Department of Internal Medicine, Section of Molecular Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [11]Center for Precision Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [12]Department of Anthropology, University of Cincinnati, Cincinnati, OH, USA. [13]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA. [14]Department of Neurosciences, Experimental Oto-Rhino-Laryngology, KU Leuven, Leuven, Belgium. [15]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA. [16]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. [17]Department of Anthropology, University of Pittsburgh, Pittsburgh, PA, USA. [18]Murdoch Children's Research Institute, Melbourne, Victoria, Australia. [19]These authors contributed equally: Julie D. White, Karlijne Indencleef. ✉e-mail: jdw345@psu.edu; karlijne.indencleef@kuleuven.be; peter.claes@kuleuven.be

Furthermore, we reveal interactions between variants at different loci affecting similar aspects of facial shape variation, identifying gene sets that work in concert to build human faces. With this work, we not only push forward our understanding of human facial genetics, but also illustrate the potential for researchers to confront Atchley and Hall's problem, by intensively characterizing complex morphological variation and using advanced methods to identify factors involved in the developmental choreography of complex morphological structures.

Results

Multivariate phenotyping and meta-analysis framework. To study facial variation at both global and local scales, we start with a set of three-dimensional (3D) facial surface scans, upon which we map a dense mesh of 7,160 homologous vertices[24]. We then apply a data-driven facial segmentation approach, defined by grouping vertices that are strongly correlated using hierarchical spectral clustering[16,25]. The configurations of each of the resulting 63 segments are then subjected independently to a Generalized Procrustes analysis, after which principal component analysis (PCA) is performed in conjunction with parallel analysis to capture the major phenotypic variation in each facial segment[26,27] (Extended Data Fig. 1). The number of principal components (PCs) kept at this stage of the analysis ranged from 7 to 70, with segments containing large numbers of quasi-landmarks generally requiring more PCs to describe the variation in that segment. The inherent shape variability in each segment also plays a role in the number of PCs retained by parallel analysis, with more variable segments retaining more PCs. For example, although segments 5 and 25 contain similar numbers of quasi-landmarks, because the variability of the nose (segment 5) is generally greater than that of the lower cheeks (segment 25), the parallel analysis for segment 5 retained 32 PCs while for segment 25 it retained only 20 PCs (Extended Data Fig. 1b).

We then tested for genetic association between the facial PCs and 7,417,619 SNPs by using a data-driven approach (Extended Data Fig. 2). Within each segment, instead of a priori selecting the PCs of interest, or treating each of the 63 segments as a single 'trait', we use canonical correlation analysis (CCA) to first identify the linear combination of components in each segment maximally correlated with the SNP being tested in the identification cohort. We call this multivariate combination of PCs the 'trait'. Thus, each SNP is associated (although not always with significance) with its own 'trait' in each segment. Subsequently, the verification cohort is projected onto each of these traits, creating univariate 'phenotype' variables that are tested for genotype–phenotype associations by using linear regression. The projection ensures that the shape variation tested in the verification step is equivalent to the 'trait' used in the identification step. The identification and verification *P* values are then meta-analyzed using Stouffer's method[28,29]. The whole process is then repeated, switching the dataset used for identification and verification, thereby resulting in 126 meta-analysis *P* values and traits (63 segments × 2 meta-analysis tracks) for each SNP. Further details are available in the Methods and Supplementary Notes 1 and 2.

Sharing of genome-wide signals between facial segments. We first assessed the degree to which variation in each facial segment shares the same patterns of association across the genome by computing the linkage disequilibrium score correlation (LDSC) based on genome-wide-association *P* values for each pair of facial segments[30,31]. This 63 × 63 matrix of correlations was visualized on top of the facial segmentation hierarchy to assess between-segment correlations within and between facial quadrants (Extended Data Fig. 3), although it is important to note that these LDSCs should not be considered 'genetic correlations' in the typical way of a univariate trait, since the *z*-scores used are unsigned. The LDSCs were highest between segments of the same facial quadrant (that is, lips, nose, lower face, upper face), validating the hierarchical clustering used to initially define the segments (Extended Data Fig. 3b). Average-linkage hierarchical clustering of the facial segments based on the correlation values gave rise to four main clusters, each corresponding primarily to segments from the same quadrant (Extended Data Fig. 4). Despite substantial within-quadrant similarity, there were notable correlations between groups of segments from different quadrants (Extended Data Fig. 3a). Some of these specific correlations reflect close physical proximity of the segments in different quadrants (for example, segments 12 and 33), but some correlations seem to reflect the shared embryological origins of groups of segments. Specifically, segments representing the nose (quadrant II) and upper face (quadrant IV) cluster together, and most segments representing the lips (quadrant I) and lower face (quadrant III) cluster together (Extended Data Fig. 4). Quadrants II and IV together approximate the frontonasal prominence, which appears earlier in development than the mandibular and maxillary prominences, which are approximated by quadrants I and III, respectively[32].

Genome-wide-association meta-analysis. In total, we identified 17,612 SNPs with *P* values ($P_{\text{Meta-US}}$ and/or $P_{\text{Meta-UK}}$) lower than the genome-wide threshold ($P \leq 5 \times 10^{-8}$). Of these, 11,398 SNPs also passed the study-wide-significance threshold ($P \leq 6.96 \times 10^{-10}$) (Supplementary Fig. 1). For each peak passing the genome-wide threshold, we designated the SNP with the lowest *P* value across all facial segments as the 'lead SNP', refining our results to 218 genome-wide-significant lead SNPs. Of these, 203 SNPs showed consistent genetic effects on the trait identified in the US- and UK-driven meta-analyses in the facial segment with the lowest *P* value for that SNP (Fig. 1 and Supplementary Table 3), and 120 of these were also below study-wide significance. Visual representations of the LocusZoom[33] and effect plots for each of the 203 genome-wide-significant SNPs are available in the FigShare repository[34].

The global-to-local approach means that we often identified associations between a single SNP and variation in many facial segments. In this article, we focus primarily on the segment in which the SNP had its lowest *P* value (the 'Best segment') and provide information on in which meta-analysis track (Meta-US or Meta-UK) the SNP reached this significance level (the 'Best meta-analysis track'). Thus, throughout the rest of the article, the reported *P* values for each SNP will be in the format of $P_{\text{Best track}}$ (Best segment) = value. By plotting the strongest association results for each segment (Fig. 1, left), segments 1 and 2 are visibly the 'Best segment' for most SNPs, with $n = 20$ SNPs reaching lowest significance in the full face (segment 1) in the US-driven meta-analysis ($n = 15$ for Meta-UK) and $n = 19$ SNPs reaching lowest significance in segment 2 in the US-driven meta-analysis ($n = 18$ for Meta-UK).

Genes near lead SNPs are enriched for both craniofacial and limb development. In a GREAT[35] analysis of the regions surrounding the 203 genome-wide-significant lead SNPs, the top ten terms (based on lowest binomial *P* values) in the mouse phenotype, human phenotype and gene ontology (GO) biological processes categories are all highly relevant to craniofacial shape and overall morphology (Extended Data Fig. 5a), with the top human phenotype being oral clefting. A FUMA[36] analysis of the same regions highlighted genes overlapping several pathways related to abnormal cellular maintenance and also included pathways highly relevant for morphological development, like the Wnt, Hedgehog and TGFβ signaling pathways (Extended Data Fig. 5b).

Facial GWAS peaks are enriched for enhancers specific to cell types across the timeline of facial development. To assess the likely cell types and developmental timepoints in which our GWAS regions are active, we compiled H3K27ac ChIP–seq signals—detecting a

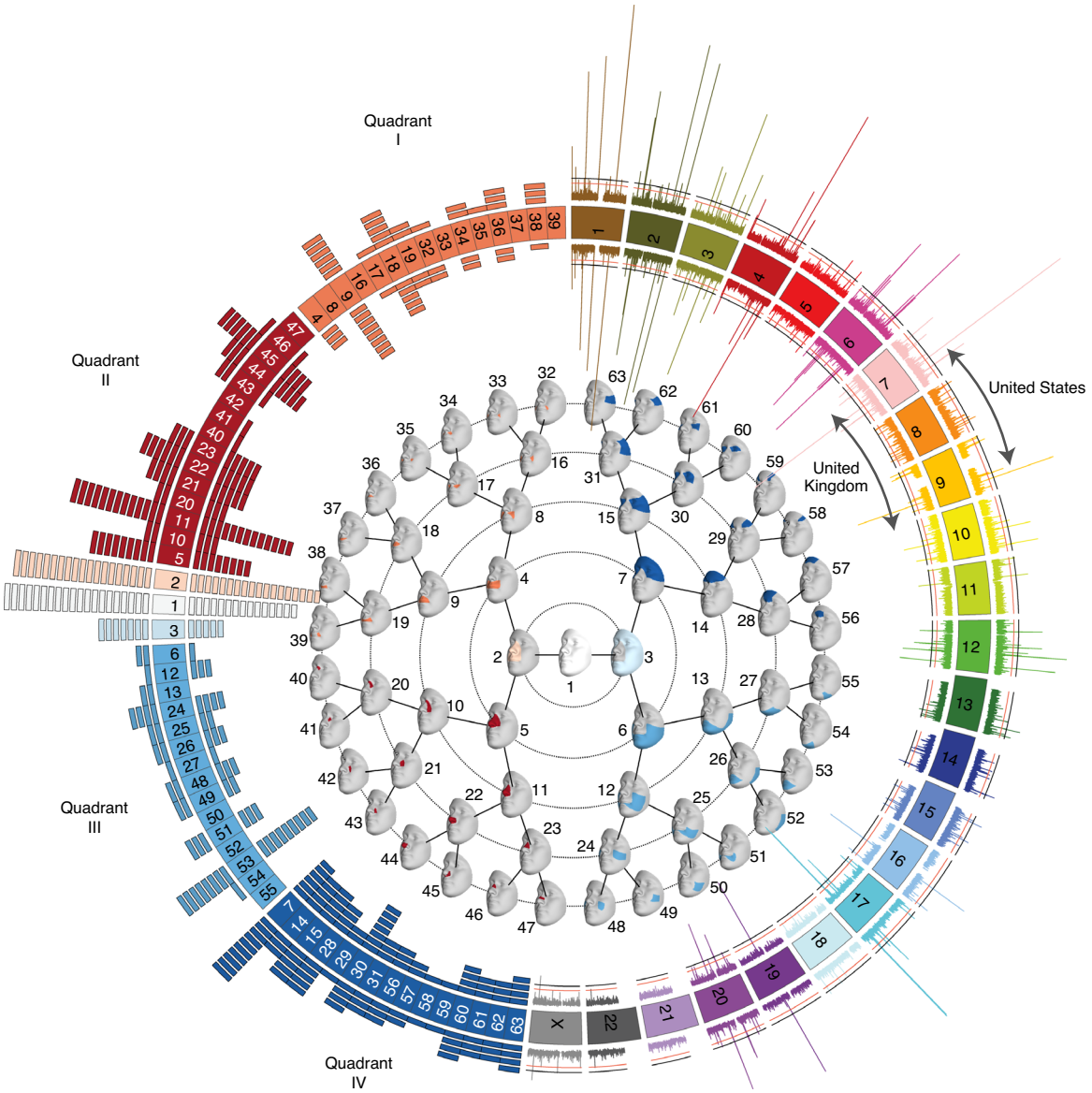


Fig. 1 | Overall results of US-driven and UK-driven meta-analyses. On the left, numbered blocks representing the 63 facial segments arranged and colored according to quadrant (I, orange; II, red; III, light blue; IV, dark blue), the full face (white) and segments 2 (light orange) and 3 (ice blue). The histogram arranged on the left side represents the number of genome-wide-significant lead SNPs reaching their lowest *P* value in each segment, with each rectangle representing one SNP. The US-driven meta-analysis results are on the outside of the circle and the UK-driven meta-analysis results are on the inside of the circle. In the center, the global-to-local facial segmentation of all 3D images included in this analysis, obtained using hierarchical spectral clustering, are colored to match with the quadrants on the left. On the right, a Miami plot of the US-driven meta-analysis *P* values on the outside and the UK-driven meta-analysis *P* values on the inside, with chromosomes colored and labeled. Values plotted are the result of Stouffer's meta-analysis of one-sided right-tailed identification and verification *P* values, detailed in the Methods, and are $-\log_{10}$ scaled (range, 0–80). The red line represents the genome-wide-significance threshold ($P=5\times10^{-8}$) and the black line represents the study-wide threshold ($P=6.96\times10^{-10}$). Created using Circos v.0.69-8 (ref. [53]).

marker of the promoters of transcriptionally active genes and active distal enhancers[37,38]—from approximately 100 different cell types and tissues, including cranial neural crest cells (CNCCs), fetal and adult osteoblasts and mesenchymal stem cell-derived chondrocytes, as well as dissected embryonic craniofacial tissues (Carnegie stages 13–20). Both CNCCs and craniofacial tissues showed the highest H3K27ac signals in the vicinity of the 203 genome-wide-significant lead SNPs, whereas no H3K27ac signal was observed for 203 random SNPs matched for allele frequency and distance to the nearest gene (Fig. 2a). The difference in H3K27ac signal between the 203 genome-wide-significant lead and random SNPs was significant based on a two-sided Wilcoxon rank-sum test for many cell types and tissues, with CNCCs and embryonic craniofacial tissues having the greatest median differences (Extended Data Fig. 6 and Supplementary Table 4).

To distinguish enrichment between coding and noncoding elements, we examined chromatin signals in CNCCs and embryonic craniofacial tissues in more detail, using ChIP–seq data on additional chromatin marks and transcription factors[39,40]. In the CNCCs, candidate regulatory regions in the vicinity of the 203 genome-wide-significant lead SNPs were enriched significantly for strong and intermediate enhancers and depleted in weak promoters (Fig. 2b). In embryonic craniofacial tissue, all developmental stages sampled were significantly enriched for the chromHMM states of

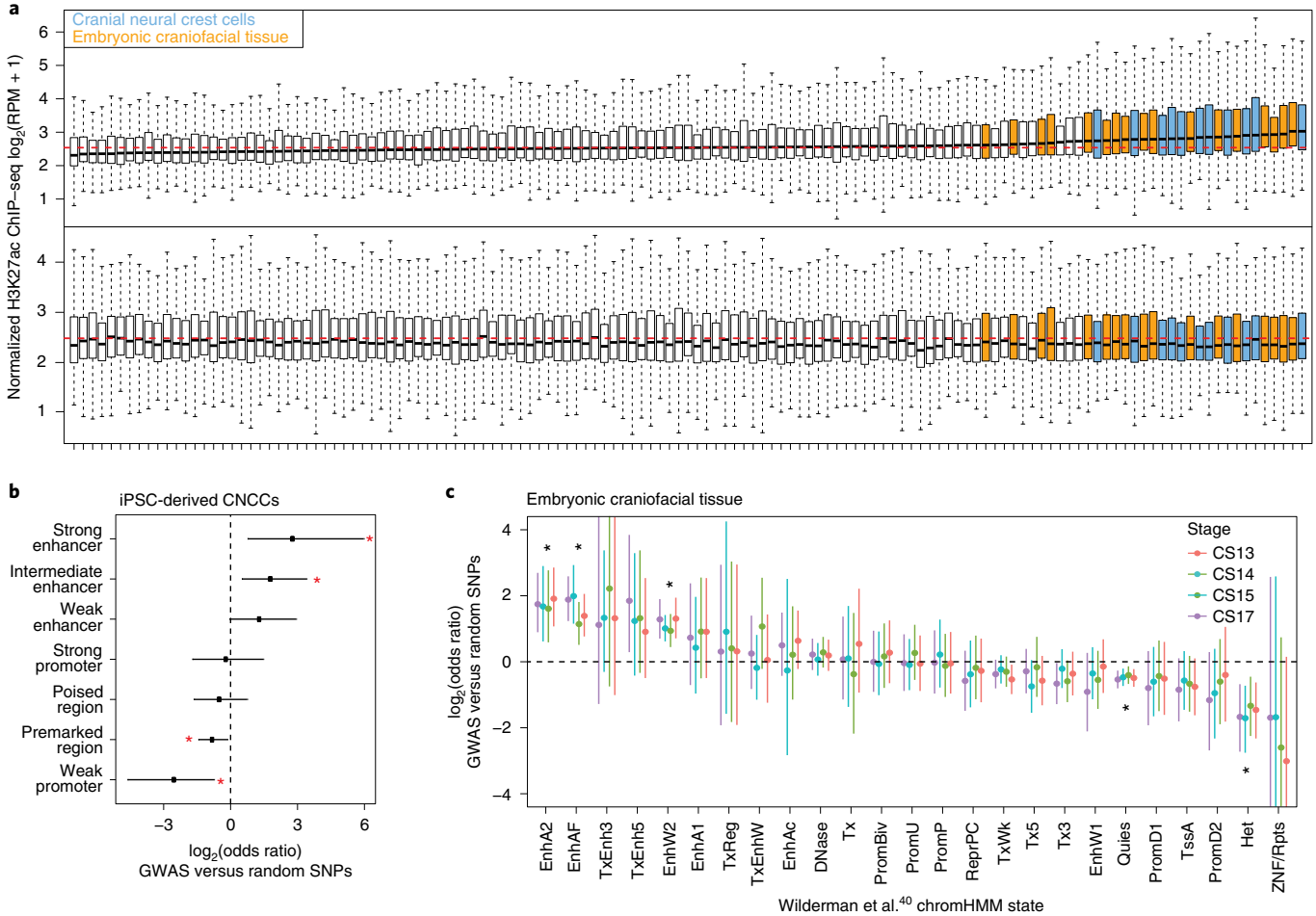


Fig. 2 | Regions near the 203 genome-wide-significant lead SNPs are enriched for enhancers preferentially active in cranial neural crest cells and embryonic craniofacial tissue. a, Each boxplot represents the distribution of H3K27ac signal in 20-kb regions around the 203 genome-wide-significant lead SNPs (top) or 203 random SNPs (bottom) in one sample, with cranial neural crest cells and embryonic craniofacial tissues highlighted. Boxplots plot the first and third quartiles, with a dark black line representing the median. Whiskers extend to the largest and smallest values no further than 1.5× the interquartile range from the first and third quartiles, respectively. The dashed red lines represent the median level of H3K27ac reads per million (RPM) signal across all cell types and tissues. A larger labeled version of **a** is available in the FigShare repository[34]. **b**,**c**, For each class of regulatory element in either CNCCs derived from induced pluripotent stem cells (iPSC) (**b**) or embryonic craniofacial tissue (**c**), the number of elements within 20 kb of the 203 genome-wide-significant lead SNPs was compared to the number within 20 kb of 203 random SNPs by using a two-sided Fisher's exact test. Points represent estimated odds ratio and surrounding bars represent 95% confidence intervals. Asterisks indicate any Benjamini–Hochberg adjusted *P* value < 0.05. For embryonic craniofacial tissue, enrichments were calculated for each Carnegie stage separately, as Wilderman et al.[40] performed chromatin state segmentation for each stage separately. Descriptions of all mnemonics can be found at: https://egg2.wustl.edu/roadmap/web_portal/imputed.html#chr_imp.

active enhancers, active enhancer flanks and weak enhancers, and depleted in quiescent/low and heterochromatin states (Fig. 2c).

Cell-type-specific activity patterns were used to further subdivide the 203 genome-wide-significant lead SNPs by using *k*-means clustering of H3K27ac signals (Fig. 3). As expected, many lead SNPs showed specific activity for CNCCs and craniofacial tissue (for example, cluster 5), representing activity at an early time point in development. Interestingly, however, some SNPs showed preferential activity for either CNCCs or craniofacial tissue (for example, clusters 1 and 2). Greater specificity for CNCCs could arise because CNCCs constitute a relatively small proportion of the cells present in craniofacial tissue at Carnegie stages 13–20, while greater specificity for craniofacial tissue could be due to activity in further differentiated cell types of the face.

Known and new loci. We identified 89 genome-wide-significant (66 also study-wide-significant) peaks that overlap with the results of prior association studies of normal-range facial phenotypes. Of these, 29 genome-wide-significant (20 also study-wide-significant) peaks were reported by studies with overlapping samples as this study and 60 genome-wide-significant (46 also study-wide-significant) peaks were previously reported by studies with completely non-overlapping sample sets. A total of 61 genome-wide (28 also study-wide) significant peaks observed in our analysis are located at loci harboring putative craniofacial genes (implicated from human malformations or animal models), but which had not yet been observed in GWAS for normal-range facial morphology. Our GWAS additionally revealed 53 genome-wide-sigificant (26 also study-wide-significant) peaks at loci harboring genes with no previously known role in facial development or disease. The annotation for each GWAS peak can be found in Supplementary Table 3.

Genomic regions harboring multiple lead SNPs. With our phenotyping and analysis framework, in many cases we are able to provide a more nuanced understanding of the underlying genetic architecture of facial variation. For example, variants at the *TBX15-WARS2*

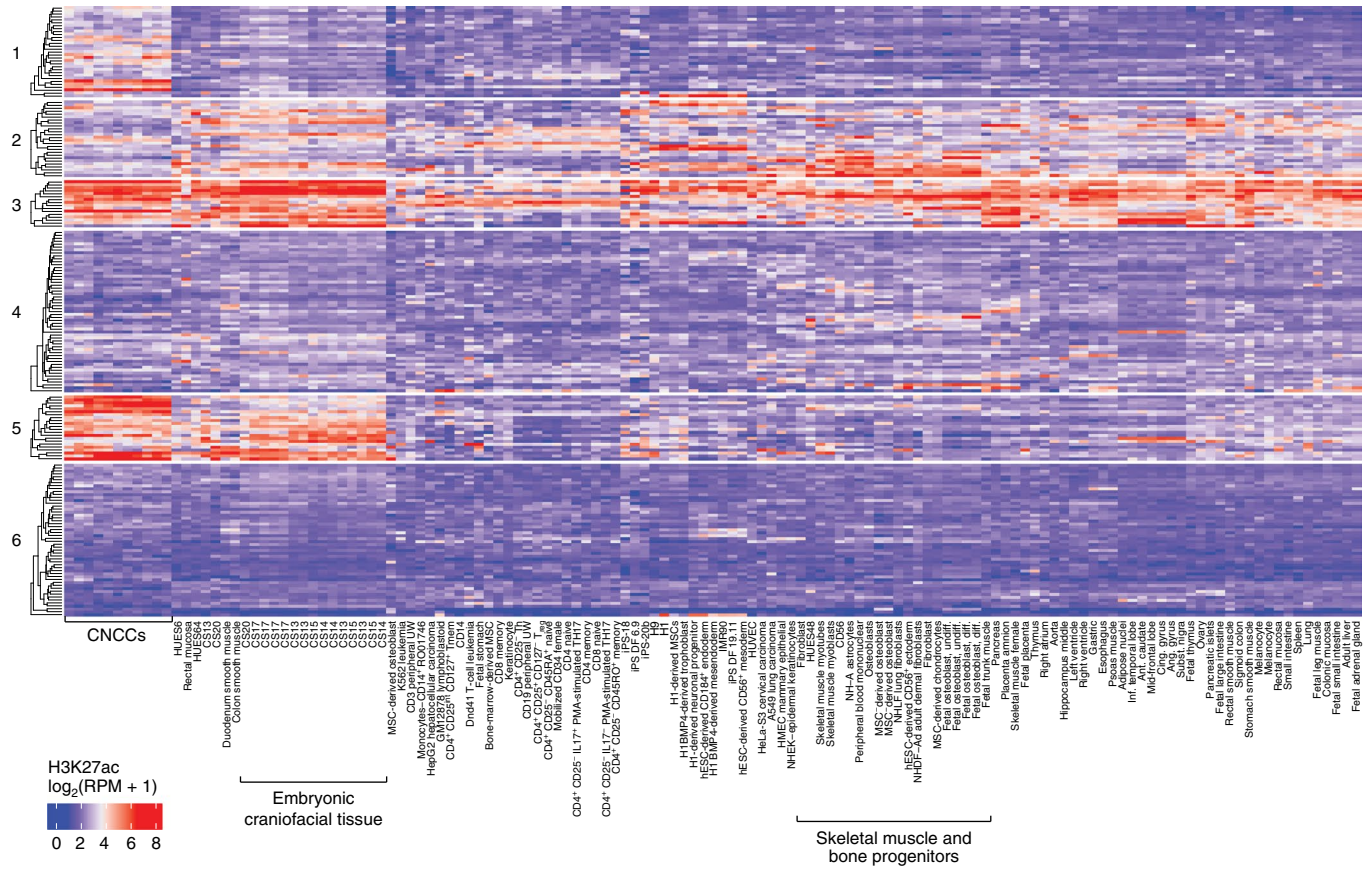


Fig. 3 | Activity of 203 genome-wide-significant lead SNPs in all cell types studied. H3K27ac signal calculation and *k*-means clustering of SNPs were performed as described in Methods. Average-linkage clustering on Euclidean distances was performed both within each of the six row clusters and for all columns. Descriptions of all abbreviations can be found at https://egg2.wustl.edu/roadmap/web_portal/meta.html.

locus (1p12; Fig. 4) were previously reported to be associated with forehead prominence[16] and self-reported chin dimples[11], already indicating that this locus has multiple spatially separated effects on the face. In our current analysis, we see the same influence on forehead morphology, as previously reported by our group[16], with lead SNP rs3936018, located in the promoter region of *WARS2*, reaching its lowest significance in segment 14 ($P_{\text{Meta-US}}$(segment 14) = 8.01×10^{-58}). Interestingly, this lead SNP overlaps in location with a SNP not originally identified in our peak selection approach, rs12027501 ($P_{\text{Meta-US}}$(segment 1) = 1.03×10^{-41}). The latter was most significant in segment 1, the full face, and is not a good proxy for the former (r^2, 0.075; normalized coefficient of linkage disequilibrium D', 0.979), indicating it is likely an independent statistical signal. Another signal, approximately 275 kb upstream of *TBX15* (rs7513680), was most significantly associated with morphology in segment 51 ($P_{\text{Meta-UK}}$(segment 51) = 7.03×10^{-13}), representing the cheek area around the corners of the mouth. Lastly, another GWAS peak is present approximately 301 kb downstream of *WARS2* (rs17023457) with an effect in the upper cheeks ($P_{\text{Meta-UK}}$(segment 48) = 3.26×10^{-15}). Of interest, we observed 24 such loci with multiple genome-wide-significant peaks that are each associated with different facial traits (Supplementary Table 5 and Supplementary Data 1).

Genetic interactions impacting facial variation. To better analyze and rank the effects of multiple genotypes on a facial trait, we utilized structural equation modeling (SEM) to refine our understanding of which groups of genome-wide-significant variants best explain the variance observed in each facial segment. SEM is a multivariate statistical analysis technique that analyzes structural relationships between measured variables (for example, genetic variants and covariates) and latent constructs (univariate phenotypes derived from the PCs of the analyzed facial segment). This was done in an iterative manner, resulting in 50 well-fitting SEM models (corresponding to 50 facial segments; Supplementary Data 2). For each of these 50 models, the output included a univariate latent variable and a list of variants ranked by their estimated contribution to that variable, highlighting the polygenic nature of facial variation captured by the latent variable. Higher correlation of cross-sample H3K27ac activity was found when comparing SNPs deemed significant by the same SEM model than when comparing SNPs nonsignificant in the same SEM model (Extended Data Fig. 7). Additionally, of the SEM-significant SNPs, four SNP combinations displayed evidence of pairwise epistatic interactions (Table 1, Fig. 5, Extended Data Fig. 8 and Supplementary Note 3).

Discussion

In their review, Atchley and Hall[1] provided a framework with which we can better understand and describe the development of complex morphological structures. In this analysis, we have focused on one part of this framework and have identified intrinsic genetic factors contributing to normal-range variation in the structure of the human face. By implementing an open-ended multivariate association method, in which the inherent morphological variation within each of these segments drives the association, and by using both standard and modified-for-multivariate follow-up bioinformatic approaches, we describe the association between SNPs and facial traits as well as the likely cellular functions of the regions

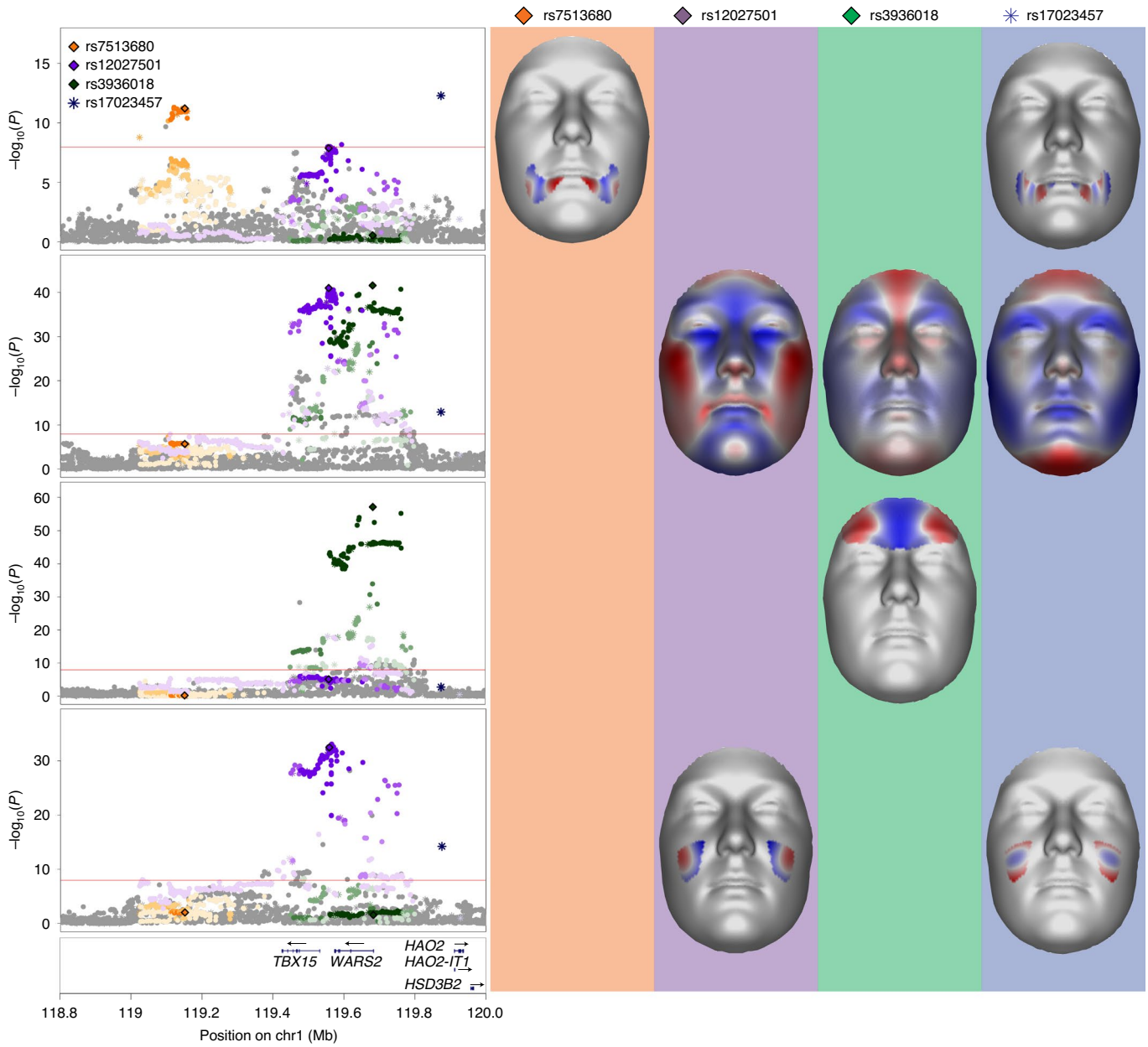


Fig. 4 | *TBX15-WARS2* multi-peak locus. LocusZoom[33] plots and facial effects for four association signals near the *TBX15-WARS2* locus. Clustering based on r^2 was performed to separate noncorrelated signals, resulting in the separation of four SNP clusters. Color for each SNP is based on cluster association, with saturation indicating r^2 correlation with the most significant SNP in the cluster. SNPs represented by diamonds are the genome-wide-significant lead SNPs also present in the 1000G Phase 3 dataset; SNPs represented by circles are adjacent SNPs also present in the 1000G Phase 3 dataset; SNPs represented by asterisks are those not present in the 1000G Phase 3 dataset. For the segment in which each lead SNP had its lowest effect, we plot the facial effects for the lead SNPs reaching significance in that segment as the normal displacement (displacement in the direction normal to the facial surface) in each quasi-landmark going from minor to major allele, with red colored areas shifting outward while blue colored areas shift inward.

surrounding these SNPs. We also highlight regions with multiple SNPs affecting different facial phenotypes as well as evidence for multiple SNPs working in concert to produce a single phenotype. Taken in summary, our results illustrate an avenue for investigating the coordinated processes underlying complex morphological structures, like the human face, at a deeper level than single associations between genotype and univariate phenotype.

Overall, our association results reflect patterns from known biological processes. For instance, linkage disequilibrium (LD) score regression correlations between segments seem to reflect the shared embryological origins of different parts of the face, indicating that the hierarchical spectral clustering of the face based on structural correlations effectively partitions underlying genetic signals into biologically coherent groups. It is additionally clear from the large number of genome-wide-significant SNPs reaching their strongest association in the full face and segment 2 (covering the nose and upper lip) that these facial regions are 'hotspots' for genomic signals (Fig. 1). In general, quadrant II (representing the nose) and quadrant IV (representing the forehead and eyes) had the most genome-wide-significant lead SNPs reaching lowest significance in segments within each quadrant. This is unsurprising, given the close relationship between visible facial features in those areas and

Table 1 | Four SNPs with evidence of epistatic interactions

Segment	SNP 1			SNP 2			Test statistic	*P* value
	rsID	Location	Gene annotation	rsID	Location	Gene annotation		
6	rs10838269	11:44378010	*ALX4*	rs11175967	12:66321344	*HMGA2*	23.9422	9.94×10^{-7}
9	rs76244841	1:2775953	*PRDM16*	rs62443772	7:42131949	*GLI3*	16.5745	4.68×10^{-6}
11	rs6740960	2:42181679	*PKDCC*	rs6795164	3:133885925	*SLCO2A1*	16.3707	5.21×10^{-5}
22	rs7373685	3:128107020	*GATA2*	rs7843236	8:121980512	*SNTB1*	15.7837	7.10×10^{-5}

For each of the 50 segments with a refined SEM model, we used the latent variables and SNP lists to test for evidence of epistasis using a two-sided linear regression epistasis test in Plink v.1.9, with Bonferroni multiple-testing correction. For the four SNP pairs with significant evidence of epistatic interactions, Table 1 lists the epistasis *P* value, rsID, GRCh37 location and gene annotation. The phenotypic and marginal distributions for the pairs are depicted as boxplots in Fig. 5 and Extended Data Fig. 8.

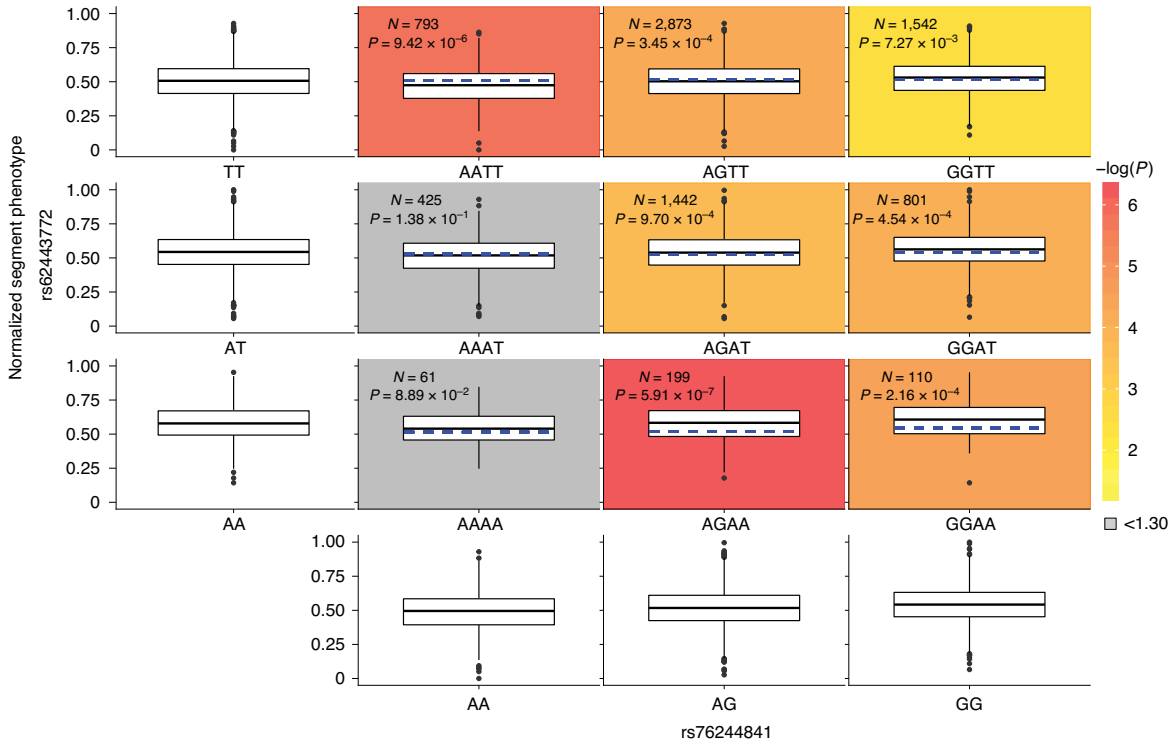


Fig. 5 | Phenotypic and marginal distributions for the rs62443772–rs76244841 epistatic pair. Plotted in the first column and last row are the marginal phenotypic distributions of the genotypes, which shows the phenotypic distribution that would occur if the two genotypes were acting alone. The median phenotype was also calculated for each diplotype as the average of the marginal medians of the singular genotypes (blue dashed lines on the colored plots). The observed diplotype median (black line on the colored plots) was compared to the expected diplotype median (blue dashed lines on the colored blots) via Mood's Median test[54] with one degree of freedom. The resulting log-transformed *P* value was used to color the boxplots to illustrate significance, unless the difference was nonsignificant, in which case the color was automatically set to gray. Within each colored boxplot is the untransformed Mood's median *P* value as well as the number of individuals used for significance testing. Boxplots plot the first and third quartiles, with a dark black line representing the median. Whiskers extend to the largest and smallest values no further than 1.5× the interquartile range from the first and third quartiles, respectively.

the underlying skeletal structure. Indeed, regions with less correspondence to underlying skeletal structure, like the upper lip (quadrant I), had many fewer lead SNPs reaching lowest significance in the contained segments, and facial regions with some structural correspondence but still greatly impacted by age and adiposity, like the lower face and cheeks (quadrant III), had only slightly more.

Reassuringly, the genes located within 500 kb of our genome-wide-significant lead SNPs were highly enriched for processes and phenotypes associated with craniofacial development and morphogenesis in humans and mice (Extended Data Fig. 5). Notably, the top human phenotype was oral clefting, indicating a substantial overlap between the genes involved in normal facial variation and those implicated in the most common craniofacial birth defect in humans. Furthermore, many of the surrounding genes to which the genome-wide-significant lead SNPs were annotated are known to be involved in pathways relevant for craniofacial development, such as the Wnt signaling and TGFβ pathways (Extended Data Fig. 5b). Our GWAS signals were also enriched for processes associated with limb development and related phenotypes, pointing to a shared genetic architecture between faces and limbs (Extended Data Fig. 5a) and a number of genes near our genome-wide-significant loci (for example, Dlx homeobox genes, BMP genes, and *FGFR2*)

have well-established roles in limb development[41]. These findings are also supported by the large number of human syndromes that present with both facial and limb malformations[42].

For the regions surrounding the 203 genome-wide-significant lead SNPs, both CNCCs and embryonic craniofacial tissues showed the highest enrichment in H3K27ac signal (Fig. 2a). These observations are consistent with (1) activity of our 203 genome-wide-significant lead SNPs in CNCCs and embryonic craniofacial tissues and (2) an embryonic origin for human facial variation across the timeline of facial development, as CNCCs represent an early time point in facial development whereas the craniofacial tissues represent progressively later timepoints. In both CNCCs and craniofacial tissue at all sampled developmental stages, regions in the vicinity of the 203 genome-wide-significant lead SNPs were significantly enriched for predicted enhancers and not promoters (Fig. 2b,c). This is an especially intriguing result, as recent evidence has described the action of multiple enhancers, each showing different tissue or timing specificity, in modulating expression levels to affect craniofacial development[43]. Complementing our GREAT analysis results, indicating that some genes near our GWAS peaks are involved in both facial and limb development, a subset of genome-wide-significant lead SNPs showed preferential activity in additional in-vitro-derived cell types relevant to both the face and the rest of the skeletal system, including osteoblasts, chondrocytes, differentiating skeletal muscle myoblasts, fibroblasts and keratinocytes (for example, cluster 3; Fig. 3). Together, these results suggest that genetic variation underlying facial morphology operates by modulating enhancer activity across multiple cell types throughout the timeline of embryonic facial development.

A total of 61 genome-wide-significant peaks from our analysis did not overlap with the results of prior GWAS for normal-range facial morphology, but were located nearby putative craniofacial genes implicated from human malformations or animal models. For instance, *MSX1* has been implicated in orofacial clefting in humans[44,45] and mice[45,46], and is also expressed widely in lip and dental tissues during development[47]. We observed two distinct peaks at the *MSX1* locus (4p16.2), one approximately 55 kb upstream of *MSX1* with a pronounced effect on the lateral upper lip (lead SNP rs13117653; $P_{\text{Meta-US}}(\text{segment } 34) = 4.2\times10^{-18}$) and a second peak, about 323 kb upstream of *MSX1* and located in the intron of *STX18*, involving the lateral lower lip and mandible (lead SNP rs3910659; $P_{\text{Meta-UK}}(\text{segment } 25) = 4.45\times10^{-9}$; Extended Data Fig. 9a–e). This result could indicate a potential role of *STX18* in craniofacial development, although the STX18 protein is important primarily for functioning of the endoplasmic reticulum. Alternatively, this result could provide further evidence that complex phenotypic effects seen in our human sample could be due to the action of multiple regulatory elements within a single locus. In support of this, Attanasio et al. demonstrated that the activity of *Msx1* in the second pharyngeal arch and maxillary process of the e11.5 mouse embryo is recapitulated by the combined activity of two separate enhancers[43].

We also identified 53 genome-wide-significant signals in regions harboring genes with no previously known role in craniofacial development or disease, although many of the implicated genes are known to have a general role in developmental processes critical to morphogenesis. For example, in the current study, variants at the *DACT1* locus are associated with mandibular morphology (Extended Data Fig. 9f–h). *DACT1* is an established antagonist of the Wnt signaling pathway, which is known to be involved in craniofacial development[48], although *DACT1* is studied mostly for its involvement in gastric cancer. However, *DACT1* has also been shown to inhibit the delamination of neural crest cells, further supporting its involvement in facial development[49]. These new signals are promising new candidates for potential roles in facial morphogenesis.

In addition to better understanding which parts of the face had the most signals, we capitalized on the utility of facial segmentation via hierarchical clustering to finely parse out the effect of a SNP even within a complex genomic region. Notably, we observed 24 loci with multiple genome-wide-significant peaks each associated with different facial traits, suggesting that these variants might overlap with or be impacted by regulatory elements that affect the face in highly specific ways (Supplementary Table 5 and Supplementary Data 1). An important consideration to our peak selection procedure is that it is statistical and heuristic in nature, being based on investigator-chosen thresholds of both distance and similarity of associated facial phenotypes, and thus is not perfect. Refining a peak selection approach based on combinations of distance, LD patterns, and trait similarity was beyond the grasp of this paper, but we believe such an approach has potential for further interrogating the complex genetic architecture of facial variation, as we have illustrated using the *TBX15-WARS2* locus (Fig. 4).

Given the complexity of the human face and its component traits, it is likely that the genetic architecture contributing to facial variation includes groups of genomic regions that contribute to the same facial trait, perhaps through actions in similar cell types or explicit interactions among variants. Importantly, genome-wide-significant SNPs that significantly explained variance in the same segment, based on the SEM for that segment, showed higher correlations of cross-sample H3K27ac activity than when compared with SNPs that did not, indicating that the SEM-refined lists of SNPs for each segment are likely those that are similar in either their spatial or temporal cellular activity (Extended Data Fig. 7). Tests for epistasis using the SEM-refined SNP lists for each segment identified four SNP combinations with significant evidence of pairwise epistatic interactions (Table 1). For example, rs76244841 (*PRDM16* associated; $P_{\text{Meta-UK}}(\text{segment } 30) = 1.48\times10^{-8}$) and rs62443772 (*GLI3* associated; $P_{\text{Meta-UK}}(\text{segment } 22) = 5.35\times10^{-16}$) were found to have a significant interaction in facial segment 9, which covers the premaxillary soft tissue from the base of the columella to the oral commissure (Table 1 and Fig. 5). Interestingly, *PRDM16* and *GLI3* are both part of a tetrameric Hedgehog signaling complex in *Drosophila melanogaster* (Supplementary Note 3)[50–52]. Overall, these results indicate that the statistical evidence of SNP groups influencing polygenic facial variation identified through SEM, and explicit variant interactions suggested by the epistasis analysis, are potentially representative of true biological relationships but must be confirmed with further study.

In conclusion, with this work we have not only reported genomic variants influencing normal-range facial variation, but have also sought to use our in-depth facial phenotyping approach and bioinformatic tools to illustrate one way in which researchers without access to functional follow-up analyses can delve deeper into the genetic architecture of complex morphological traits. These results illustrate the potential to highlight spatial and temporal connections between SNPs, representing a major step forward in our ability to characterize the polygenic genetic architecture of complex morphological structures. In performing an open-ended and minimally restrictive study, we are optimistic that our results will be useful for other research efforts to better understand the biological forces that shape human and nonhuman morphology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-00741-7.

Received: 10 October 2019; Accepted: 23 October 2020;
Published online: 7 December 2020

References

1. Atchley, W. R. & Hall, B. K. A model for development and evolution of complex morphological structures. *Biol. Rev.* **66**, 101–157 (1991).
2. Gratten, J., Wray, N. R., Keller, M. C. & Visscher, P. M. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.* **17**, 782–790 (2014).
3. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* **19**, 110–124 (2018).
4. Weinberg, S. M. et al. Hunting for genes that shape human faces: initial successes and challenges for the future. *Orthod. Craniofac. Res.* **22**, 207–212 (2019).
5. Weinberg, S. M., Cornell, R. & Leslie, E. J. Craniofacial genetics: where have we been and where are we going? *PLoS Genet.* **14**, e1007438 (2018).
6. Dixon, M. J., Marazita, M. L., Beaty, T. H. & Murray, J. C. Cleft lip and palate: understanding genetic and environmental influences. *Nat. Rev. Genet.* **12**, 167–178 (2011).
7. Paternoster, L. et al. Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am. J. Hum. Genet.* **90**, 478–485 (2012).
8. Liu, F. et al. A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet.* **8**, e1002932 (2012).
9. Jacobs, L. C. et al. Intrinsic and extrinsic risk factors for sagging eyelids. *JAMA Dermatol.* **150**, 836–843 (2014).
10. Adhikari, K. et al. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nat. Commun.* **7**, 11616 (2016).
11. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
12. Shaffer, J. R. et al. Genome-wide association study reveals multiple loci influencing normal human facial morphology. *PLoS Genet.* **12**, e1006149 (2016).
13. Cole, J. B. et al. Genome-wide association study of African children identifies association of SCHIP1 and PDE8A with facial size and shape. *PLoS Genet.* **12**, e1006174 (2016).
14. Lee, M. K. et al. Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One* **12**, e0176566 (2017).
15. Crouch, D. J. M. et al. Genetics of the human face: identification of large-effect single gene variants. *Proc. Natl Acad. Sci. USA* **115**, E676–E685 (2018).
16. Claes, P. et al. Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat. Genet.* **50**, 414–423 (2018).
17. Endo, C. et al. Genome-wide association study in Japanese females identifies fifteen novel skin-related trait associations. *Sci. Rep.* **8**, 8974 (2018).
18. Cha, S. et al. Identification of five novel genetic loci related to facial morphology by genome-wide association studies. *BMC Genomics* **19**, 481 (2018).
19. Howe, L. J. et al. Investigating the shared genetics of non-syndromic cleft lip/palate and facial morphology. *PLoS Genet.* **14**, e1007501 (2018).
20. Qiao, L. et al. Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. *J. Genet. Genomics* **45**, 419–432 (2018).
21. Wu, W. et al. Whole-exome sequencing identified four loci influencing craniofacial morphology in northern Han Chinese. *Hum. Genet.* **138**, 601–611 (2019).
22. Li, Y. et al. EDAR, LYPLAL1, PRDM16, PAX3, DKK1, TNFSF12, CACNA2D3 and SUPT3H gene variants influence facial morphology in a Eurasian population. *Hum. Genet.* **138**, 681–689 (2019).
23. Xiong, Z. et al. Novel genetic loci affecting facial shape variation in humans. *eLife* **8**, e49898 (2019).
24. White, J. D. et al. MeshMonk: open-source large-scale intensive 3D phenotyping. *Sci. Rep.* **9**, 6085 (2019).
25. Sero, D. et al. Facial recognition from DNA using face-to-DNA classifiers. *Nat. Commun.* **10**, 2557 (2019).
26. Hayton, J. C., Allen, D. G. & Scarpello, V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organ. Res. Methods* **7**, 191–205 (2004).
27. Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T. & Fralish, J. S. Parallel analysis: a method for determining significant principal components. *J. Veg. Sci.* **6**, 99–106 (1995).
28. Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A. & Williams, R. M. Jr. *The American Soldier: Adjustment During Army Life*. Vol. 1 (Princeton Univ. Press, 1949).
29. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma. Oxf. Engl.* **26**, 2190–2191 (2010).
30. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
31. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
32. Som, P. M., Streit, A. & Naidich, T. P. Illustrated review of the embryology and development of the facial region, part 3: an overview of the molecular interactions responsible for facial development. *Am. J. Neuroradiol.* **35**, 223–229 (2014).
33. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
34. White, J. & Indencleef, K. Insights into the genetic architecture of the human face. *FigShare* https://doi.org/10.6084/m9.figshare.c.4667261 (2020).
35. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
36. Watanabe, K., Taskesen, E., Bochoven, Avan & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
37. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
38. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
39. Prescott, S. L. et al. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
40. Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J. P. & Cotney, J. High-resolution epigenomic atlas of human embryonic craniofacial development. *Cell Rep.* **23**, 1581–1597 (2018).
41. Kraus, P. & Lufkin, T. Dlx homeobox gene control of mammalian limb and craniofacial development. *Am. J. Med. Genet. A* **140**, 1366–1374 (2006).
42. Hennekam, R. C. M., Krantz, I. D. & Allanson, J. E. *Gorlin's Syndromes of the Head and Neck* (Oxford Univ. Press, 2010).
43. Attanasio, C. et al. Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**, 1241006 (2013).
44. Beaty, T. H. et al. Testing candidate genes for non-syndromic oral clefts using a case-parent trio design. *Genet. Epidemiol.* **22**, 1–11 (2002).
45. Alappat, S., Zhang, Z. Y. & Chen, Y. P. Msx homeobox gene family and craniofacial development. *Cell Res* **13**, 429–442 (2003).
46. Satokata, I. & Maas, R. Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat. Genet.* **6**, 348–356 (1994).
47. Nakatomi, M. et al. Genetic interactions between Pax9 and Msx1 regulate lip development and several stages of tooth morphogenesis. *Dev. Biol.* **340**, 438–449 (2010).
48. Wang, J.-L. et al. TGF-β signaling regulates DACT1 expression in intestinal epithelial cells. *Biomed. Pharmacother.* **97**, 864–869 (2018).
49. Rabadán, M. A. et al. Delamination of neural crest cells requires transient and reversible Wnt inhibition mediated by Dact1/2. *Development* **143**, 2194–2205 (2016).
50. Stegman, M. A. et al. Identification of a tetrameric hedgehog signaling complex. *J. Biol. Chem.* **275**, 21809–21812 (2000).
51. Méthot, N. & Basler, K. Suppressor of fused opposes hedgehog signal transduction by impeding nuclear accumulation of the activator form of Cubitus interruptus. *Development* **127**, 4001–4010 (2000).
52. Monnier, V., Dussillol, F., Alves, G., Lamour-Isnard, C. & Plessis, A. Suppressor of fused links fused and Cubitus interruptus on the hedgehog signalling pathway. *Curr. Biol. CB* **8**, 583–586 (1998).
53. Krzywinski, M. I. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
54. Brown, G. W. & Mood, A. M. On median tests for linear hypotheses. In *Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probability* (ed. Neyman, J.) 159–166 (Univ. of California Press, 1951).

Methods

Sample and recruitment. The samples used for analysis included a combination of three independently collected datasets from the United States (US; $n_{US}=4{,}680$) and one dataset from the United Kingdom (UK; $n_{UK}=3{,}566$), for a total sample size of $n=8{,}246$. The US samples originated from the 3D Facial Norms cohort[55] (3DFN) and studies at the Pennsylvania State University (PSU) and Indiana University-Purdue University Indianapolis (IUPUI). The UK dataset included samples from the Avon Longitudinal Study of Parents and their Children (ALSPAC)[56,57]. Institutional review board approval was obtained at each recruitment site, and all participants gave their written informed consent before participation. For children, written consent was obtained from a parent or legal guardian. Some individuals from the 3DFN and PSU samples were tested previously for associations with facial morphology in our prior work[16]. A breakdown of the samples used for analysis is shown in Supplementary Table 2 and further details are available in the Supplementary Methods. In all datasets, participants with missing information in sex, age, height or weight, or with insufficient image quality were removed.

Genotyping and imputation. Due to the several genotyping platforms used for the US cohort (details in the Supplementary Methods), we chose to impute the samples from each platform separately, then combine the imputed results[58]. For each dataset, standard data cleaning and quality assurance practices were performed based on the GRCh37 genome assembly. Phasing was performed using SHAPEIT2 (v.2.r900)[59] and imputation to the 1000G Phase 3 reference panel[60] performed using the positional Burrows-Wheeler Transform[61] pipeline (v.3.1) of the Sanger Imputation Server (v.0.0.6)[62]. After post-imputation quality control and intersection of imputed SNPs, a single merged dataset of all US participants was created with 7,417,619 SNPs for analysis.

The raw genotype data from ALSPAC were not available, and restrictions are in place against merging the ALSPAC genotypes with any others. For this reason, ALSPAC genotypes, phased using SHAPEIT2[59] and imputed to the 1000G Phase 1 reference panel (Version 3)[63] using IMPUTE2[64], were obtained directly from the ALSPAC database and held separately during the analysis. After post-imputation quality control, the ALSPAC dataset contained 8,629,873 SNPs for analysis.

For both datasets, SNPs on the X chromosome were coded 0/2 for hemizygous males, to match with the 0/1/2 coding for females[12].

Ancestry axes and selection of European participants. From the post-imputation merged dataset of US participants, we identified the European participants by projecting them into a PC space constructed using the 1000G Phase 3 dataset, first filtered for LD and SNPs shared between both datasets. Further details are available in the Supplementary Methods. In the combined PC space, we calculated the ancestry axes for the US participants and the Euclidean distance between all US participants and the 1000G samples. Using a kth nearest neighbor algorithm, we identified the five nearest 1000G neighbors for each US participant. The most common 1000G population label from these five nearest neighbors was then assigned to the US participant, and participants assigned the 1000G European population labels of CEU, TSI, FIN, GBR and IBS were selected for analysis.

Ancestry axes were calculated for the UK participants by projecting them into the 1000G Phase 3 dataset in a manner similar to that described for the US participants. Since all ALSPAC participants available for this analysis were European, no additional ancestry refinement was performed.

3D image acquisition. For all datasets, 3D images were captured using either a digital facial stereophotogrammetry system or a laser scanning system. All participants were asked to have closed mouths and to maintain a neutral facial expression during image capture[65]. For the 3DFN sample, facial surfaces were acquired using the 3dMDface (3dMD) camera system. PSU images were obtained with either the 3dMDface or Vectra H1 system (Canfield Scientific). The IUPUI sample was fully imaged using Vectra H1. The ALSPAC sample was imaged using a Konica Minolta Vivid 900 laser scanner (Konica Minolta Sensing Europe). For this system, two high-resolution facial scans were taken and then processed, merged and registered using a macro algorithm in Rapidform 2004 software (INUS Technology Inc.).

3D image registration and quality control. The 3D surface images and their reflections were registered using the MeshMonk registration framework (v.0.0.6)[24] in Matlab 2017b. This process results in a homologous configuration of 7,160 spatially dense quasi-landmarks, allowing the image data from different individuals and camera systems to be standardized[24]. Images differing greatly from the norm or with large holes were investigated manually and either removed or re-processed, with details available in the Supplementary Methods. Although variation in asymmetric facial features is of interest, in this work we sought only to study variation in symmetric facial shape.

Segmentation of facial shape. To study global and local effects on facial variation, we performed a data-driven facial segmentation on the UK and US datasets combined, as described previously[16]. Before segmentation, images in the two datasets were separately adjusted for sex, age, age-squared, height, weight, facial size, the first four genomic ancestry axes and the camera system, using PLSR (function plsregress from Matlab 2017b). As an illustration, the age adjustment is visualized in Supplementary Fig. 2. After adjustment, facial segments were defined by grouping vertices that are correlated strongly using hierarchical spectral clustering[16,25]. The strength of covariation between quasi-landmarks was defined using Escoufier's RV coefficient[66,67]. The RV coefficient was then used to build a structural similarity matrix that defined the hierarchical construction of 63 facial segments, broken into five levels (Extended Data Fig. 1a). The configurations of each segment were then subjected independently to a Generalized Procrustes analysis[68], after which a PCA was performed in combination with parallel analysis to capture the major variance in the facial segments with fewer variables[26,27] (Extended Data Fig. 1b).

Multivariate genome-wide-association meta-analyses. The meta-analysis framework utilized consists of three steps performed separately for each of the 63 segments: identification, verification, and meta-analysis (Extended Data Fig. 2). For all analyses, the genotypes were coded additively based on the presence of the major allele. In the identification step, for each of the 63 facial segments, each SNP was associated with phenotypic variation using CCA (canoncorr in Matlab 2017b). CCA is a multivariate analysis that extracts the linear combination of PCs, which represent the direction of phenotypic effect in shape space (which we call a 'trait') that are maximally correlated with a SNP, and returns a correlation value between those PCs and the SNP tested. Because CCA does not accommodate adjustments for covariates, we removed the effect of relevant covariates (sex, age, age-squared, height, weight, facial size, the first four genomic ancestry axes and the camera system), on both the independent (SNP) and the dependent (facial shape) variables using PLSR (plsregress from Matlab 2017b), and thus performed the CCA under a reduced model with residualized variables. The correlation value between PCs and SNPs is tested for significance based on Rao's F-test approximation[69] (right tail, one-sided test). In summary, for each of the 63 segments, the CCA component of the identification step identifies the phenotypic trait most correlated with each SNP ($Trait_{US}$ and $Trait_{UK}$ in Extended Data Fig. 2) and Rao's F-test provides a P value ($P_{CCA\text{-}US}$ and $P_{CCA\text{-}UK}$) representing the strength of the correlation. CCA has also been implemented in 'mv-PLINK'[70]. Performance tests of mv-PLINK have shown that it outperforms univariate methods and has similar power to other multivariate methods of association[70–72], which generally have higher statistical power than univariate methods[70–76].

In the verification step, the shape PCs of the nonidentification dataset were projected onto the trait found in the identification stage, which returns a univariate variable (which we call a 'phenotype'; $UniVar_{US}$ and $UniVar_{UK}$). These univariate variables were then tested for genotype–phenotype associations in a standard linear regression (regstats in Matlab 2017b) with the SNP genotypes of the verification dataset as independent variable and the univariate trait projection score as the dependent variable. This function employs a t-statistic and a one-sided (right tail) P value was obtained with the Student's t cumulative distribution function[77] (function tcdf in Matlab 2017b).

In the meta-analysis step, the identification P value (from Rao's F-test on the canonical correlation) and the verification P value (from the univariate regression) were combined using Stouffer's method[28,29], which was chosen because a meta-analysis of beta values was not possible given that the CCA returns a positive correlation value, not a beta statistic. The entire process was repeated, resulting in two meta-analysis P values ($P_{Meta\text{-}US}$ and $P_{Meta\text{-}UK}$) accompanied by two identified traits per segment and per SNP: first using US data in the identification stage and UK data as verification ($META_{US}$ or US-driven), then using UK data in the identification stage and US data as verification ($META_{UK}$ or UK-driven). A validation of our analysis pipeline is available in Supplementary Note 1.

Sharing of genome-wide signal between facial segments. To assess the extent to which genome-wide signals of association with facial variation were shared between a pair of facial segments, LD score regression[30,31] was applied to the meta-analysis, after converting the meta P values to z-scores and ignoring the sign or direction of effect. The former was required because of the multivariate nature of our results and the latter was needed since CCA is a one-sided test with canonical correlations always between [0 1]. As a result, all resulting genetic correlations reported here are restricted to be positive as well. Further details on the calculation of LDSC values are available in the Supplementary Methods. This process was done twice, once each for the US- and UK-driven meta-analyses. A high degree of congruence ($r_s = 0.95$) between the results based on the US- and UK-driven meta-analyses was observed, and the average correlation of both between each pair of facial segments was reported. The 63×63 matrix of average correlations was visualized on top of the facial segmentation hierarchy to assess correlation both within and between facial quadrants (Extended Data Fig. 3) and used to perform average-linkage hierarchical clustering (Extended Data Fig. 4).

GWAS peak selection. The analysis strategy yielded 126 meta-analysis P values and 126 traits for every SNP, representing the 63 segments × two meta-analysis tracks. Per SNP, the lowest P value was selected, and we noted in which meta-analysis track ($META_{US}$ or $META_{UK}$; 'Best meta-analysis track') and segment ('Best segment') this P value occurred. The study-wide Bonferroni threshold

($P \leq 6.96 \times 10^{-10}$) was calculated as $5 \times 10^{-8}/(1.0042 \times 1.6631 \times 43.0145)$, with the denominator values representing the number of independent tests per SNP, across both meta-analysis tracks, and across all segments, respectively. These values were calculated using 10,000 permutations each of 1,000 random SNPs, with more details available in Supplementary Note 2 and the permutation outcomes available in the FigShare repository for this article[34]. Although a study-wide threshold was calculated, we chose to annotate lead SNPs reaching at least genome-wide threshold to retain as many potentially biologically meaningful results as possible. The FigShare repository also provides information on all SNPs reaching suggestive significance ($P = 5 \times 10^{-7}$) as well as QQ plots for each segment in all stages of the analysis[34]. For the initial peak selection, we chose to group SNPs below genome-wide threshold by genomic position, and the SNP with the lowest *P* value per genomic region was selected as the lead SNP. Within a ±500-kb window of the resulting genome-wide-significant lead SNPs, we further refined the selection by performing a regression of slopes on the traits defined in the identification stage (in Best meta-analysis track and Best segment) to determine if adjacent SNPs showed consistent effects with the lead SNP, resulting in 218 genome-wide-significant lead SNPs. Of these 218 lead SNPs, 203 showed consistent traits in the US and UK datasets in the Best segment (Supplementary Table 3), with more details in the Supplementary Methods. Visual representations of the LocusZoom[33] and effect plots for each of the 203 genome-wide-significant SNPs are available in the FigShare repository[34]. The 203 lead SNPs were mapped to 138 cytogenetic bands (loci) using the Ensembl GRCh37 locations[78]. This method of peak selection is statistical in nature and is thus not perfect. For example, our inspection of the LocusZoom[33] plots for the *TBX15-WARS2* locus led to the identification of two clusters of SNPs, based on r^2 correlation, sharing the same genomic positions and affecting different facial segments, but separating these two clusters was not possible in our initial peak selection and they were considered a single signal until manual investigation. To comprehensively identify SNPs within a locus contributing to facial morphology, and the specific facial segments affected, fine mapping and other detailed investigations are needed.

Gene annotation. Genes ±500 kb of the genome-wide-significant lead SNPs were identified using the Table Browser of the UCSC Genome Browser[79]. The most likely candidate gene per lead SNP was identified based on a three-step system using first literature searches, then the results from Hooper et al. on the transcriptomics of mouse facial development[80], then the FUMA gene prioritization algorithm (v.1.3.3)[36]. Further details are available in the Supplementary Methods. Using the available literature, we classified the lead SNP into one of five categories: 'Region previously implicated in normal-range facial morphology', 'Region previously implicated in normal-range facial morphology using other analyses of these data', 'Candidate gene implicated in craniofacial morphology through animal model', 'Region or candidate gene implicated in craniofacial morphology through human dysmorphology' and 'No previous association'. To the best of our knowledge, all links with facial morphology from the literature are provided in Supplementary Table 3.

To investigate the potential roles of the identified genome-wide-significant lead SNPs, analyses using FUMA (v.1.3.3)[36], which can test for enrichment of a set of genes in predefined pathways, and GREAT (v.3.0.0)[35], which predicts the function of cis-regulatory regions, were performed using preset parameters (Extended Data Fig. 5). In this article, we focus on the top FUMA and GREAT results, based on *P* value, and have provided the full export of GREAT results in the FigShare repository[34].

Cell-type-specific enhancer enrichment. To assess activity of the 203 genome-wide-significant lead SNPs in various cell types and tissues (further details in the Supplementary Methods), we analyzed signals of acetylation of histone H3 on lysine 27 (H3K27ac). Across cell types and tissues, we compared 20-kb windows containing the 203 genome-wide-significant lead SNPs, 203 random SNPs matched for minor allele frequency and distance to the distance to the nearest gene by using SNPsnap[81], or 619 Crohn's disease-associated SNPs from the National Center for Biotechnology Information-European Bioinformatics Institute (NCBI-EBI) GWAS catalog[82]. Regions in the vicinity of SNPs associated with Crohn's disease showed the highest H3K27ac signal in various immune cell types, serving as a positive control for both our approach and dataset (Extended Data Fig. 10). A two-sided Wilcoxon rank-sum test was used to compare the H3K27ac signal between the 203 genome-wide-significant lead and random SNPs, within each cell type and tissue analyzed. *K*-means clustering was performed on the lead SNP H3K27ac signal across all cell types and tissues with $k = 6$, as we found that this value maximized the number of clusters without significantly impacting cluster quality, as measured by silhouette width (Fig. 3).

Chromatin state association in CNCCs and embryonic craniofacial tissue. Lists of human CNCC regulatory elements were annotated based on multiple chromatin marks by Prescott et al.[39] and embryonic craniofacial chromHMM states were computed in combined data from each Carnegie stage by Wilderman et al.[40]. For each set of regulatory regions, all regions within 20 kb of either genome-wide-significant lead SNPs or the above-described 203 random SNPs were considered. Enrichment/depletion of each class of regulatory region for lead SNPs versus random SNPs was computed using a two-sided Fisher's exact test (Fig. 2b,c).

Structural equation modeling. To better define the cause–effect relationships between the significant genotypes and their collective traits, both the US and UK participants were used as input for SEM using the lavaan package (v.0.6-3) in R (≥3.5.0)[83], which reports a two-sided *P* value. For our analyses, separate SEM models were constructed for each segment using each of the 203 genome-wide-significant lead SNPs and the shape PCs for all participants, with additional information available in the Supplementary Methods.

For each of the 50 SEM models where the refinement process was successful (details in the Supplementary Methods), final model fit indices and model parameter estimates are provided in Supplementary Data 2. Reassuringly, for segments that are closely related in the segmentation hierarchy (segments 5, 11, 23 and 47) there is an average overlap of 46% of the variants meeting the $P < 0.05$ cutoff for SEM significance, compared to 13.6% average overlap for nonhierarchically related segments (segments 5 and 6). The H3K27ac activity across all cell types was compared for significant variants both within and between segments using Spearman's rho using two-sided Kruskal–Wallis tests (Extended Data Fig. 7).

Epistasis analysis. We additionally used the univariate latent variable and the variants passing the $P < 0.05$ significance cutoff from the final 50 refined SEM models ($P < 0.1$ for segments 7, 16 and 25) to assess whether interactions between genotypes increase or decrease the distribution of the latent variable. For each segment, the effect on the latent variable of all diplotype combinations of variants were assessed via a linear regression epistasis analysis in Plink v.1.9 (ref. [84]). After Bonferroni correction for multiple testing, four SNP pairs were significant at $P < 0.05$ (Table 1). For these four pairs, the nine diplotype combinations and their normalized phenotypic and marginal distributions were plotted (Fig. 5 and Extended Data Fig. 8) to assess the genotypic contribution to epistatic masking (the combination of two variants reduce the phenotype) and boosting (the combination of two variants increase the phenotype). For each diplotype combination, the marginal phenotypic medians of the singular genotypes were averaged to visualize the predicted phenotypic distribution that would occur if the two genotypes were acting independently, and this average median was compared to the medians of the combined diplotypes. Significance testing was performed using a two-sided Mood's Median test[54] with one degree of freedom. These steps were performed using the R packages agricolae (v.1.3-0), cowplot (v.1.0.0), ggplot2 (v.3.1.1), ggpubr (v.0.2), gridExtra (v.2.3), gtable (v.0.3.0), grid (v.3.6.2), Hmisc (v.4.2-0), psych (v.1.8.12) and data.table (v.1.12.0).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All of the genotypic markers for the 3DFN dataset are available to the research community through the dbGaP controlled-access repository (http://www.ncbi.nlm.nih.gov/gap) at accession no. phs000949.v1.p1. The raw source data for the phenotypes—the 3D facial surface models in.obj format—are available through the FaceBase Consortium (https://www.facebase.org) at accession no. FB00000491.01. Access to these 3D facial surface models requires proper institutional ethics approval and approval from the FaceBase data access committee. Additional details can be requested from S.M.W.

The participants making up the PSU and IUPUI datasets were not collected with broad data sharing consent. Given the highly identifiable nature of both facial and genomic information and unresolved issues regarding risk to participants, we opted for a more conservative approach to participant recruitment. Broad data sharing of the raw data from these collections would thus be in legal and ethical violation of the informed consent obtained from the participants. This restriction is not because of any personal or commercial interests. Additional details can be requested from M.D.S. and S.W. for the PSU and IUPUI datasets, respectively. The ALSPAC (UK) data will be made available to bona fide researchers on application to the ALSPAC Executive Committee (http://www.bris.ac.uk/alspac/researchers/data-access). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

Publicly available data used were the 1000G Phase 3 data (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), the list of HapMap 3 SNPs excluding the MHC region (http://ldsc.broadinstitute.org/static/media/w_hm3.noMHC.snplist.zip), and ChIP–seq files from Prescott et al.[39] (GSE70751), Najafova et al.[85] (GSE82295), Baumgart et al.[86] (GSE89179), Nott et al.[87] (https://genome.ucsc.edu/s/nottalexi/glassLab_BrainCellTypes_hg19), Pattison et al.[88] (GSE119997), Wilderman et al.[40] (GSE97752) and the Roadmap Epigenomics Project[89] (https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/). Meta-analysis GWAS statistics are available on GWAS Catalog (GCP000044). All data relevant to run future replications and meta-analysis efforts are provided in the FigShare repository for this work[34], along with additional figures (https://doi.org/10.6084/m9.figshare.c.4667261). Items available in the FigShare repository are (1) anthropometric mask: a Matfile of the anthropometric mask used; (2) association statistics and effects of the 203 lead SNPs: facial effects, LocusZoom plots and association statistics from each stage of the analysis for the 203 lead SNPs; (3) calculation of study-wide-significance

threshold: script and permutation outcomes needed to replicate the calculation of the study-wide-significance threshold; (4) facial segment assignments: segment assignments for each quasi-landmark in the anthropometric mask; (5) Fig. 2a labeled: a larger version of Fig. 2a, with all cell types and tissues labeled; (6) GREAT Export: raw output of the GREAT analysis; (7) PCA shape constructs: PCA shape spaces for all 63 facial segments; (8) QQ plots: QQ plots for each segment in all stages of the analysis; (9) script to explore facial segments and GWAS hits: MatLab script for select data exploration functions; (10) SNPs reaching suggestive significance in either meta-analysis track: association statistics of all SNPs with $P < 5 \times 10^{-7}$ in METAUS or METAUK tracks; (11) source data for manuscript figures: source data in Excel format for all figures, where possible.

Code availability

KU Leuven provides the MeshMonk (v.0.0.6) spatially dense facial-mapping software, free to use for academic purposes (https://github.com/TheWebMonks/meshmonk). Matlab 2017b implementations of the hierarchical spectral clustering to obtain facial segmentations are available from a previous publication[25] (https://doi.org/10.6084/m9.figshare.7649024).

The statistical analyses in this work were based on functions of the statistical toolbox in Matlab 2017b, SHAPEIT2 (v.2.r900), Sanger Imputation Server (v.0.0.6), PBWT pipeline (v.3.1), MeshMonk (v.0.0.6), LDSC (v.1.0.1), FUMA (v.1.3.3), GREAT (v.3.0.0), Plink v.1.9, lavaan (v.0.6-3), R (>v.3.4), agricolae (v.1.3-0), cowplot (v.1.0.0), ggplot2 (v.3.1.1), ggpubr (v.0.2), gridExtra (v.2.3), gtable (v.0.3.0), grid (v.3.6.2), Hmisc (v.4.2-0), psych (v.1.8.12), data.table (v.1.12.0), Genotype Harmonizer (v.1.4.20), KING (v.2.1.3), bowtie2 (v.2.3.4.2), bedtools (v.2.27.1) and Bioconductor (v.3.7), as mentioned throughout the Methods.

References

55. Weinberg, S. M. et al. 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 (2016).
56. Boyd, A. et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon longitudinal study of parents and children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
57. Fraser, A. et al. Cohort profile: the Avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
58. Verma, S. S. et al. Imputation and quality control steps for combining multiple genome-wide datasets. *Front. Genet.* **5**, 370 (2014).
59. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
60. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
61. Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinforma. Oxf. Engl.* **30**, 1266–1272 (2014).
62. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
63. 1000 Genomes Project Consortium. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
64. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 Genes Genomics Genet.* **1**, 457–470 (2011).
65. Heike, C. L., Upson, K., Stuhaug, E. & Weinberg, S. M. 3D digital stereophotogrammetry: a practical guide to facial image acquisition. *Head. Face Med.* **6**, 18 (2010).
66. Robert, P. & Escoufier, Y. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **25**, 257–265 (1976).
67. Klingenberg, C. P. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a priori hypotheses. *Evol. Dev.* **11**, 405–421 (2009).
68. Rohlf, F. J. & Slice, D. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Biol.* **39**, 40–59 (1990).
69. Olson, C. L. On choosing a test statistic in multivariate analysis of variance. *Psychol. Bull.* **83**, 579–586 (1976).
70. Ferreira, M. A. R. & Purcell, S. M. A multivariate test of association. *Bioinformatics* **25**, 132–133 (2009).
71. Galesloot, T. E., van Steen, K., Kiemeney, L. A. L. M., Janss, L. L. & Vermeulen, S. H. A comparison of multivariate genome-wide association methods. *PLoS One* **9**, e95923 (2014).
72. Porter, H. F. & O'Reilly, P. F. Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci. Rep.* **7**, 38837 (2017).
73. O'Reilly, P. F. et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One* **7**, e34861 (2012).
74. Korte, A. et al. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* **44**, 1066–1071 (2012).
75. Stephens, M. A unified framework for association analysis with multiple related phenotypes. *PLoS One* **8**, e65245 (2013).
76. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
77. Devroye, L. *Non-uniform Random Variate Generation* (Springer, 1986).
78. Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
79. Karolchik, D. et al. The UCSC table browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
80. Hooper, J. E. et al. Systems biology of facial development: contributions of ectoderm and mesenchyme. *Dev. Biol.* **426**, 97–114 (2017).
81. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31**, 418–420 (2015).
82. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
83. Rosseel, Y. lavaan: an R package for structural equation modeling. *J. Stat. Softw.* **48**, 1–36 (2012).
84. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
85. Najafova, Z. et al. BRD4 localization to lineage-specific enhancers is associated with a distinct transcription factor repertoire. *Nucleic Acids Res.* **45**, 127–141 (2017).
86. Baumgart, S. J. et al. CHD1 regulates cell fate determination by activation of differentiation-induced genes. *Nucleic Acids Res.* **45**, 7722–7735 (2017).
87. Nott, A. et al. Brain cell type-specific enhancer-promoter interactome maps and disease risk association. *Science* **366**, 1134–1139 (2019).
88. Pattison, J. M. et al. Retinoic acid and BMP4 cooperate with TP63 to alter chromatin dynamics during surface epithelial commitment. *Nat. Genet.* **50**, 1658–1665 (2018).
89. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

Acknowledgements

We are extremely grateful to all the individuals and families who took part in this study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We are also very grateful to all of the US participants for generously donating their time to our research, and to present and former laboratory members who worked tirelessly to make these analyses possible. Pittsburgh personnel, data collection and analyses were supported by the National Institute of Dental and Craniofacial Research (U01-DE020078, program director/principal investigators (PD/PIs): M.L.M./S.M.W.; R01-DE016148, PD/PIs: M.L.M./S.M.W.; and R01-DE027023, PD/PIs: S.M.W./J.R.S.). Funding for genotyping by the National Human Genome Research Institute (X01-HG007821 and X01-HG007485, PD/PI: M.L.M.) and funding for initial genomic data cleaning by the University of Washington provided by contract HHSN268201200008I from the National Institute for Dental and Craniofacial Research awarded to the Center for Inherited Disease Research (https://www.cidr.jhmi.edu/). Penn State personnel, data collection and analyses were supported by Procter & Gamble, Company (UCRI-2015-1117-HN-532, PD/PIs: H.L.N.), the Center for Human Evolution and Development at Penn State, the Science Foundation of Ireland Walton Fellowship (04.W4/B643, PD/PI: M.D.S.), the US National Institute of Justice (2008-DN-BX-K125, PD/PI: M.D.S.; and 2018-DU-BX-0219, PD/PIs: S.W.) and by the US Department of Defense. IUPUI personnel, data collection and analyses were supported by the National Institute of Justice (2015-R2-CX-0023, 2014-DN-BX-K031 and 2018-DU-BX-0219, PD/PI: S.W.). University of Cincinnati personnel and data collection were supported by Procter & Gamble, Company (UCRI-2015-1117-HN-532, PD/PI: H.L.N.). The UK Medical Research Council and Wellcome (grant no. 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. The publication is the work of the authors and K.I. and P.C. will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). ALSPAC GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. The KU Leuven research team and analyses were supported by the National Institute of Dental and Craniofacial Research (R01-DE027023, PD/PIs: S.M.W./J.R.S.), The Research Fund KU Leuven (BOF-C1, C14/15/081 and C14/20/081, PD/PI: P.C.), The Research Program of the Research Foundation—Flanders (FWO, G078518N, PD/PI: P.C.) and a Senior Clinical Investigator Fellowship of The Research Foundation—Flanders (G078714N, PD/PI: G.H.). Stanford University personnel and analyses were supported by the National Institute of Dental and Craniofacial Research (R01-DE027023, PD/PIs: S.M.W./J.R.S.; and U01-DE024430, PD/PIs: J.W./L. Selleri), the Howard Hughes Medical Institute and the March of Dimes Foundation (1-FY15-312, PD/PI: J.W.).

Author contributions

P.C., M.D.S., S.M.W., J.R.S., J.W. and S.W. conceptualized the study (ideas; formulation or evolution of overarching research goals and aims). J.D.W., K.I., R.J.E., M.K.L., J.L., S.W. and P.C. carried out the data curation (management activities to annotate (produce metadata), scrub data and maintain research data for initial use and later re-use). J.D.W., K.I., S.N., R.J.E., H.H., J.R., J.L. and P.C. carried out the formal analysis (application of statistical, mathematical, computational or other formal techniques to analyze or

synthesize study data). S.R., H.L.N., E.F., T.S., M.L.M., J.R.S., J.W., S.W., S.M.W., M.D.S. and P.C. were responsible for funding acquisition (acquisition of the financial support for the project leading to this publication). J.D.W., K.I., S.N., R.J.E., H.H., J.R., M.K.L., J.L. and P.C. carried out the investigation (conducting a research and investigation process, specifically performing the experiments or data/evidence collection). J.D.W., S.N., R.J.E., J.M., S.R., E.E.Q., H.L.N., T.S., M.L.M., J.W., S.W., S.M.W. and M.D.S. provided the resources (provision of study materials, computing resources or other analysis tools). P.C., S.M.W., M.D.S., S.W., J.W., J.R.S., M.L.M., T.S., H.P. and G.H. carried out the supervision (oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team). J.D.W., K.I., S.N., R.J.E., H.H., J.R., M.K.L. and P.C. did the visualization (preparation, creation and/or presentation of the published work, specifically visualization/data presentation). J.D.W., K.I., S.N., R.J.E. and J.R. wrote the original draft. J.D.W., K.I., S.N., R.J.E., H.H., J.R., S.R., E.E.Q., M.L.M., H.P., J.R.S., J.W., S.W., S.M.W., M.D.S. and P.C. reviewed and edited the final manuscript.

Competing interests

H.L.N. has received $6,000 in consulting fees from Procter & Gamble, Company. Procter & Gamble, Company had no role in the conceptualization, design, data analysis, decision to publish or preparation of this manuscript. All other authors declare no competing interests.

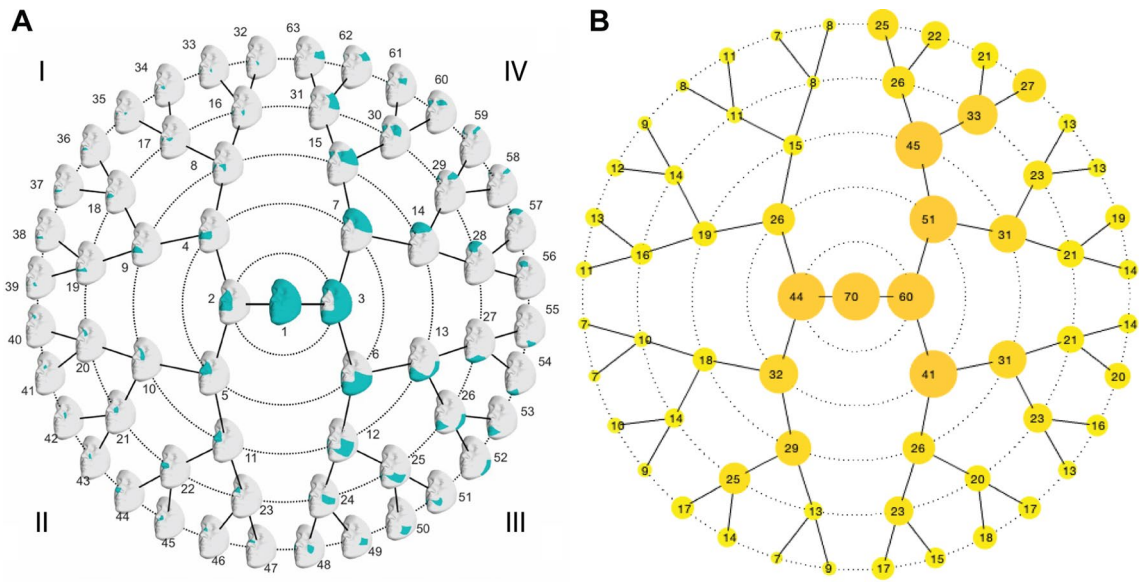
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-00741-7.

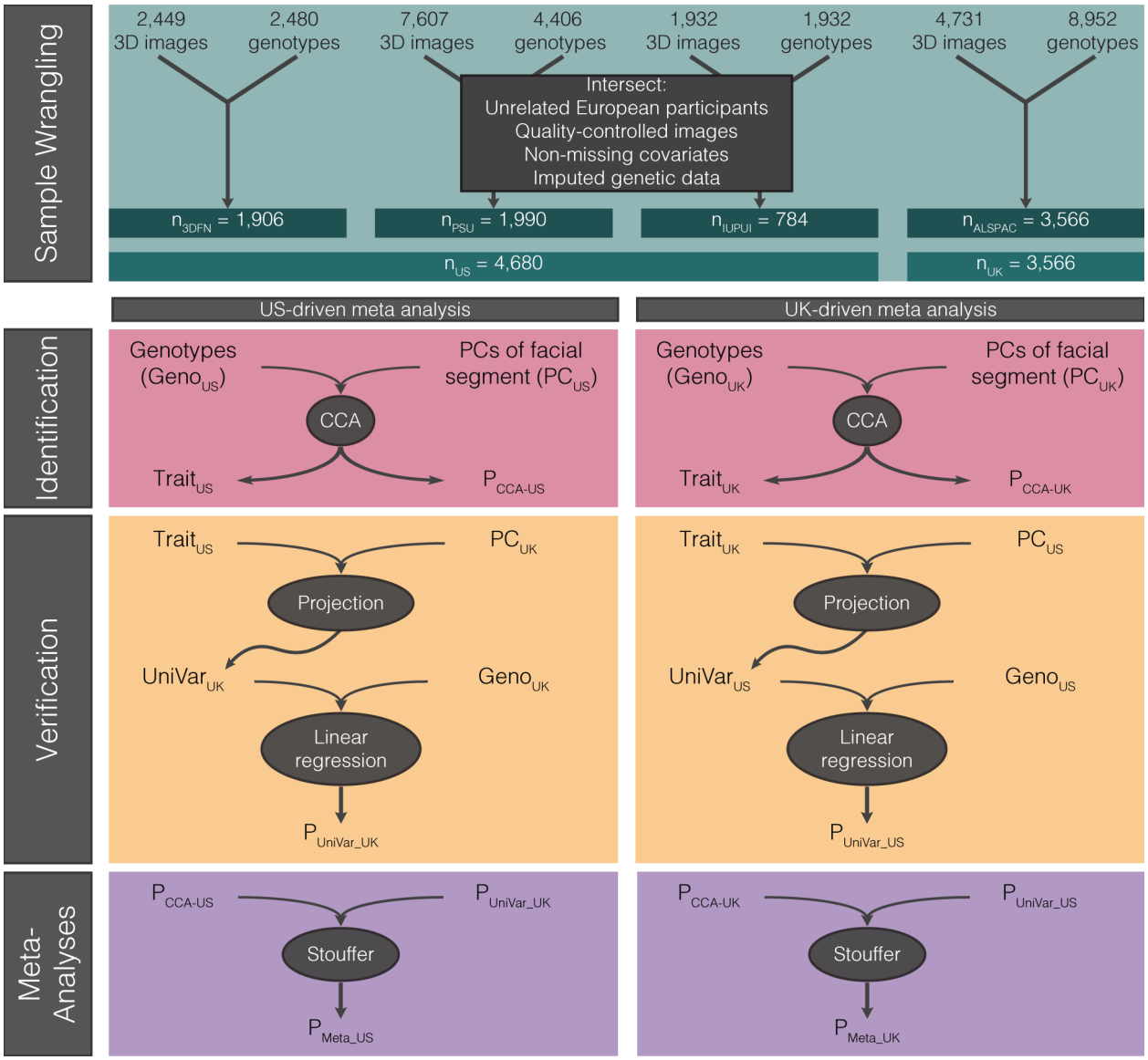
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-00741-7.

Correspondence and requests for materials should be addressed to J.D.W., K.I. or P.C.

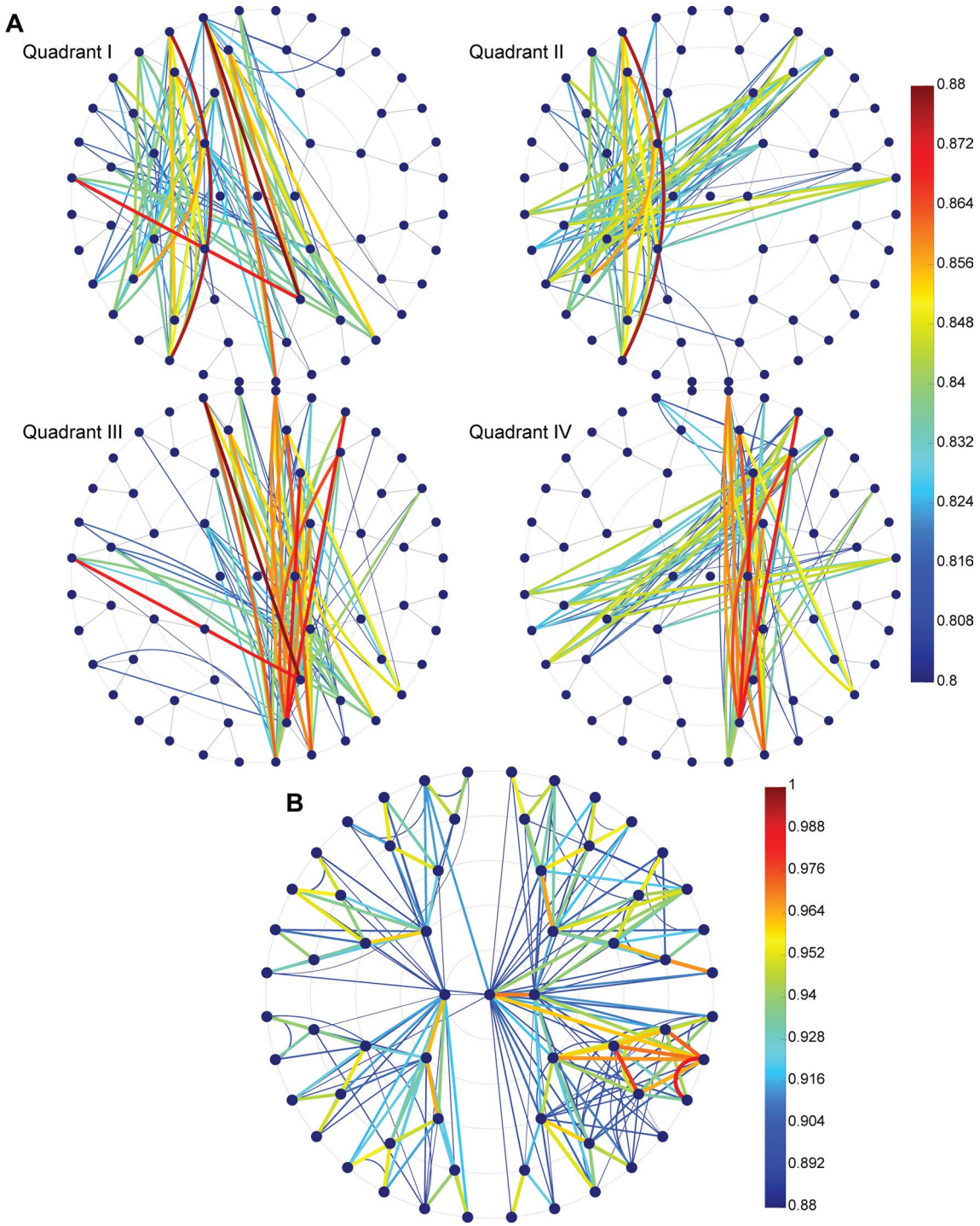
Reprints and permissions information is available at www.nature.com/reprints.



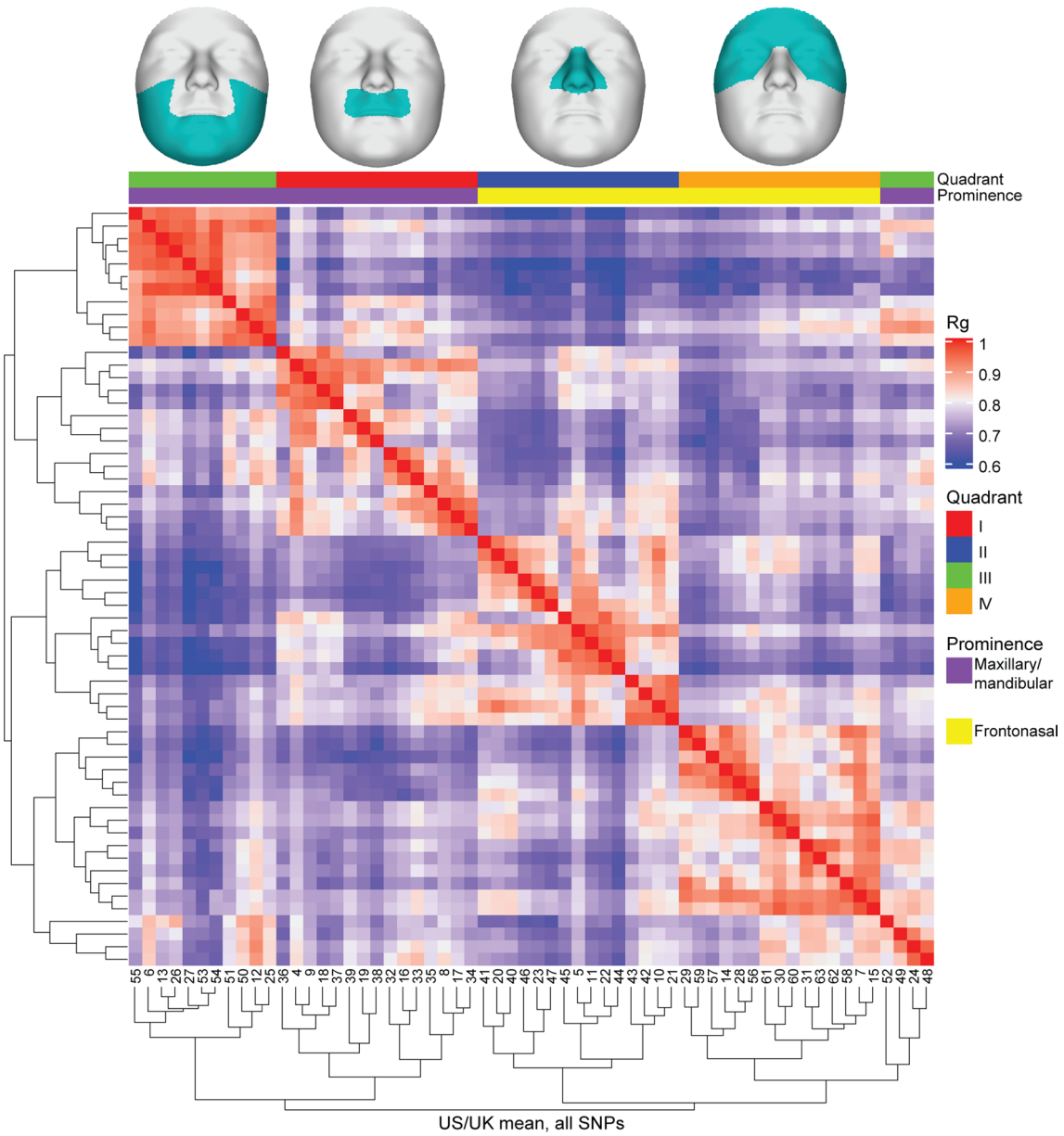
Extended Data Fig. 1 | Hierarchical spectral clustering of facial shape. a, Global-to-local facial segmentation of all 3D images (n_{Total} = 8,246), obtained using hierarchical spectral clustering. Segments are colored in teal and identical to those in Fig. 1. Roman numerals represent 'quadrants' of facial segments. **b**, The number of principal components retained after parallel analysis for each facial segment.



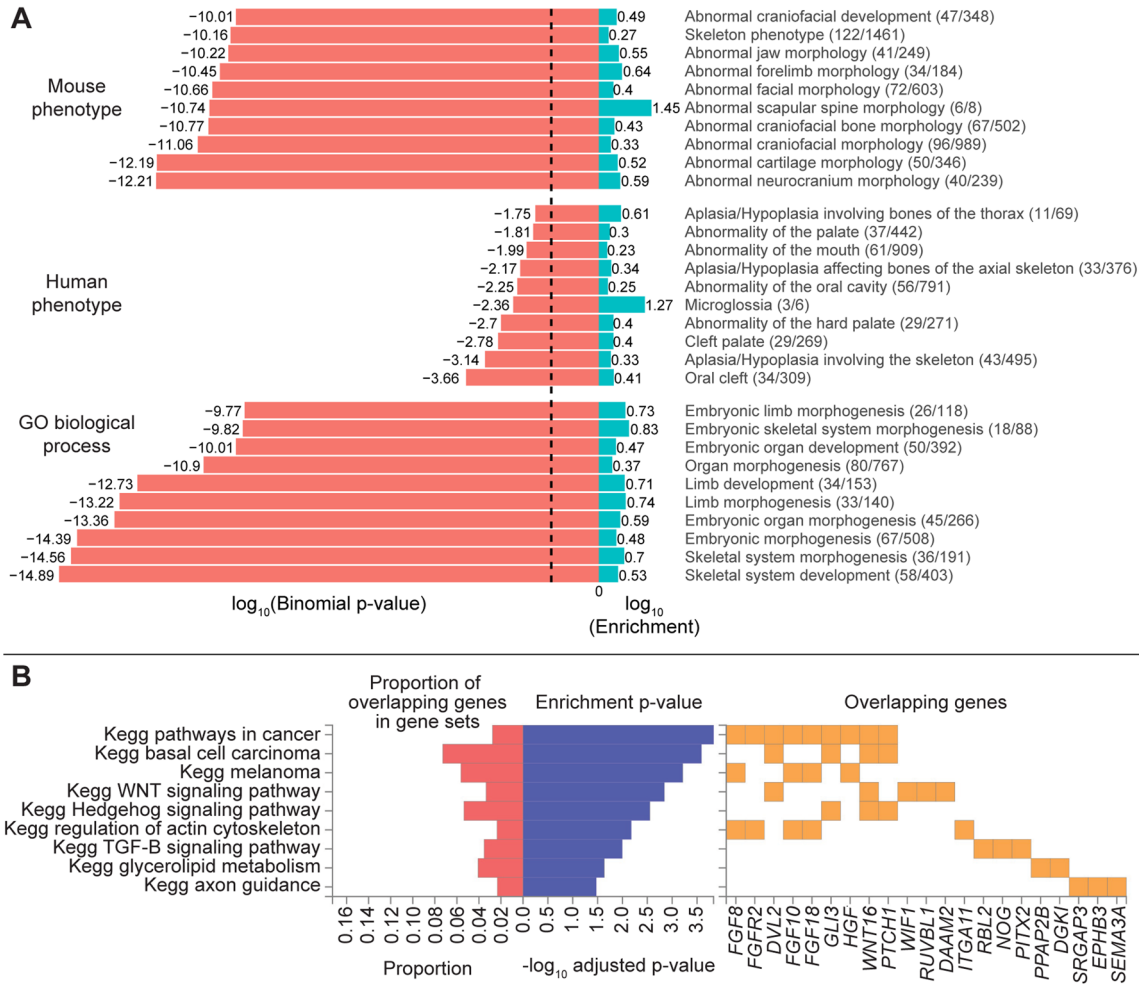
Extended Data Fig. 2 | Study design. *Sample Wrangling*: Images and genotypes from each study were intersected and unrelated participants of European ancestry, with quality-controlled images, covariates, and imputed genetic data were selected to obtain the analyzed data. *Identification*: For each facial segment, canonical correlation analysis (CCA) and Rao's F-test approximation was used to identify the multivariate combination of facial principal components most correlated with the genotypes, which led to a *P* value ($P_{CCA\text{-}US}$ or $P_{CCA\text{-}UK}$) and multivariate phenotypic trait most correlated with each SNP ($Trait_{US}$ and $Trait_{UK}$). *Verification*: The principal components of the other dataset were then projected onto this trait to obtain a univariate variable representing the distribution of participants from the verification dataset for the trait identified in the identification dataset ($UniVar_{UK}$ and $UniVar_{US}$). The genotypes of the verification dataset are then tested against this variable via linear regression, resulting in an additional *P* value ($P_{UniVar\text{-}UK}$ and $P_{UniVar\text{-}US}$). *Meta-Analysis*: The *P* values from identification and verification are meta-analyzed using Stouffer's method, resulting in the final set of *P* values from each meta-analysis track ($P_{META\text{-}US}$ and $P_{META\text{-}UK}$).



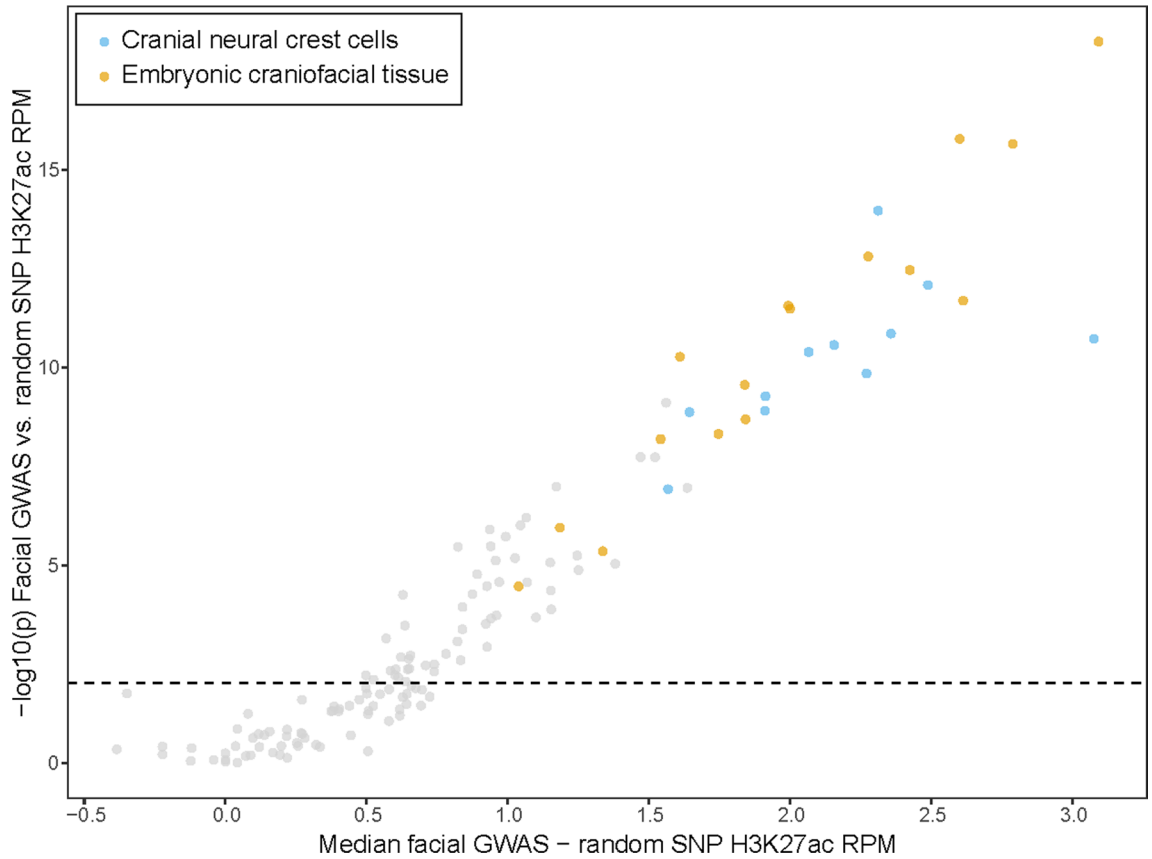
Extended Data Fig. 3 | Genomic signal correlations. LDSC correlations between segments. **a**, Correlations between segments from different quadrants, ranging from 0.8 to 0.88, which seem to reflect both physical proximity of segments on the face and shared embryological origins. **b**, Correlations ranging from 0.88 to 1, which are mostly between segments within the same facial quadrant.



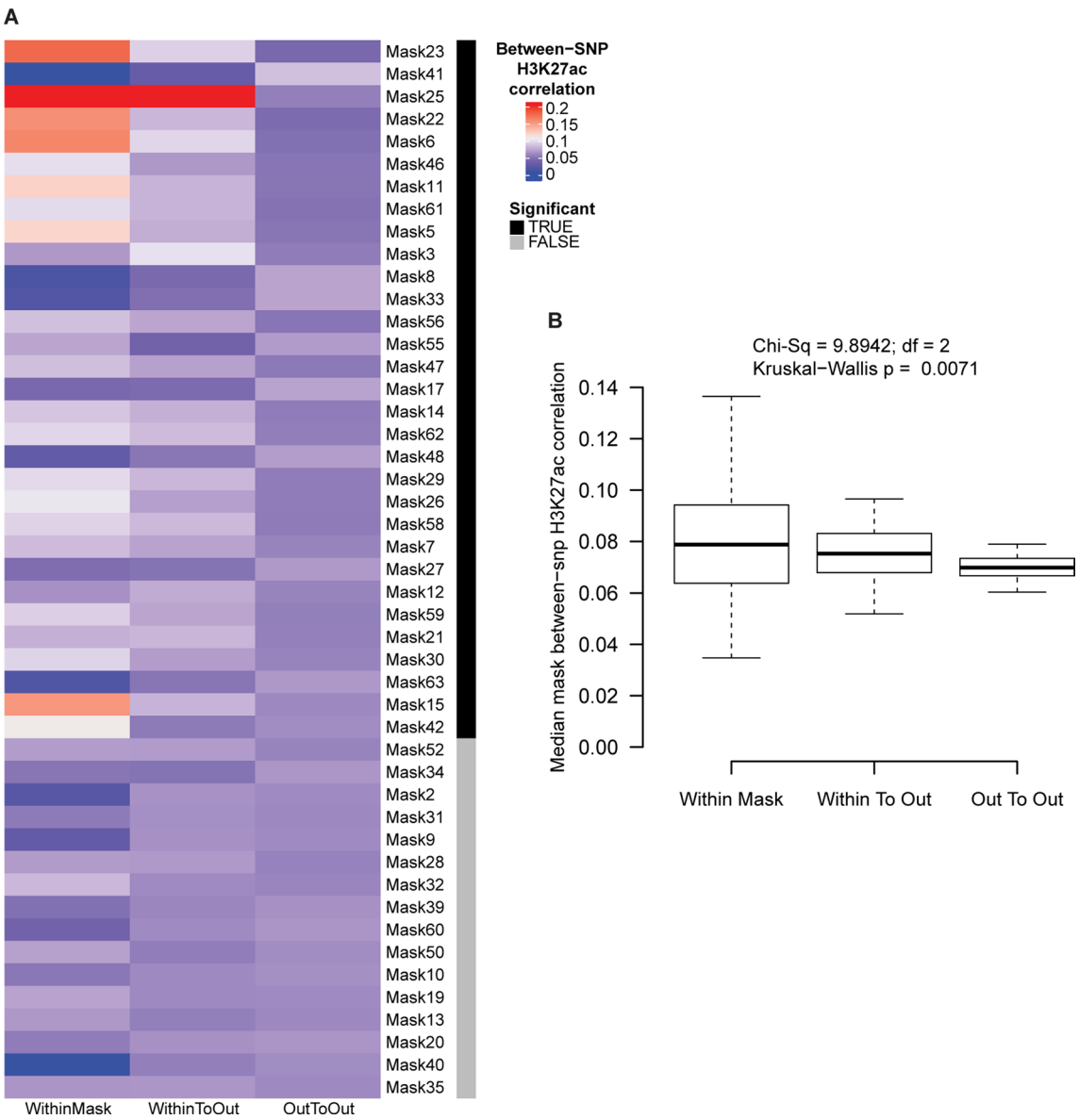
Extended Data Fig. 4 | Clustering of facial segments on the basis of shared genetic signals. Correlations between facial segments on the basis of SNP *P* values were calculated using LDSC, as described in Methods, and average-linkage hierarchical clustering was performed using the matrix of correlation values. Quadrant colors in legend refer to the quadrant of the polar dendrogram in which the facial segment lies in, also represented by the facial images at the top, and embryonic facial prominences are assigned to each facial segment.



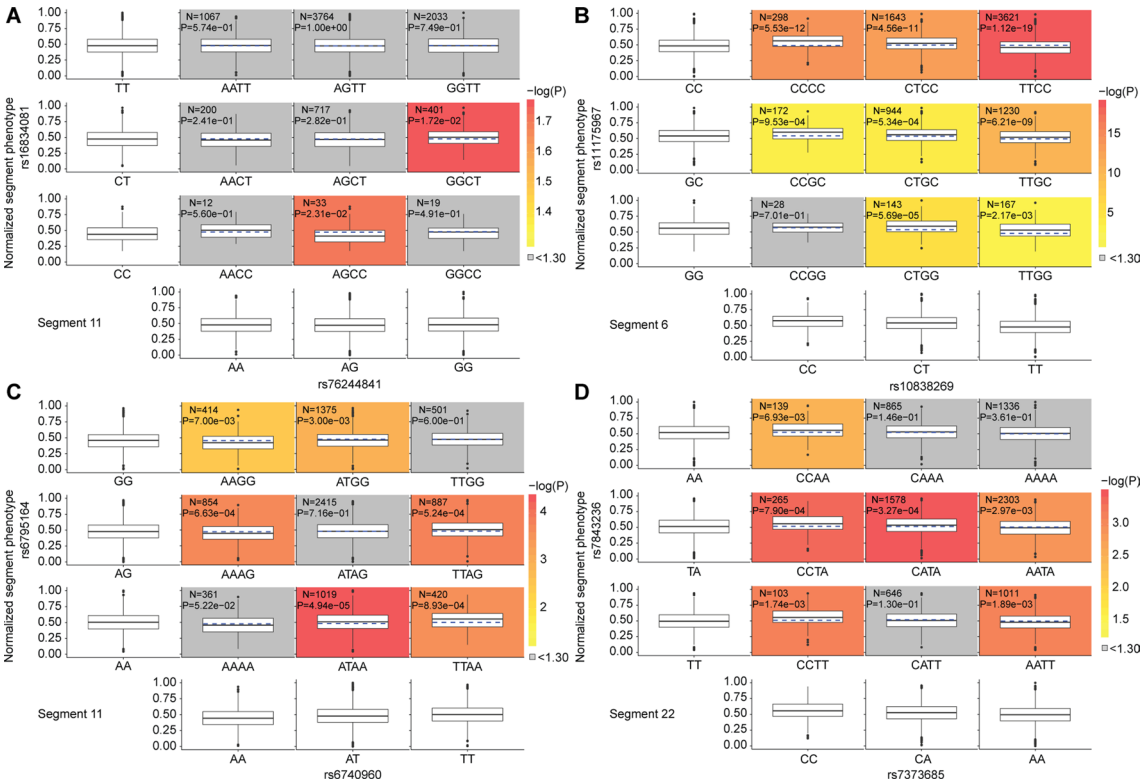
Extended Data Fig. 5 | GREAT and FUMA analyses showing enrichment for craniofacial and limb development. a, GREAT analysis. For the top ten GO terms in each category, plotted is the binomial test Bonferroni-corrected *P* value (red; negative values) and binomial region fold enrichment (blue; positive values). Behind every GO term, in parentheses we indicate the number of genes in the test set with the annotation (Observed) and the total number of genes in the genome with the annotation (Total), with the format (Observed/Total). Dashed line represents significance at $P = \log_{10}(0.05) = -1.3$. **b**, FUMA analysis, indicating the KEGG pathways that were significantly enriched in our results. Multiple pathways are relevant for craniofacial development. The right panel shows the genes that are involved in the pathways.



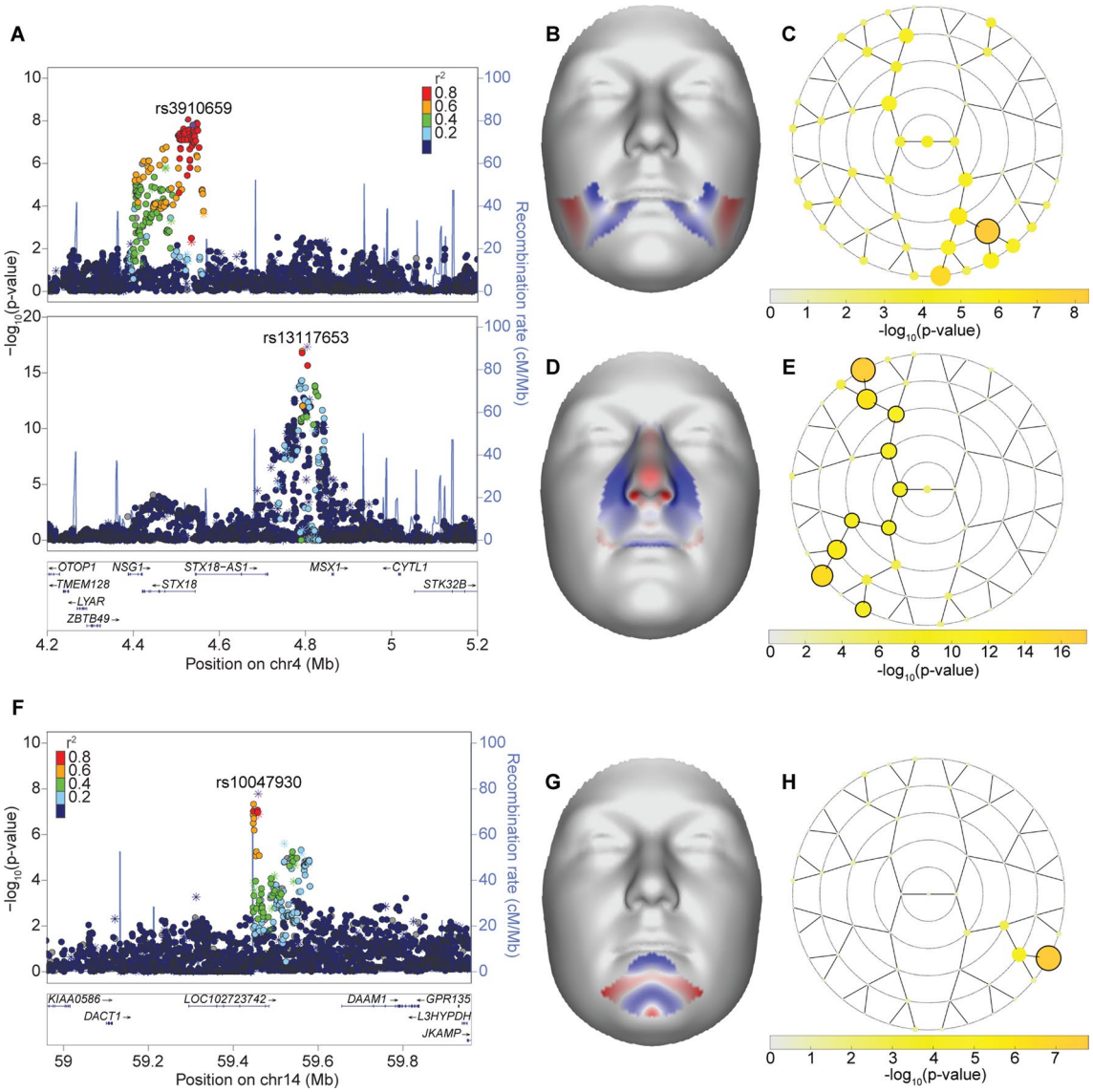
Extended Data Fig. 6 | H3K27ac signal is significantly different in 203 lead vs. 203 random SNPs for relevant facial tissues. For all cell types and tissues, each represented by a point above, the median difference between H3K27ac RPM signal between the 203 lead SNPs vs. 203 random SNPs was tested for significance using a two-sided Wilcoxon rank-sum test. The thin dashed line represents the 5% false discovery rate P value of 0.0094, using the Benjamini–Hochberg method. Relative to the random, MAF-matched SNPs, the lead SNPs are significantly enriched for H3K27ac signal in many cell types, with the highest magnitude differences being from CNCCs (blue) and embryonic craniofacial tissues (orange). Test statistics used to create this plot are available in Supplementary Table 4.



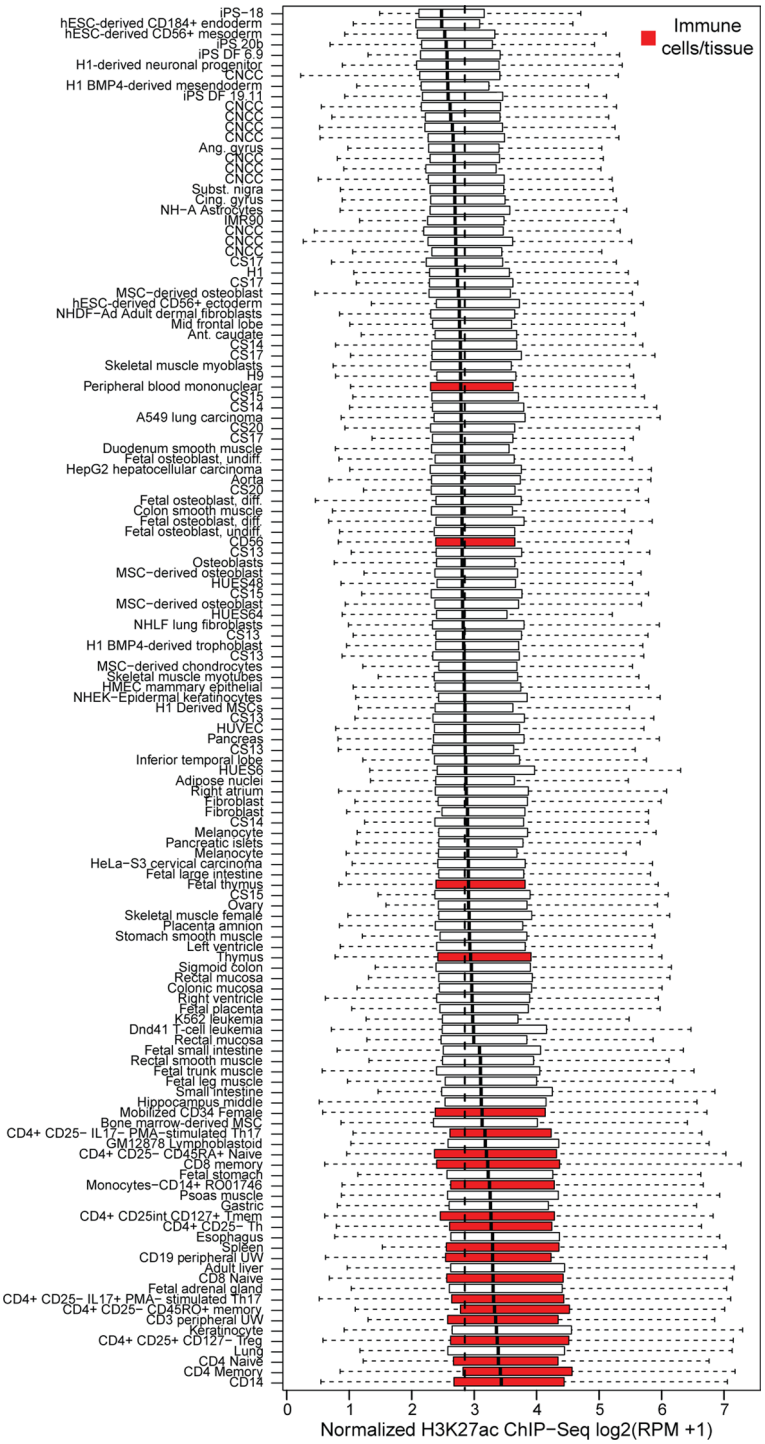
Extended Data Fig. 7 | Correlation of H3K27ac activity among SEM models. a, For all segments (aka 'masks'), we compared the H3K27ac activity for significant SNPs from the refined SEM model for variation in that facial segment. Plotted is the Spearman's rho correlation between pairs of SNPs significant in the same SEM model ('Within Mask'); pairs of SNPs where one is from the SEM model and the other is not ('Within To Out'), and where both SNPs in the pair are from a different SEM model ('Out To Out'). Segments where the distribution of correlation across all cell types was significantly different (Benjamini–Hochberg adjusted $P < 0.05$) based on a two-sided Kruskal–Wallis test are indicated in black. **b**, For all cell types, the median correlation across all segments is plotted for each of the three SNP groupings. Significance between the means was determined using a two-sided Kruskal–Wallis test. Boxplots plot the first and third quartiles, with a dark black line representing the median. Whiskers extend to the largest and smallest values no further than 1.5 × the inter-quartile range from the first and third quartiles, respectively.



Extended Data Fig. 8 | Phenotypic and marginal distributions for diplotype combinations. For a random SNP pairing (**a**) and each significant epistasis pair (**b**–**d**), boxplots are plotted to visualize the epistatic effect on the phenotype. The marginal phenotypic medians of the singular genotypes (non-shaded boxplots) were used to calculate and visualize the predicted diplotype phenotypic distribution that would occur if the two genotypes were acting alone. The median phenotype was also calculated for each diplotype as the average of the marginal medians of the singular genotypes (blue dashed lines on the colored plots). This median was compared to the observed medians of the diplotypes (solid black lines; colored boxplots) via Mood's Median test with one degree of freedom. Log-transformed *P* values were used to color boxplots if there was a significant ($P < 0.05$; $\log(P) > 1.30$) difference between the expected phenotype of the combined genotype and observed diplotype. Boxplots plot the first and third quartiles, with a dark black line representing the median. Whiskers extend to the largest and smallest values no further than 1.5 × the inter-quartile range from the first and third quartiles, respectively.



Extended Data Fig. 9 | *MSX1* and *DACT1* loci. LocusZoom plots for the two association signals nearby *MSX1* (**a**), which has previously been implicated in orofacial clefting in humans and mice, and *DACT1* (**f**), which is a novel result. Points represent one-sided -$\log_{10}(P)$ of the $META_{UK}$ meta-analysis track for the facial segment illustrated in the normal displacement figures (**b**, **d**, **g**) and are colored based on linkage disequilibrium with the labeled SNP. Asterisks indicate genotyped SNPs and circles indicate imputed SNPs. Facial effects for the two association signals nearby *MSX1*: rs3910659 (**b**) and rs13117653 (**d**) and the signal nearby *DACT1*: rs10047930 (**g**). Effects are the normal displacement (displacement in the direction locally normal to the facial surface) in each quasi landmark of the lowest facial segment reaching genome-wide significance in $META_{UK}$, going from the minor to the major allele. Blue indicates inward depression; red indicates outward protrusion. Yellow rosette plots depict the -$\log_{10}(P)$ of the meta-analysis *P* value (one-sided, right-tailed) per facial segment in $META_{UK}$ track. Black-encircled facial segments have reached genome-wide significance ($P = 5 \times 10^{-8}$). (**c**) rs3910659; (**e**) rs13117653; (**h**) rs10047930.



Extended Data Fig. 10 | Regions nearby previously published SNPs associated with risk for Crohn's disease are preferentially active in immune cells and tissues. Each boxplot represents the distribution of H3K27ac signal in 20 kb regions around 619 Crohn's disease-associated SNPs from the NCBI-EBI GWAS catalog in one sample. See Methods for details on calculation of H3K27ac signal. Samples corresponding to immune cells and tissues are highlighted in red. Thin dashed line at ~2.9 is the median level of signal across all cell types and tissues. Boxplots plot the first and third quartiles, with a dark black line representing the median. Whiskers extend to the largest and smallest values no further than 1.5 × the inter-quartile range from the first and third quartiles, respectively.

nature research

Corresponding author(s): Julie D. White

Last updated by author(s): Oct 16, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Three-dimensional images composed of surface and texture maps were taken using the 3dMD Face (3dMD, Atlanta, GA) and Vectra H1 (Canfield Scientific, Parsippany, NJ) 3D imaging systems, or the Konica Minolta Vivid 900 (Konica Minolta Sensing Europe, Milton Keynes, UK) laser scanner. Images gathered using the Konica laser scanner were processed using a macro algorithm in Rapidform 2004 software (INUS Technology Inc., Seoul, South Korea). Genotyping was performed using the Illumina OmniExpress + Exome v1.2 array, the 23andMe v3 and v4 arrays (Mountain View, CA), the Illumina Infinium Multi-Ethnic Global-8 v1 array, or the Illumina Human Hap550 quad array.
Data analysis	KU Leuven provides the MeshMonk spatially dense facial mapping software (v0.0.6), free to use for academic purposes (https://github.com/TheWebMonks/meshmonk). Matlab 2017b implementations of the hierarchical spectral clustering to obtain facial segmentations are available from a previous publication (https://doi.org/10.6084/m9.figshare.7649024). The statistical analyses in this work were based on functions of the statistical toolbox in Matlab 2017b, SHAPEIT2 (v2.r900), Sanger Imputation Server (v0.0.6), PBWT pipeline (v3.1), MeshMonk (v0.0.6), LDSC (v1.0.1), FUMA (v1.3.3), GREAT (v3.0.0), Plink 1.9, lavaan (v0.6-3), R (>3.4), agricolae (v1.3-0), cowplot (v1.0.0), ggplot2 (v3.1.1), ggpubr (v0.2), gridExtra (v2.3), gtable (v0.3.0), grid (v3.6.2), Hmisc (v4.2-0), psych (v1.8.12), data.table (v1.12.0), Genotype Harmonizer (v1.4.20), KING (v2.1.3), bowtie2 (v2.3.4.2), bedtools (v2.27.1), and bioconductor (v3.7) as mentioned throughout the Methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All of the genotypic markers for the 3DFN dataset are available to the research community through the dbGaP controlled-access repository (http://www.ncbi.nlm.nih.gov/gap) at accession #phs000929.v1.p1. The raw source data for the phenotypes - the 3D facial surface models in .obj format - are available through the FaceBase Consortium (https://www.facebase.org) at accession #FB00000491.01. Access to these 3D facial surface models requires proper institutional ethics approval and approval from the FaceBase data access committee. Additional details can be requested from SMW [smwst46@pitt.edu].

The participants making up the PSU and IUPUI datasets were not collected with broad data sharing consent. Given the highly identifiable nature of both facial and genomic information and unresolved issues regarding risk to participants, we opted for a more conservative approach to participant recruitment. Broad data sharing of the raw data from these collections would thus be in legal and ethical violation of the informed consent obtained from the participants. This restriction is not because of any personal or commercial interests. Additional details can be requested from MDS [mds17@psu.edu] and SW [walshsus@iupui.edu] for the PSU and IUPUI datasets, respectively.

The ALSPAC (UK) data will be made available to bona fide researchers on application to the ALSPAC Executive Committee (http://www.bris.ac.uk/alspac/researchers/data-access). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

KU Leuven provides the MeshMonk (v0.0.6) spatially dense facial mapping software, free to use for academic purposes (https://github.com/TheWebMonks/meshmonk). Matlab 2017b implementations of the hierarchical spectral clustering to obtain facial segmentations are available from a previous publication (https://doi.org/10.6084/m9.figshare.7649024.v1).

Publicly available data used were: 1000G Phase 3 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), the list of HapMap 3 SNPs excluding the MHC region provided by LDSC (http://ldsc.broadinstitute.org/static/media/w_hm3.noMHC.snplist.zip), and ChIP-seq files from Prescott et al. (GSE70751), Najafova et al. (GSE82295), Baumgart et al. (GSE89179), Nott et al. (https://genome.ucsc.edu/s/nottalexi/glassLab_BrainCellTypes_hg19), Pattison et al. (GSE119997), Wilderman et al. (GSE97752) and the Roadmap Epigenomics Project (https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/).

Meta-analysis GWAS statistics are available on GWAS Catalog (GCP000044). All relevant data to run future replications and meta-analysis efforts are provided in the FigShare repository for this work34, along with additional figures (https://doi.org/10.6084/m9.figshare.c.4667261). Items available in the FigShare repository are: (1) Anthropometric mask: a Matfile of the anthropometric mask used; (2) Association statistics and effects of the 203 lead SNPs: Facial effects, LocusZoom plots, and association statistics from each stage of the analysis for the 203 lead SNPs; (3) Calculation of study-wide significance threshold: Script and permutation outcomes needed to replicate the calculation of the study-wide significance threshold; (4) Facial segment assignments: Segment assignments for each quasi landmark in the anthropometric mask; (5) Figure 2A labeled: A larger version of Figure 2A, with all cell types and tissues labeled; (6) GREAT Export: Raw output of the GREAT analysis; (7) PCA shape constructs: PCA shape spaces for all 63 facial segments; (8) QQ plots: QQ plots for each segment in all stages of the analysis; (9) Script to explore facial segments and GWAS hits: MatLab script for select data exploration functions; (10) SNPs reaching suggestive significance in either meta-analysis track: Association statistics of all SNPs with $P < 5 \times 10\text{-}7$ in METAUS or METAUK tracks; (11) Source data for manuscript figures: Source data in Excel format for all figures, where possible.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was determined by the amount of 3DFN data available in the public facial data repository (Facebase.org), the amount of data available in the B2261 ALSPAC study, and by the number of individuals of European descent with genotype data and 3D facial images that were collected with informed consent as part of several studies based at The Pennsylvania State University and Indiana University Purdue University Indianapolis. More information is found in Methods.
Data exclusions	For both US and UK participants, outlier 3D facial images, likely caused by image mapping errors, were identified using two approaches. First, as described in prior work, outlier faces were identified by calculating z-scores from the Mahalanobis distance between the average face and each individual face. Faces with z-scores higher than two were manually investigated. Second, a score was calculated that reflects the missing data present in the image due to holes, spikes, and other mesh artifacts, which can be caused by facial hair or errors during the preprocessing steps. Images with high scores, indicating large gaps in the mesh, were manually investigated. During the manual check, the images were either classified as poor quality and removed or were preprocessed and mapped again. For US cohorts, genotype samples were excluded if there was poor concordance of genetic and reported sex, evidence of chromosomal aberrations, missing genotype call rate > 10%, and heterozygosity values ±3 standard deviations from the sample mean. For the 3DFN sample, 3D images and genotype data were obtained from the 3D Facial Norms repository. Recruitment was limited to

individuals aged 3 to 40 years old and of self-reported European ancestry. Individuals were excluded if they reported a personal or family history of any birth defect or syndrome affecting the head or face, a personal history of any significant facial trauma or facial surgery, or any medical condition that might alter the structure of the face. The intersection of unrelated participants with quality-controlled images, covariates, and genotype data from individuals of European descent resulted in 1,906 individuals for analysis.

The PSU sample included 3D images and genotypes of participants recruited through several studies at the Pennsylvania State University. Individuals were excluded from the analysis if they were below 18 years of age and if they reported a personal history of significant trauma or facial surgery, or any medical condition that might alter the structure of the face. No restriction on ancestry or ethnicity was imposed during recruitment, but only individuals of European descent were used in this study. The intersection of unrelated European participants with quality-controlled images, covariates, and genotype data resulted in 1,990 individuals for analysis.

The IUPUI sample includes 3D images and genotypic data from individuals recruited in Indianapolis, IN and Twinsburg, OH. Individuals who were below 18 years of age were recruited if they had a parent or legal guardian's signature. Similar to the PSU sample cohort, no restrictions were placed on the recruitment of participants, but only unrelated individuals of European descent, without significant facial injury or medical condition, and those meeting all quality control criteria were used in this study (n = 784).

The UK sample was derived from the ALSPAC dataset, a longitudinal birth cohort in which pregnant women residing in Avon with an expected delivery date between 1 April 1991 and 31 December 1992 were recruited. At the time, 14,541 pregnant women were recruited and DNA samples were collected for 11,343 children. Genome-wide data was available for 8,952 subjects and of the B2261 study, titled "Exploring distinctive facial features and their association with known candidate variants." In addition to this, 4,731 3D images were available. UK genotype samples were excluded on the basis of genetic sex and reported gender mismatches, minimal or excessive heterozygosity, disproportionate levels of individual missingness (>3%), and insufficient sample replication (IBD <0.8). The intersection of unrelated participants of European ancestry with quality-controlled images, covariates, and genotype data included 3,566 individuals.

Replication

Replication was achieved by proper separation of the data into identification and verification datasets, based on completely separate sampling, imaging, genotyping, and imputation. More information is found in Methods.

Randomization

No randomization took place, group membership of identification and verification was determined by the separately obtained datasets available. Because canonical correlation analysis does not accommodate adjustments for covariates, we removed the effect of relevant covariates (sex, age, age-squared, height, weight, facial size, the first four genomic ancestry axes, and the camera system), on both the independent (SNP) and the dependent (facial shape pre segmentation) variables using partial least squares regression (plsregress from Matlab 2017b), and thus performed the canonical correlation analysis under a reduced model with residualized variables.

Blinding

Blinding was not relevant to this study, as no treatment outcomes were assessed and data analysis procedures were standardized across all individuals. Two independent datasets were constructed from the sampling efforts of four different research centers, and the analysis was done by yet another research center.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Cranial neural crest cells (CNCCs) originated from WiCell (H9 ESC) and the Fred Gage laboratory (iPSC; Salk Institute), available from Prescott et al. (GSE70751)
Fetal osteoblast cell line, undifferentiated and differentiated, originated from Najafova et al. (GSE82295)
Mesenchymal stem cell-derived osteoblasts originated from Baumgart et al. (GSE89179)
Various brain cell types originated from Nott et al. (https://genome.ucsc.edu/s/nottalexi/glassLab_BrainCellTypes_hg19)
Surface ectoderm samples originated from Pattison et al. (GSE119997)
Embryonic craniofacial tissue originated from Wilderman et al. (GSE97752)
All other fetal and adult cell tissues and cell types originated from the Roadmap Epigenomics Project (https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/)

Sample type, ID, and URL for each cell type/tissue is available in Supplementary Table 4

Authentication	For the CNCCs, we analyzed the genomic sequence data from the lines. Please refer to the original publications for the authentication of the other cell types and tissues.
Mycoplasma contamination	For the CNCCs, we used PCR tests to test for mycoplasma contamination. Pease refer to the original publications for the testing of the other cell types and tissues.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	For the 3DFN sample, 3D images and genotype data were obtained from the 3D Facial Norms repository. The repository includes 3D facial surface images and self-reported demographic descriptors as well as basic anthropometric measurements from individuals recruited at four US sites: Pittsburgh, PA (PITT IRB PRO09060553 and RB0405013); Seattle, WA (Seattle Children's IRB 12107); Houston, TX (UT Health Committee for the Protection of Human Subjects HSC-DB-09-0508); and Iowa City, IA (University of Iowa Human Subjects Office IRB (200912764 and 200710721). Recruitment was limited to individuals aged 3 to 40 years old and of self-reported European ancestry. Individuals were excluded if they reported a personal or family history of any birth defect or syndrome affecting the head or face, a personal history of any significant facial trauma or facial surgery, or any medical condition that might alter the structure of the face. The intersection of unrelated participants with quality-controlled images, covariates, and genotype data from individuals of European descent resulted in 1,906 individuals for analysis (Female N = 1,172; Male N = 734). Average height of these participants was 163.43 cm (sd = 20.57 cm). Average weight of these participants was 64.33 kg (sd = 22.38 kg) The PSU sample included 3D images and genotypes of participants recruited through several studies at the Pennsylvania State University and sampled at the following locations: Urbana-Champaign, IL (PSU IRB 13103); New York, NY (PSU IRB 45727); Cincinnati, OH (UC IRB 2015-3073); Twinsburg, OH (PSU IRB 2503); State College, PA (PSU IRB 44929 and 4320); Austin, TX (PSU IRB 44929); and San Antonio, TX (PSU IRB 1278). Participants self-reported information on age, ethnicity, ancestry, and body characteristics, and data were gathered on height and weight. Individuals were excluded from the analysis if they were below 18 years of age and if they reported a personal history of significant trauma or facial surgery, or any medical condition that might alter the structure of the face. No restriction on ancestry or ethnicity was imposed during recruitment, but only individuals of European descent were used in this study. The intersection of unrelated European participants with quality-controlled images, covariates, and genotype data resulted in 1,990 individuals for analysis (Female N = 1,380; Male N = 610). Age ranged from 18 to 88 years old. Average height of these participants was 168.75 cm (sd = 9.23 cm). Average weight of these participants was 73.88 kg (sd = 17.05 kg). The IUPUI sample includes 3D images and genotypic data from individuals recruited in Indianapolis, IN and Twinsburg, OH (IUPUI IRB 1409306349). Participants self-reported information on age, height, weight, and ancestry at the time of the collection. Individuals who were below 18 years of age were included if they had a parent or legal guardian's signature. Similar to the PSU sample cohort, no restrictions were placed on the recruitment of participants, but only n = 784 individuals of European descent and those meeting all quality control criteria were used in this study (Female N = 539; Male N = 245). Age ranged from 7 to 78 years old. Average height of these participants was 169.24 cm (sd = 11.30 cm). Average weight of these participants was 71.88 kg (sd = 18.65 kg). The UK sample was derived from the ALSPAC dataset, a longitudinal birth cohort in which pregnant women residing in Avon with an expected delivery date between 1 April 1991 and 31 December 1992 were recruited. At the time, 14,541 pregnant women were recruited and DNA samples were collected for 11,343 children. Genome-wide data was available for 8,952 subjects of the B2261 study, titled "Exploring distinctive facial features and their association with known candidate variants." In addition to this, 4,731 3D images were available along with information on sex, age, weight, height, ancestry, and other body characteristics. The ALSPAC study website contains details of all the data that is available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/our-data/). The intersection of participants of European ancestry with quality-controlled images, covariates, and genotype data included 3,566 individuals (Female N = 1,884; Male N = 1,682). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). Age ranged from 14 to 17 years old. Average height of these participants was 169.38 cm (sd = 8.42 cm). Average weight of these participants was 61.52 kg (sd = 11.75 kg). For all datasets, there was no imbalance in gender, and weight and height distributions follow typical distributions seen in a European-derived population. We removed the effect of relevant covariates (sex, age, age-squared, height, weight, facial size, the first four genomic ancestry axes, and the camera system), on both the independent (SNP) and the dependent (facial shape pre segmentation) variables using partial least squares regression (plsregress from Matlab 2017b), and thus performed the canonical correlation analysis under a reduced model with residualized variables.
Recruitment	Two publicly available datasets, one from FaceBase and one from the ALSPAC project were used and did not involve any recruitment specific to this work. The other datasets used from PSU and IUPUI, and their characteristics, as mentioned above, were recruited over different studies and did not contain any specific selection or bias that might influence this work.
Ethics oversight	We have complied with all relevant ethical regulations for work with human participants and informed consent was obtained. Institutional review board (IRB) approval was obtained at each recruitment site and all participants gave their written informed consent prior to participation; for children, written consent was obtained from a parent or legal guardian. For the 3DFN sample, the following local ethics approvals were obtained: Pittsburgh, PA (PITT IRB PRO09060553 and RB0405013); Seattle, WA (Seattle Children's IRB 12107); Houston, TX (UT Health Committee for the Protection of Human Subjects HSC-DB-09-0508); and Iowa City,

IA (University of Iowa Human Subjects Office IRB (200912764 and 200710721). For the Penn State sample, the following local ethics approvals were obtained: Urbana-Champaign, IL (PSU IRB 13103); New York, NY (PSU IRB 45727); Cincinnati, OH (UC IRB 2015-3073); Twinsburg, OH (PSU IRB 2503); State College, PA (PSU IRB 44929 and 4320); Austin, TX (PSU IRB 44929); and San Antonio, TX (PSU IRB 1278). For the IUPUI sample, the following local ethics approvals were obtained: Indianapolis, IN and Twinsburg, OH (IUPUI IRB 1409306349). For the ALSPAC sample, approval was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Consent for biological samples was collected in accordance with the Human Tissue Act (2004).

Note that full information on the approval of the study protocol must also be provided in the manuscript.