Open access • Journal Article • DOI:10.1038/S41588-020-00741-7

Insights into the genetic architecture of the human face. — Source link

Julie D. White, Karlijne Indencleef, Sahin Naqvi, Ryan J. Eller ...+19 more authors

Institutions: Pennsylvania State University, Katholieke Universiteit Leuven, Stanford University, Indiana University – Purdue University Indianapolis ...+4 more institutions

Published on: 31 Jan 2021 - Nature Genetics (Nature Publishing Group)

Topics: Genetic architecture, Cranial neural crest and Genome-wide association study

Insights into the genetic architecture of the human face

Julie D. White[1†*], Karlijne Indencleef[2,3†*], 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], Peter Claes[2,3,7,18*]

Affiliations:
[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, United Kingdom
[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

†These authors contributed equally to this work.

*Correspondence to: jdw345@psu.edu; karlijne.indencleef@kuleuven.be; peter.claes@kuleuven.be

Author Contributions

Conceptualization (Ideas; formulation or evolution of overarching research goals and aims): P.C., M.D.S., S.M.W., J.R.S., J.W., S.W.

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., R.J.E., M.K.L., J.L., S.W., P.C.

Formal analysis (Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data): J.D.W., K.I., S.N., R.J.E., H.H., J.R., J.L., P.C.

Funding acquisition (Acquisition of the financial support for the project leading to this publication): 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., P.C.

Investigation (Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection): J.D.W., K.I., S.N., R.J.E., H.H., J.R., M.K.L., J.L., P.C.

Resources (Provision of study materials, computing resources, or other analysis tools): 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., M.D.S.

Supervision (Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team): P.C., S.M.W., M.D.S., S.W., J.W., J.R.S., M.L.M., T.S., H.P., G.H.

Visualization (Preparation, creation and/or presentation of the published work, specifically visualization/data presentation): J.D.W., K.I., S.N., R.J.E., H.H., J.R., M.K.L., P.C.

Writing (original draft): J.D.W., K.I., S.N., R.J.E., J.R.

Writing (review and editing): 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., P.C.

Abstract: The human face is complex and multipartite, and characterization of its genetic architecture remains challenging. Using a multivariate genome-wide association study (GWAS) 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 sum, our analyses provide insights for understanding how complex morphological traits are shaped by both individual and coordinated genetic actions.

Introduction:

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 non-coding 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 US and UK (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 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. 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 independently subjected to a Generalized Procrustes analysis, after which principal components 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, though 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 single-nucleotide polymorphisms (SNPs) 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 (though 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 which 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), though 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 (i.e. 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 primarily corresponding 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 (e.g. 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[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; 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].

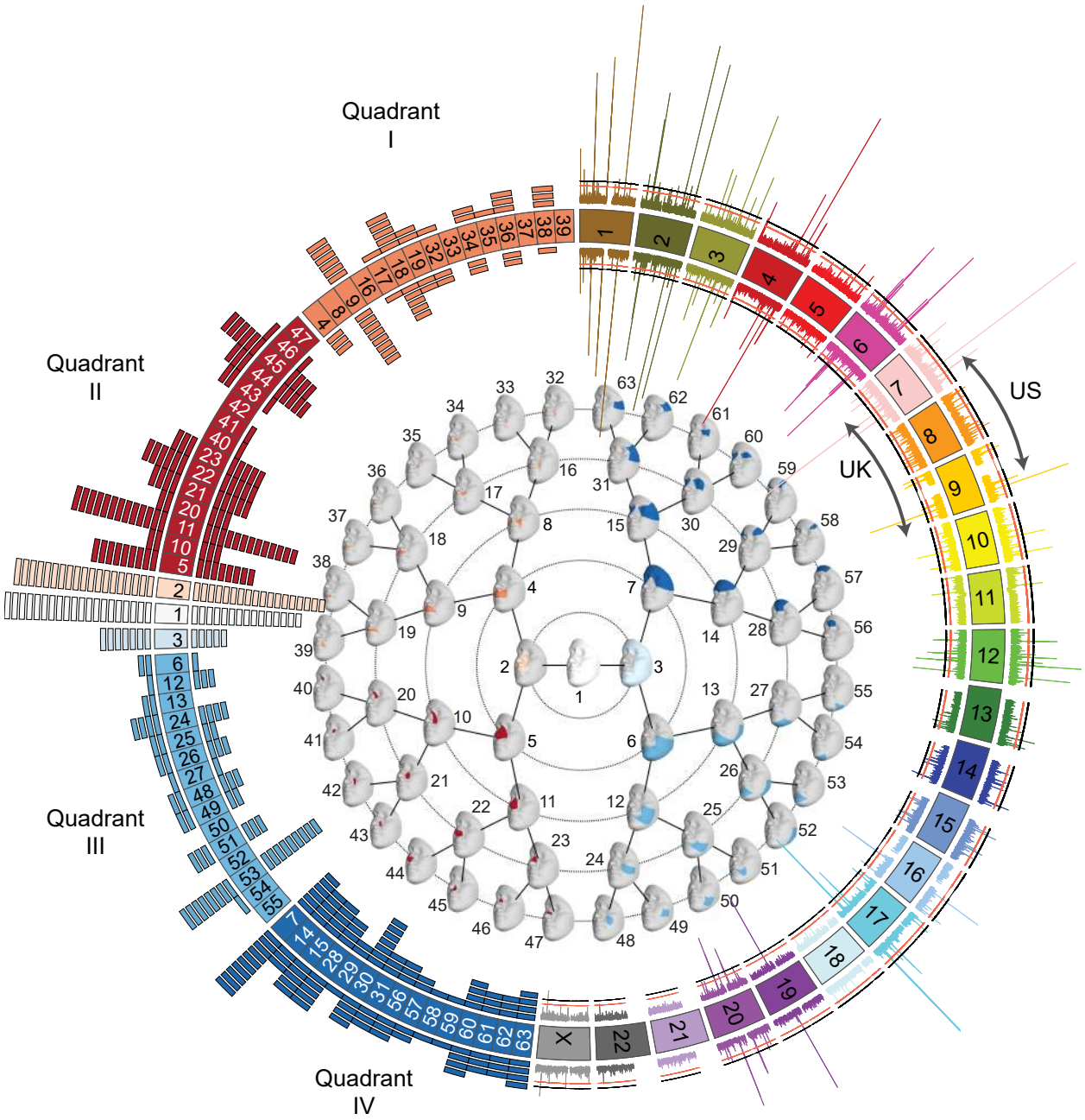


Figure 1. Overall results of US-driven and UK-driven meta-analyses. On the left, numbered blocks representing the 63 facial segments arranged and colored according to quadrant (I =

orange; II = red; III = light blue; IV = dark blue), and 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, 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 ($P = 5 \times 10^{-8}$) and the black line represents the study-wide threshold ($P = 6.96 \times 10^{-10}$). Created using Circos v0.69-8[35].

The global-to-local approach means that we often identified associations between a single SNP and variation in many facial segments. In this manuscript, we primarily focus on the segment in which the SNP had its lowest *P* value (the 'Best segment') and provide information on which meta-analysis track (Meta-US or Meta-UK) in which the SNP reached this significance level (the 'Best meta-analysis track'). Thus, throughout the rest of the manuscript, the reported *P* values for each SNP will be in the format of $P_{Best\ track\ (Best\ segment)} = value$. 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[36] 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[37] 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 marker of the promoters of transcriptionally active genes and active distal enhancers[38,39], from approximately 100 different cell types and tissues, including cranial neural crest cells (CNCCs), fetal and adult osteoblasts, 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; Supplementary Table 4).

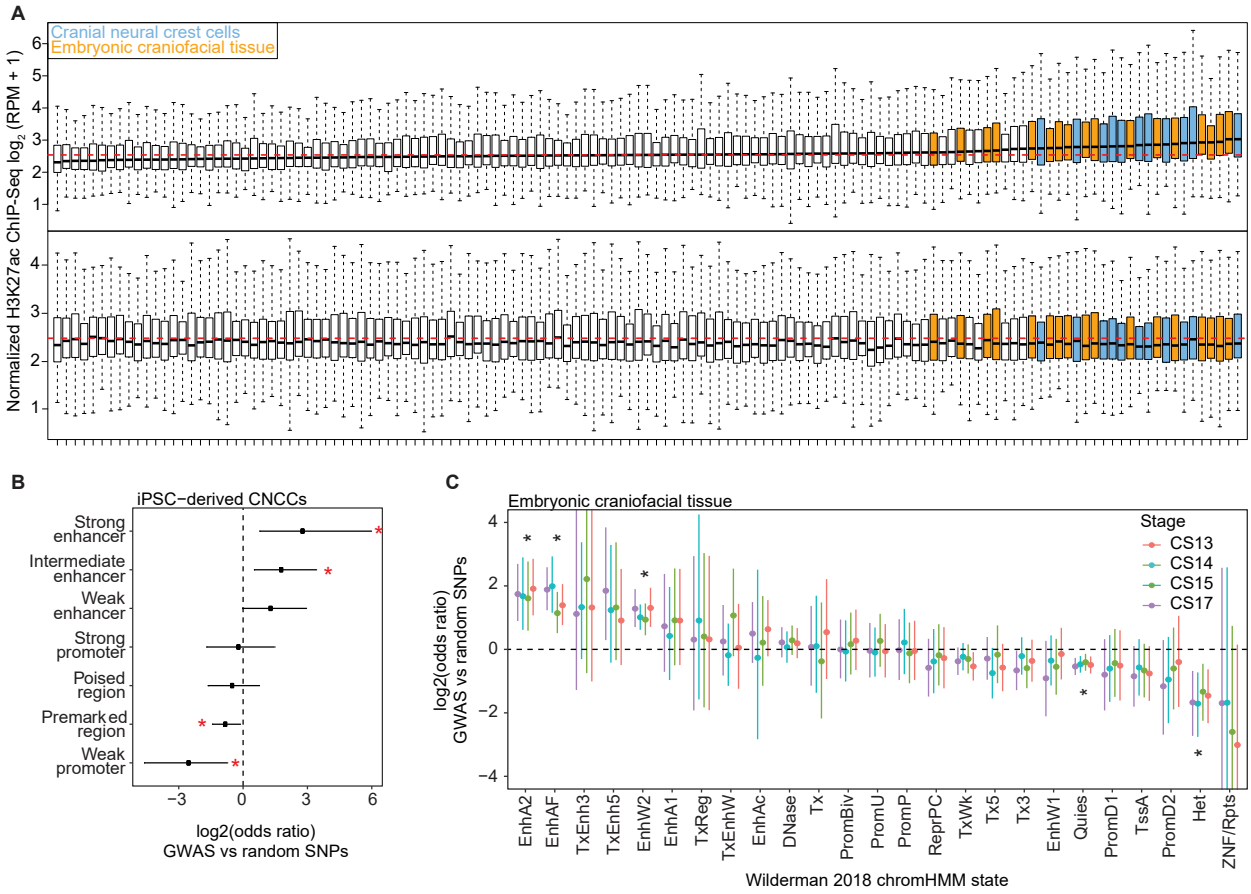


Figure 2. Regions near the 203 genome-wide significant lead SNPs are enriched for enhancers preferentially active in cranial neural crest cells and embryonic craniofacial tissue. (A) Each boxplot represents the distribution of H3K27ac signal in 20-kb regions around the 203 genome-wide significant lead SNPs (*top*) or 203 random SNPs (*bottom*) in one sample, with cranial neural crest cells and embryonic craniofacial tissue 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 inter-quartile range from the first and third quartiles, respectively. The dashed red lines represent the median level of H3K27ac RPM signal across all cell types and tissues. A larger labeled version of (A) is available in the FigShare repository[34]. For each class of regulatory element in either (**B**) cranial neural crest cells or (**C**) embryonic craniofacial tissue, 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 using a two-sided Fisher's exact test. Points represent estimated odds ratio and surrounding bars represent 95% confidence intervals. Asterisk indicates 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.

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[40,41]. In the CNCCs, candidate regulatory regions in the vicinity of the 203 genome-wide significant lead SNPs were significantly enriched 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 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 (e.g. cluster 5), representing activity in an early time point in development. Interestingly, however, some SNPs showed preferential activity for either CNCCs or craniofacial tissue (e.g. 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.

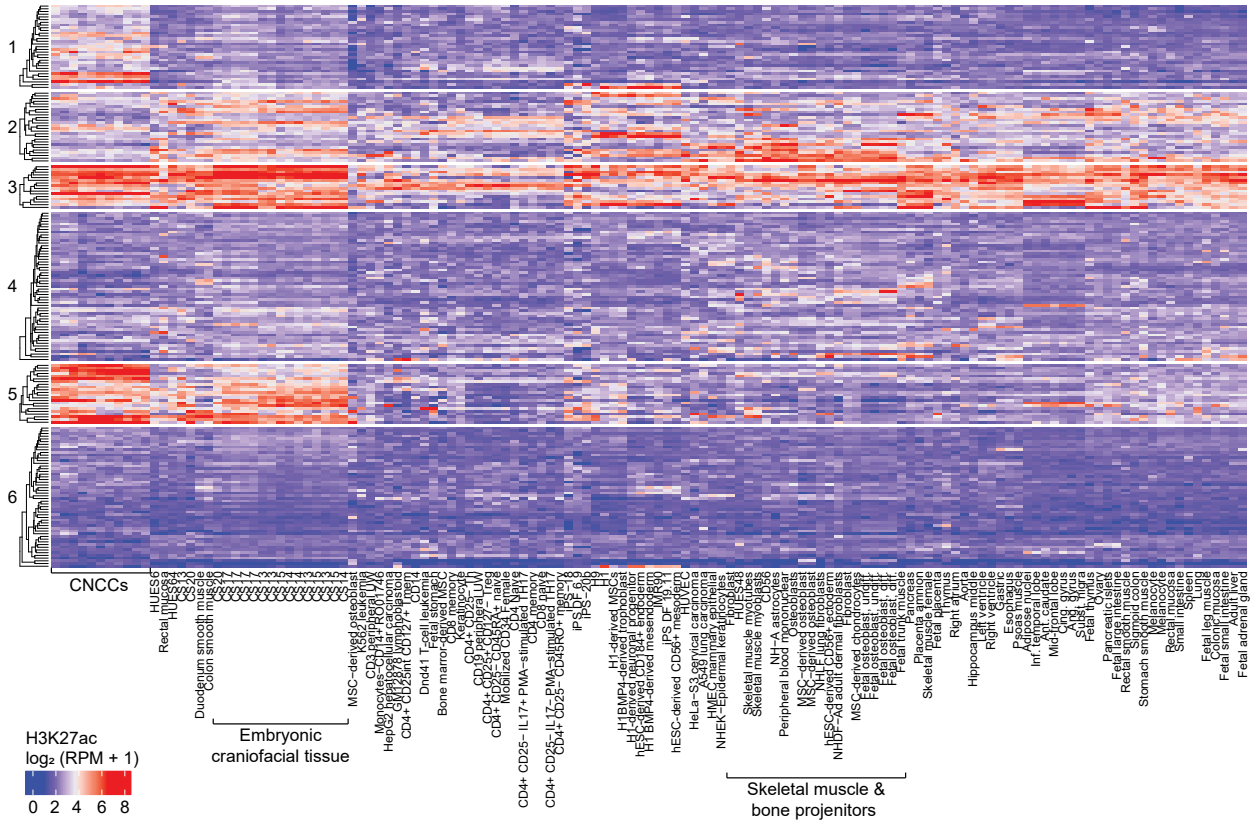


Figure 3. Activity of 203 genome-wide-significant lead SNPs in all cell-types studied. H3K27ac signal calculation and k-means clustering of SNPs were performed as described in Methods. Average linkage clustering on Euclidean distances was performed both within each of the 6 row clusters and for all columns. Descriptions of all mnemonics can be found at: https://egg2.wustl.edu/roadmap/web_portal/meta.html

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 significant (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-significant (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* 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 ($P_{Meta\text{-}US(Seg.\ 14)} = 8.01 \times 10^{-58}$). Interestingly, this lead SNP overlaps in location with a SNP not originally identified in our peak selection approach, rs12027501 ($P_{Meta\text{-}US(Seg.\ 1)} = 1.03 \times 10^{-41}$). 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 ($P_{Meta\text{-}UK(Seg.\ 51)} = 7.03 \times 10^{-13}$), 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 ($P_{Meta\text{-}UK(Seg.\ 48)} = 3.26 \times 10^{-15}$). Of interest, we observed twenty-four

such loci with multiple genome-wide significant peaks that are each associated with different facial traits (Supplementary Table 5, Supplementary Data 1).

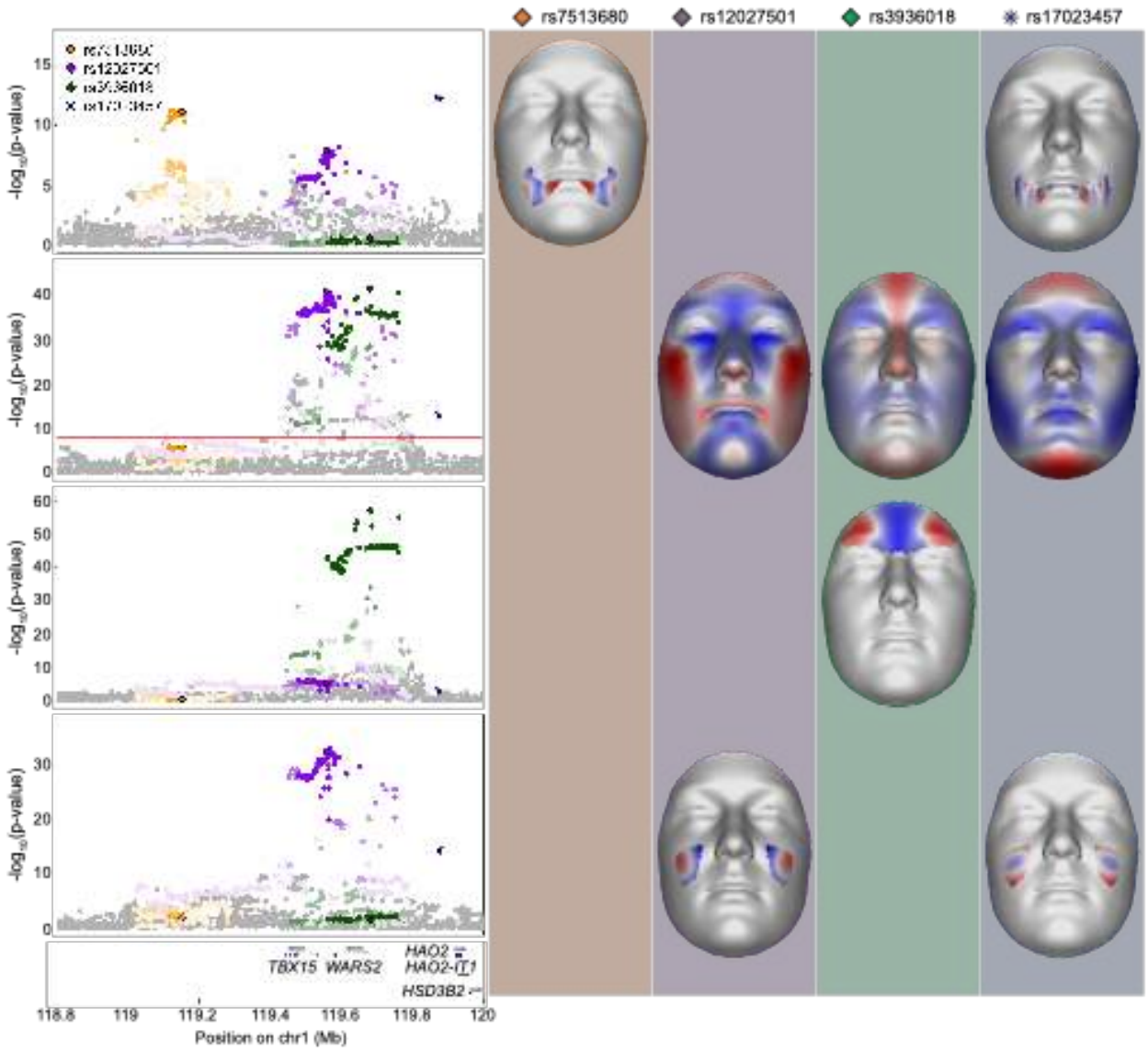


Figure 4. *TBX15-WARS2* multi-peak locus. LocusZoom[33] plots and facial effects for four association signals near the *TBX15-WARS2* locus. Clustering based on r^2 was performed to separate non-correlated signals, resulting in the separation of four SNPs. Color for each SNP is based on cluster association, 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 inwards.

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 (e.g. 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 non-significant 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; Supplementary Note 3).

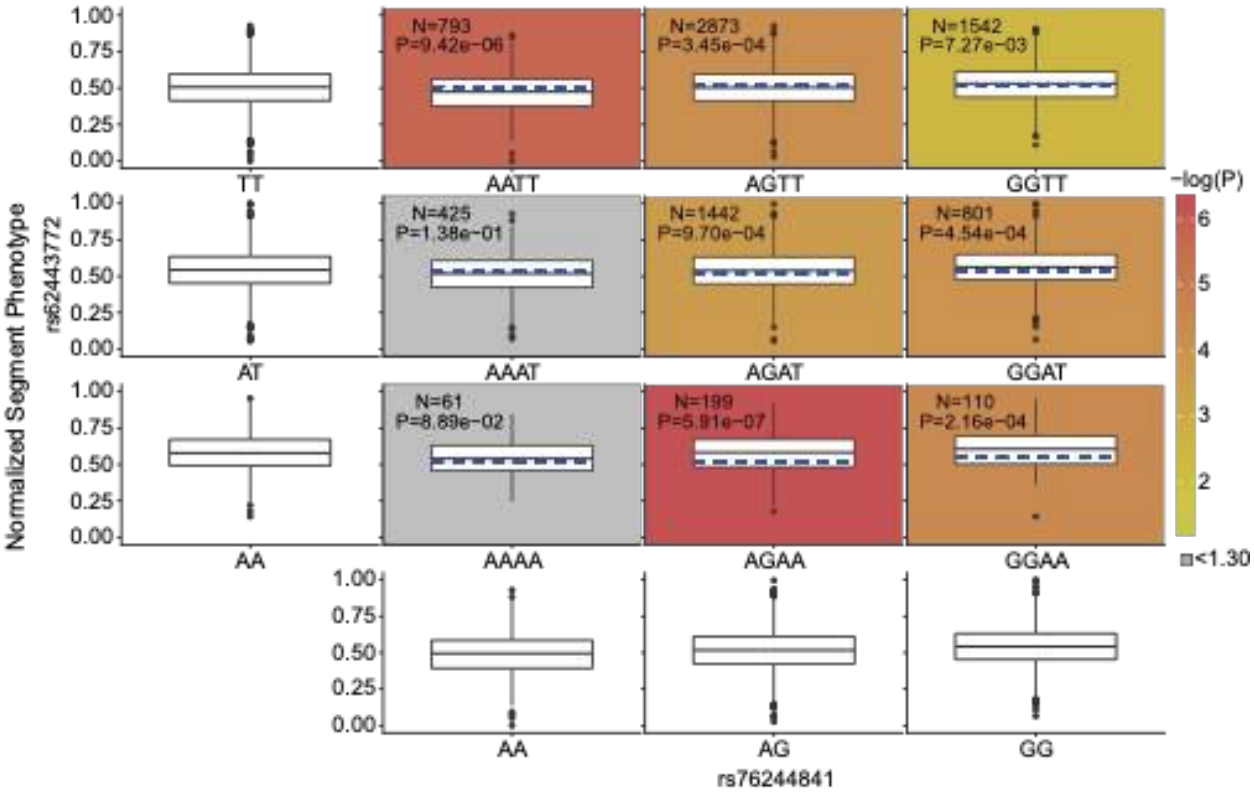


Figure 5. Phenotypic and marginal distributions for the rs62443772 - rs76244841 epistatic pair. Plotted in the first column and last row are the marginal phenotypic distributions of the genotypes, which shows the phenotypic distribution that would occur if the two genotypes were acting alone. The median 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[42] with one degree of freedom. The resulting log transformed *P* value was used to color the boxplots to illustrate significance, unless the difference was non-significant, in which the color was automatically set to grey. 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 inter-quartile range from the first and third quartiles, respectively.

Table 1. Four SNPs with evidence of epistatic interactions. 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 1.9, with Bonferroni multiple testing correction. For the four SNP pairs with significant evidence of epistatic interactions, this table lists the epistasis *P* value, rsID, GRCh37 location, and gene annotation. The phenotypic and marginal distributions for the pairs are depicted as boxplots in Figure 5 and Extended Data Fig. 8.

	SNP 1	SNP 2	Test	*P* value

Segment	RSID	Location	Annot. Gene	RSID	Location	Annot. Gene	statistic	
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}

Discussion

In their review, Atchley and Hall 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 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 sum, 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, 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 "hot spots" 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 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 (e.g. Dlx homeobox genes, BMP genes, and *FGFR2*) have well-established roles in limb development[43]. These findings are also

supported by the large number of human syndromes that present with both facial and limb malformations[44].

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 (a) activity of our 203 genome-wide-significant lead SNPs in CNCCs and embryonic craniofacial tissues and (b) 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 time points. 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 and 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[45]. 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 (e.g. 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.

Sixty-one 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[46,47] and mice[47,48], and is also widely expressed in lip and dental tissues during development[49]. 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; $P_{Meta\text{-}US(Seg.\ 34)} = 4.2 \times 10^{-18}$) 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; $P_{Meta\text{-}UK(Seg.\ 25)} = 4.45 \times 10^{-9}$; Extended Data Fig. 9A-E). This result could indicate a potential role of *STX18* in craniofacial development, though the STX18 protein is primarily important 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[45].

We also identified 53 genome-wide-significant signals in regions harboring genes with no previously known role in craniofacial development or disease, though 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[50], though *DACT1* is mostly studied 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[51]. These novel signals are promising new candidates of 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, 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, linkage disequilibrium (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 structural equation model (SEM) for that segment, showed higher correlations of cross-sample H3K27ac activity than when compared to SNPs which 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; $P_{Meta\text{-}UK(Seg.\ 30)} = 1.48 \times 10^{-8}$) and rs62443772 (*GLI3* associated; $P_{Meta\text{-}UK(Seg.\ 22)} = 5.35 \times 10^{-16}$) 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; Fig. 5). Interestingly, *PRDM16* and *GLI3* are both part of a tetrameric Hedgehog signaling complex in *Drosophila melanogaster* (Supplementary Note 3)[52–54]. 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 non-human morphology.

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 prior to participation. For children, written consent was obtained from a parent or legal guardian. Some individuals from the 3DFN and PSU samples were previously tested 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, 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 (v2.r900)[59] and imputation to the 1000G Phase 3 reference panel[60] performed using the positional Burrows-Wheeler Transform[61] pipeline (v3.1) of the Sanger Imputation Server (v0.0.6)[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 principal component (PC) space constructed using the 1000G Phase 3 dataset, first filtered for linkage disequilibrium 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 k-th 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 similar manner as 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, Atlanta, GA) camera system. PSU images were obtained with either the 3dMDface or Vectra H1 system (Canfield Scientific, Parsippany, NJ). 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, Milton Keynes, UK). 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., Seoul, South Korea).

3D image registration and quality control

The 3D surface images and their reflections were registered using the MeshMonk registration framework (v0.0.6)[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 greatly differing from the norm or with large holes were manually investigated 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 to only 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 strongly correlated 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 independently subjected 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 canonical correlation analysis (CCA, canoncorr in Matlab 2017b). CCA is a multivariate analysis which 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 sum, 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 non-identification 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], chosen because a meta-analysis of beta values was not possible given that the CCA returns a positive correlation value, not beta statistic. The entire process was repeated, resulting in two meta-analysis *P* values ($P_{\text{Meta-US}}$ and $P_{\text{Meta-UK}}$) accompanied by two identified traits per segment and per SNP: first using US in the identification stage and UK as verification (META_{US} or US-driven), then using UK in the identification stage and US 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 is 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 5×10^{-8} / (1.0042 × 1.6631 × 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 manuscript[34]. Though 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 (i.e. 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 (v1.3.3)[37]. 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 (v1.3.3)[37], which can test for enrichment of a set of genes in pre-defined pathways, and GREAT (v3.0.0)[36], which predicts the function of cis-regulatory regions, were performed using preset parameters (Extended Data Fig. 5). In this manuscript, 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 nearest gene using SNPsnap[81], or 619 Crohn's disease-associated SNPs from the 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.[41] 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 structural equation modeling (SEM) using the lavaan package (v0.6-3) in R ($\geq$ 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 (*i.e.* 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 non-hierarchically related segments (*i.e.* 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 1.9[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; Extended Data Fig. 8) to assess the genotypic contribution to epistatic masking (*i.e.* the combination of two variants reduce the phenotype) and boosting (*i.e.* 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[42] with one degree of freedom. These steps were performed using the R packages 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), and data.table (v1.12.0).

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 #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 #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.[41] (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 relevant data 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) *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 $P < 5 \times 10^{-7}$ in $META_{US}$ or $META_{UK}$ tracks; (11) *Source data for manuscript figures*: Source data in Excel format for all figures, where possible.

Code availability

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[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 (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 (>v3.4), 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.

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, PD/PIs: Marazita/Weinberg; R01-DE016148, PD/PIs: Marazita/Weinberg; and R01-DE027023, PD/PIs: Weinberg/Shaffer). Funding for genotyping by the National Human Genome Research Institute (X01-HG007821 and X01-HG007485, PD/PI: Marazita) 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: Norton), the Center for Human Evolution and Development at Penn State, the Science Foundation of Ireland Walton Fellowship (04.W4/B643, PD/PI: Shriver), the US National Institute of Justice (2008-DN-BX-K125, PD/PI: Shriver; and 2018-DU-BX-0219, PD/PIs: Walsh), 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: Walsh).

University of Cincinnati personnel and data collection were supported by Procter & Gamble, Company (UCRI-2015-1117-HN-532, PD/PI: Norton).

The UK Medical Research Council and Wellcome (Grant ref: 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: Weinberg/Shaffer), The Research Fund KU Leuven (BOF-C1, C14/15/081 and C14/20/081, PD/PI: Claes), The Research Program

of the Research Foundation – Flanders (FWO, G078518N, PD/PI: Claes), and a Senior Clinical Investigator Fellowship of The Research Foundation – Flanders (G078714N, PD/PI: Hens).

Stanford University personnel and analyses were supported by the National Institute of Dental and Craniofacial Research (R01-DE027023, PD/PIs: Weinberg/Shaffer; and U01-DE024430, PD/PIs: Wysocka/Selleri), the Howard Hughes Medical Institute, and the March of Dimes Foundation (1-FY15-312, PD/PI: Wysocka).

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.

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**, 1–11 (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**, 1–21 (2016).
13. Cole, J. B. *et al.* Genomewide 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**, 1–13 (2017).
15. Crouch, D. J. M. *et al.* Genetics of the human face: Identification of large-effect single gene variants. *Proc. Natl. Acad. Sci. U. S. A.* **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**, (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**, 601–9611 (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 Jr., R. M. *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. (2020) doi:10.6084/m9.figshare.c.4667261.
35. Krzywinski, M. I. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
36. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

37. Watanabe, K., Taskesen, E., Bochoven, A. van & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

38. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).

39. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).

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. Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).

42. Brown, G. W. & Mood, A. M. On Median Tests for Linear Hypotheses. in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (ed. Neyman, J.) 159–166 (University of California Press, 1951).

43. Kraus, P. & Lufkin, T. Dlx homeobox gene control of mammalian limb and craniofacial development. *Am. J. Med. Genet. A.* **140**, 1366–1374 (2006).

44. Hennekam, R. C. M., Krantz, I. D. & Allanson, J. E. *Gorlin's Syndromes of the Head and Neck*. (Oxford University Press, Inc., 2010).

45. Attanasio, C. *et al.* Fine Tuning of Craniofacial Morphology by Distant-Acting Enhancers. *Science* **342**, 1241006 (2013).

46. 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).

47. Alappat, S., Zhang, Z. Y. & Chen, Y. P. Msx homeobox gene family and craniofacial development. *Cell Res.* **13**, 429–442 (2003).

48. Satokata, I. & Maas, R. Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat. Genet.* **6**, 348–356 (1994).

49. 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).

50. Wang, J.-L. *et al.* TGF-β signaling regulates DACT1 expression in intestinal epithelial cells. *Biomed. Pharmacother.* **97**, 864–869 (2018).

51. 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).

52. Stegman, M. A. *et al.* Identification of a tetrameric hedgehog signaling complex. *J. Biol. Chem.* **275**, 21809–21812 (2000).

53. 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).

54. 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).

Methods-only 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., Steen, K. van, 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-Verlag, 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**, D493-496 (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).

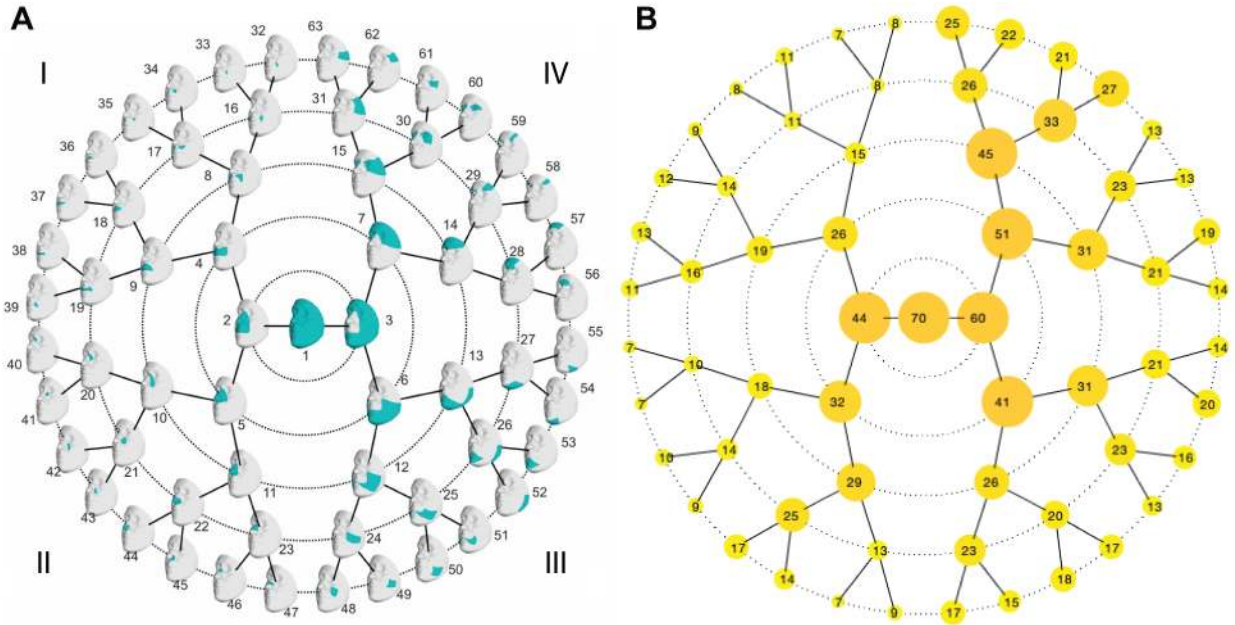
86. Baumgart, S. J. *et al.* CHD1 regulates cell fate determination by activation of differentiation-induced genes. *Nucleic Acids Res.* **45**, 7722–7735 (2017).

87. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease risk association. *Science* **366**, 1134–1139 (2019).

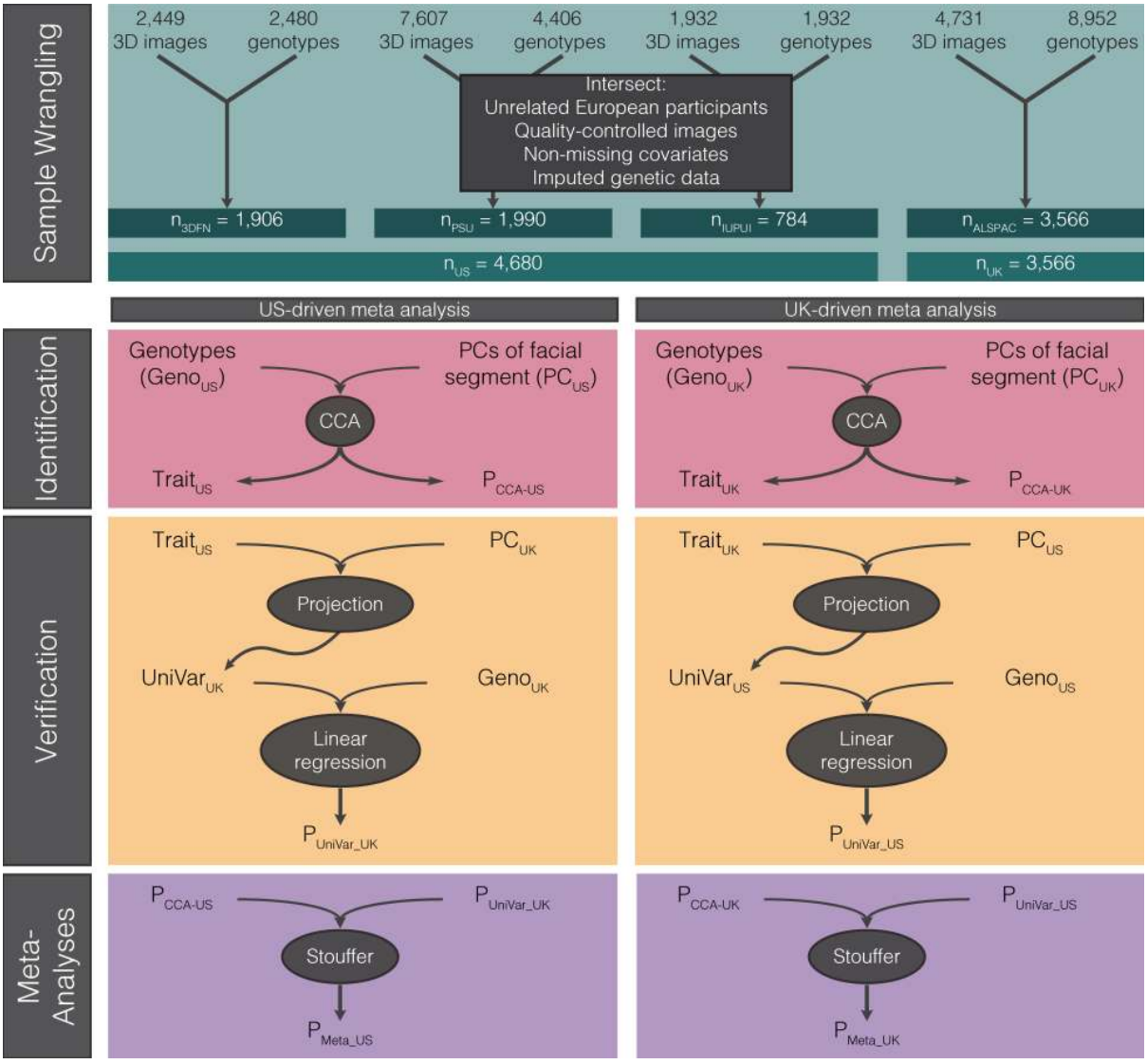
88. Pattison, J. M. *et al.* Retinoic Acid and BMP4 Cooperate with TP63 to alter Chromatin Dynamics during Surface Epithelial Commitment. *Nat. Genet.* **50**, 1658–1665 (2018).

89. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

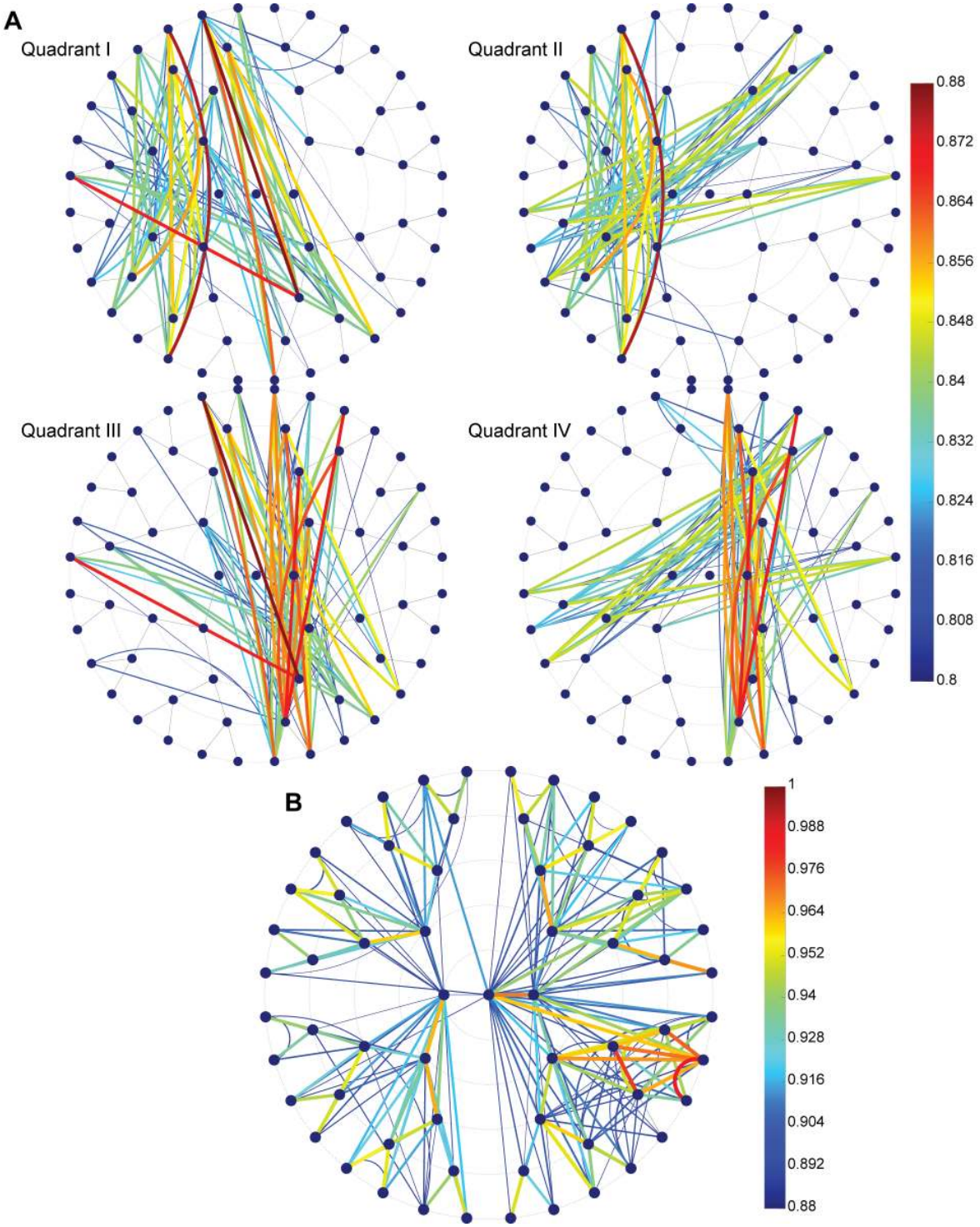
Extended Data



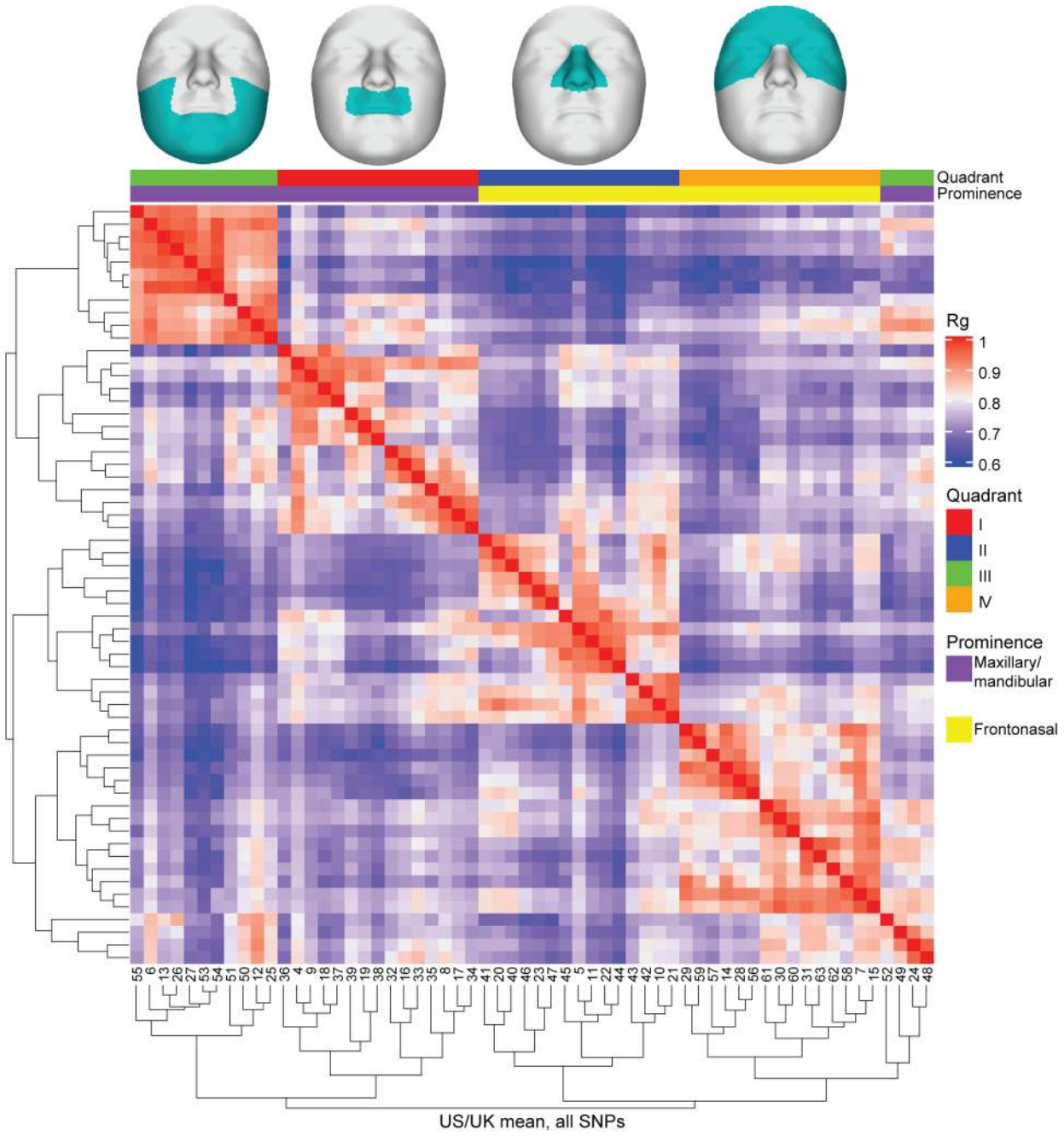
Extended Data Figure 1. (A) Global to local facial segmentation of all 3D images (n_{Total} = 8,246), obtained using hierarchical spectral clustering. Segments are colored in teal and identical to those in Figure 1. Roman numerals represent "quadrants" of facial segments. (B) The number of principal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



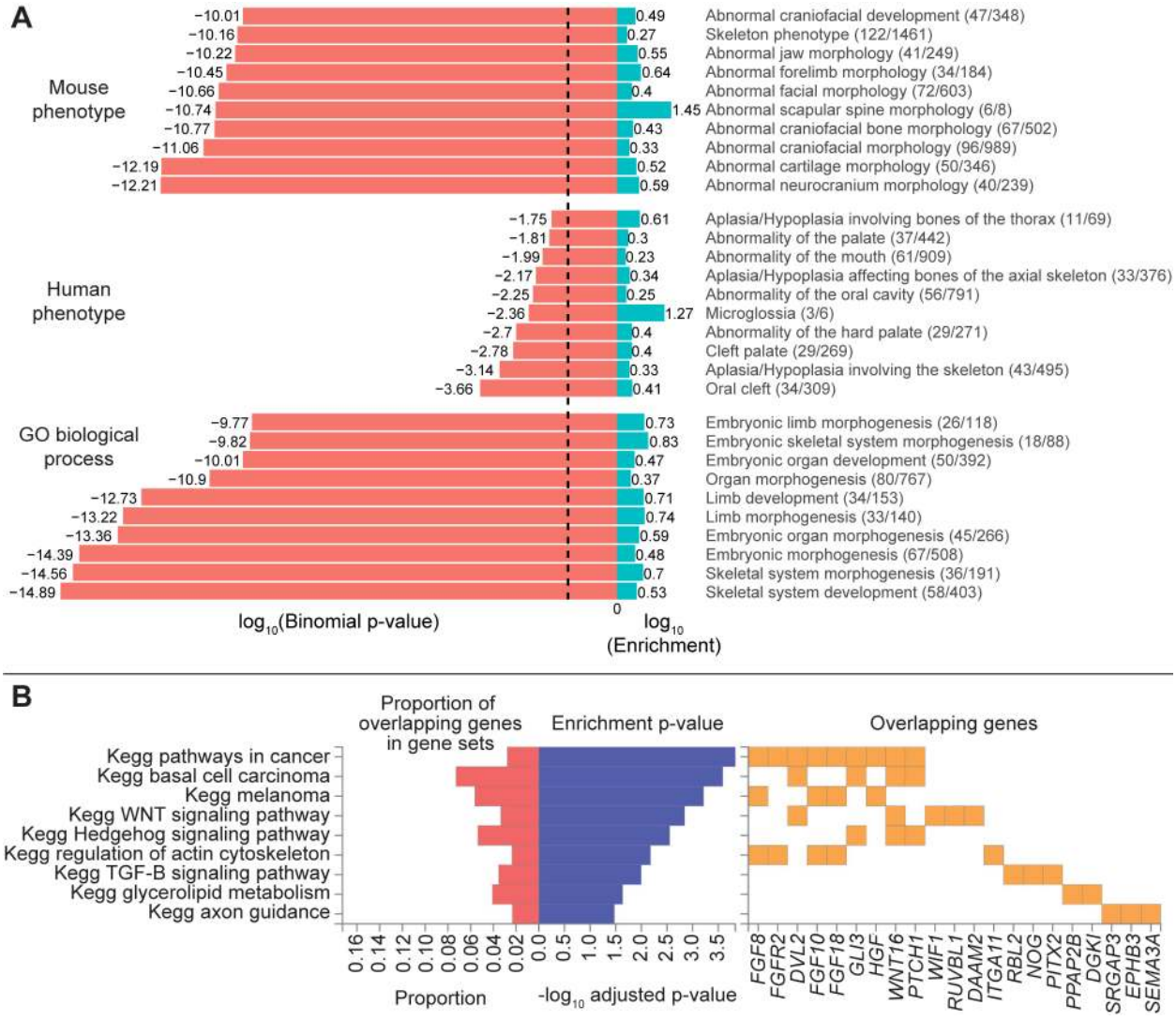
Extended Data Figure 2. *Sample Wrangling*: Images and genotypes from each study were intersected and unrelated participants of European ancestry, with quality-controlled images, covariates, and imputed genetic data were selected to obtain the analyzed data. *Identification*: For each facial segment, 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 $P_{META\text{-}UK}$).



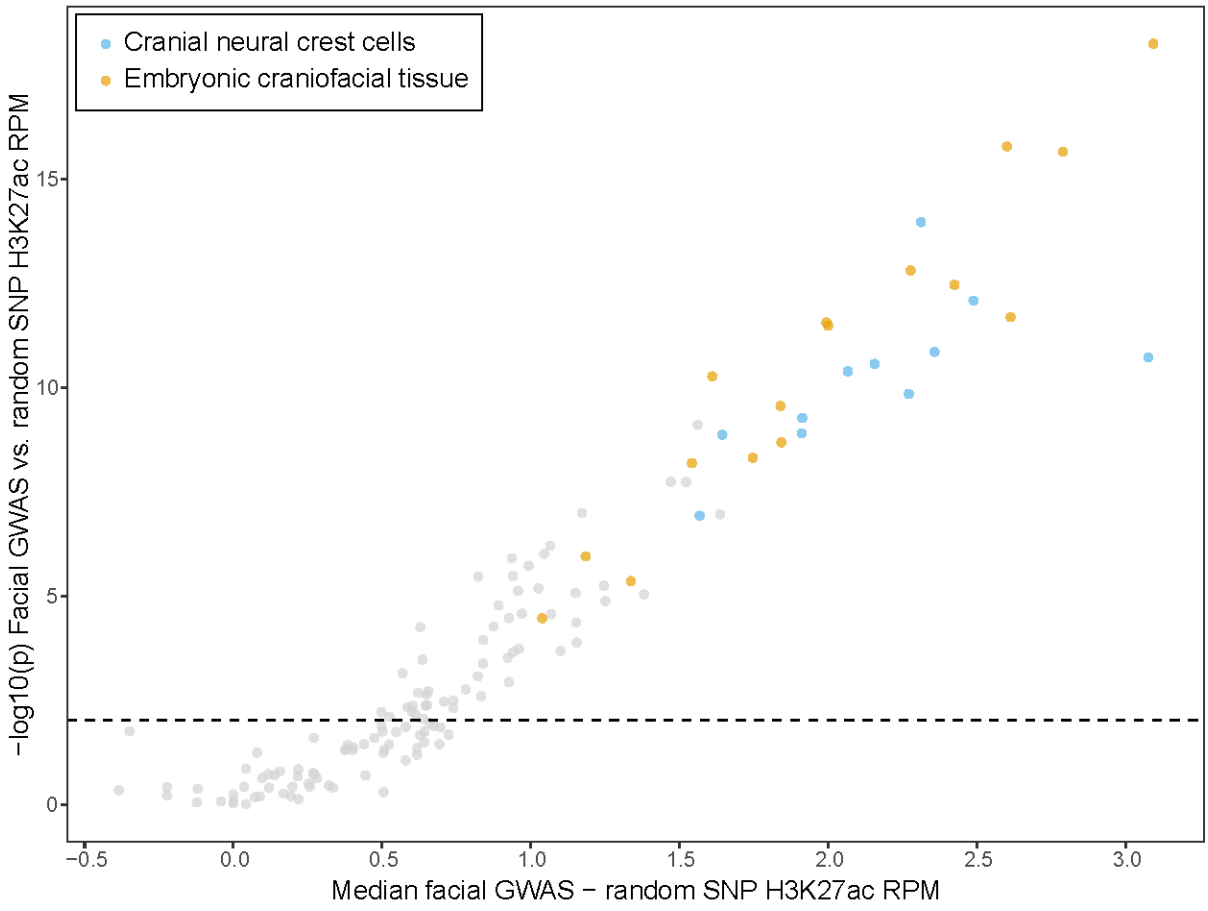
Extended Data Figure 3. LDSC correlations between segments. (A) Correlations between segments from different quadrants, ranging from 0.8 to 0.88, which seem to reflect both physical proximity of segments on the face and shared embryological origins. (B) Correlations ranging from 0.88 to 1, which are mostly between segments within the same facial quadrant.



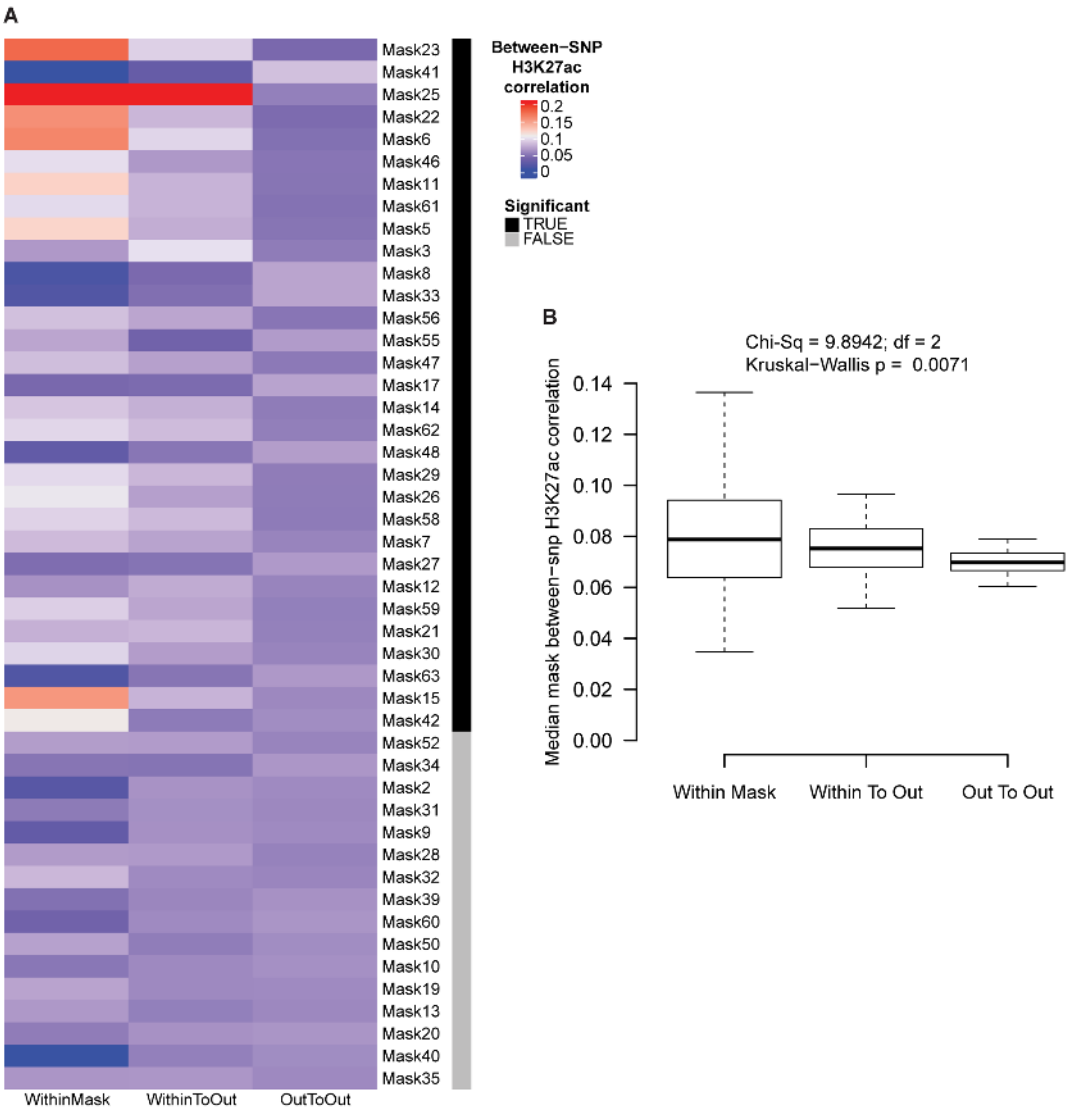
Extended Data Figure 4. Correlations between facial segments on the basis of SNP *P* values were calculated using LDSC, as described in Methods, and average linkage hierarchical clustering was performed using the matrix of correlation values. Quadrant colors in legend refer to the quadrant of the polar 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.



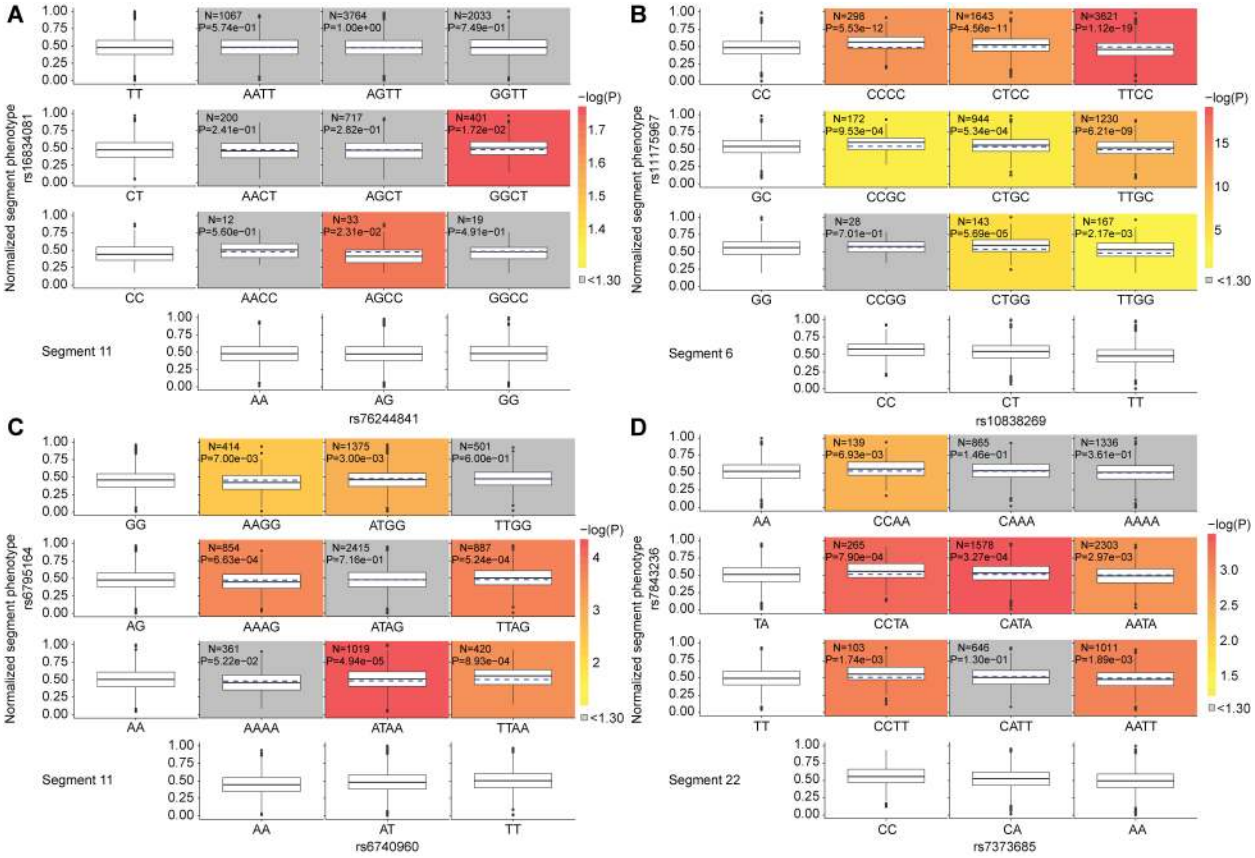
Extended Data Figure 5. (A) GREAT analysis. For the top ten GO terms in each category, plotted is the binomial test Bonferroni-corrected *P* value (red; negative values) and binomial region fold enrichment (blue; positive values). Behind every GO term, in parentheses we indicate the number of genes in 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* = log10(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.



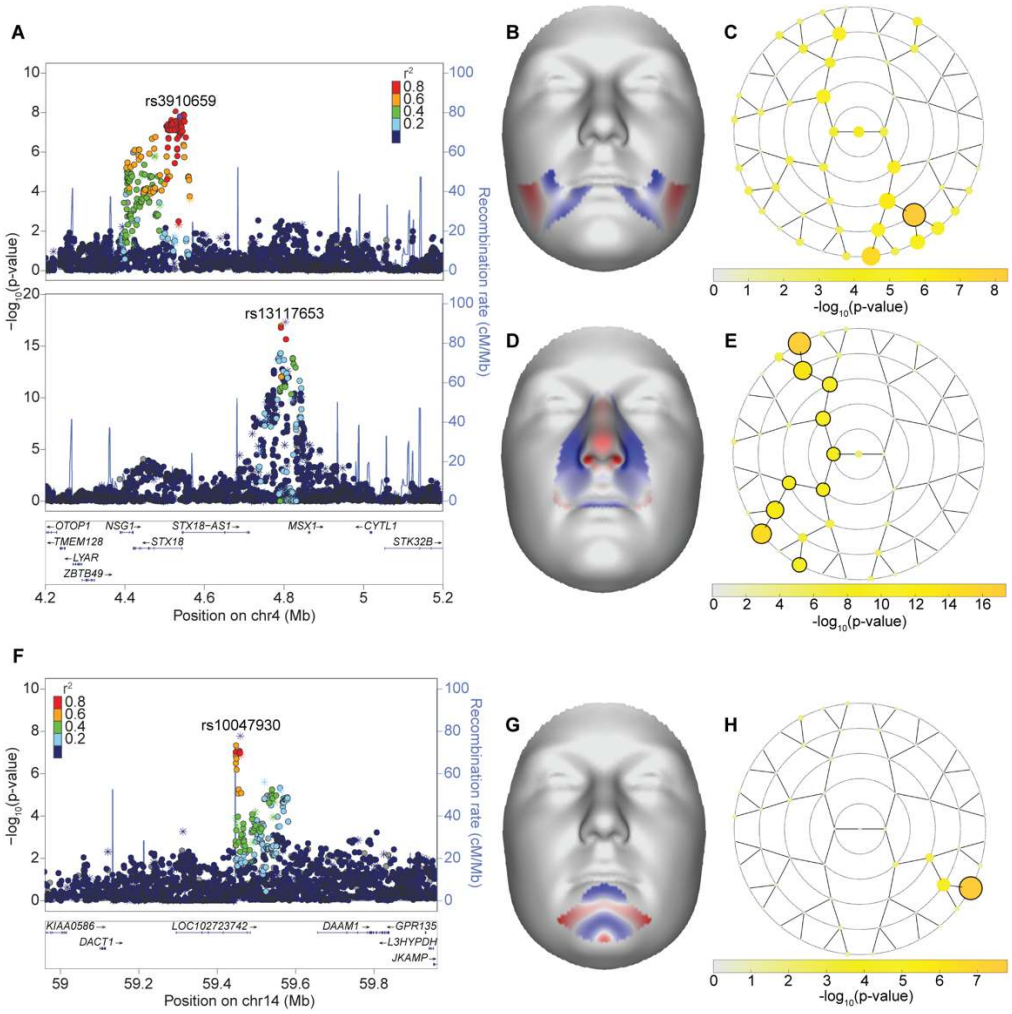
Extended Data Figure 6. For all cell-types and tissues, each represented by a point above, the median difference between H3K27ac RPM signal between the 203 lead SNPs vs. 203 random SNPs was tested for significance using a two-sided Wilcoxon rank-sum test. The thin dashed line represents the 5% false 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.



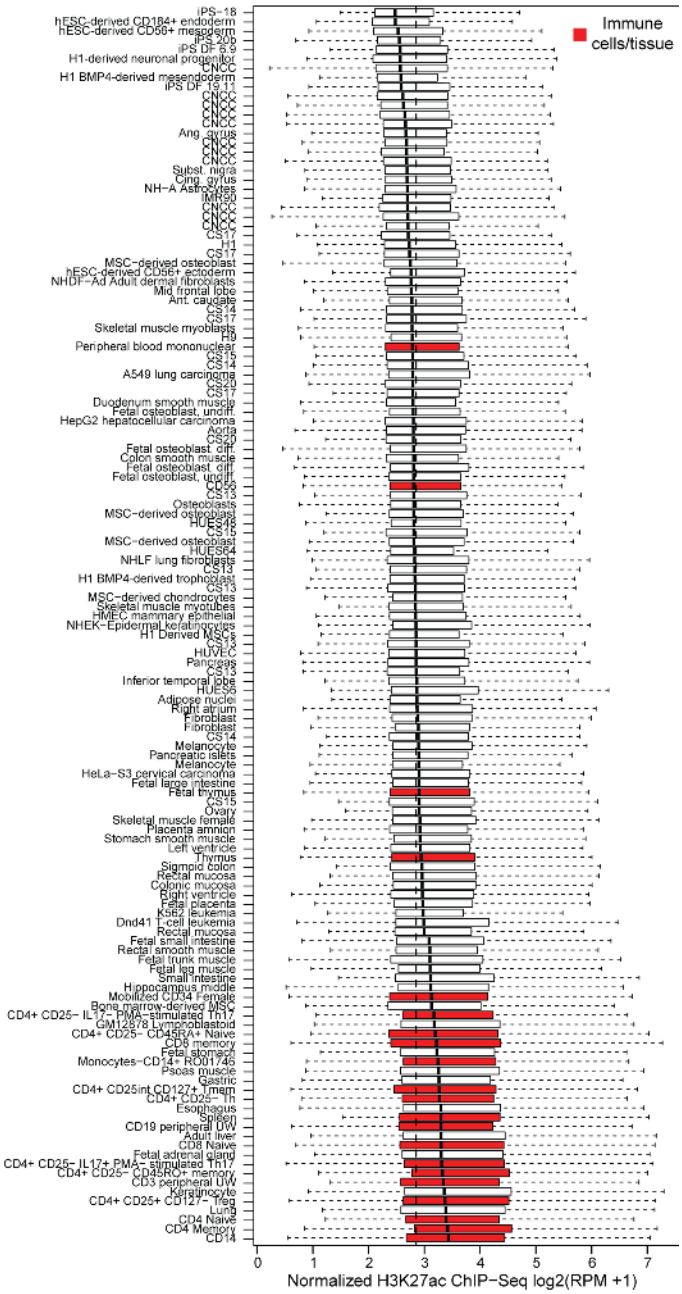
Extended Data Figure 7. (A) For all segments (aka "masks"), we compared the H3K27ac activity for significant SNPs from the refined SEM model for variation in that facial segment. Plotted is the Spearman's rho correlation between pairs of SNPs significant in the same SEM model ("Within Mask"); pairs of 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 Figure 8. 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 Figure 9. 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 -log10(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 Figure 10. 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.

Supplementary Materials for

Insights into the genetic architecture of the human face

Julie D. White[1†*], Karlijne Indencleef[2,3†*], 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], Peter Claes[2,3,7,18*]

*Correspondence to: jdw345@psu.edu; karlijne.indencleef@kuleuven.be; peter.claes@kuleuven.be

†These authors contributed equally to this work.

This PDF file includes:

Supplementary Note 1

Canonical correlation analysis to analyze multivariate data.

In CCA, 'canonical' is the statistical term for analyzing latent variables (which are not directly observed), obtained using a singular value decomposition on a correlation matrix, that represent multiple variables (which are directly observed), and the primary outputs are canonical correlation values. In our work, we use CCA to evaluate the latent associations between genotype and principal components of shape variation. The identification of "traits," or the linear combination of principal components most associated with genotypes for the SNP being tested, is a by-product of the test which we have opportunistically used. CCA has also been implemented in the PLINK genetic analysis software as `mv-PLINK` (https://genepi.qimr.edu.au/staff/manuelF/multivariate/main.html)[1].

Validity of the meta-analysis

To illustrate the validity of our meta-analysis pipeline, we performed the following simulation:

1. For each of the 203 lead SNPs, randomly rearrange the genotypes in both datasets, essentially creating genotypes that have no underlying association with multivariate facial shape in any of the 63 facial segments.
2. Identification step in the meta-analysis: For the US dataset, use canonical correlation analysis and Rao's F-test approximation to identify the linear combination of principal components representing shape variation in segment 1 that is maximally correlated with the SNP, producing a *P* value and a latent facial trait. Repeat this for all 63 facial segments, separately, ultimately generating 63 latent facial traits and 63 associated *P* values.
3. Verification step in the meta-analysis: In segment 1, project the principal components of the UK individuals onto the latent facial trait from Step 2 to obtain univariate projection scores for the UK individuals on the US-derived latent trait. Subject these scores to a standard linear regression against SNP variation to verify or replicate the SNP-trait association. Separately repeat for all 63 latent traits.
4. Meta-analysis step: Combine the independent *P* values from the US-based identification step and the UK-based verification step using Stouffer's method. Do this for each of the 63 segments.
5. Reverse the identification and verification datasets: Reverse the roles of the US and UK datasets and repeat Steps 2 - 4, this time with the UK as identification and the US as verification, to obtain a second meta-analysis *P* value. Note that this meta-analysis outcome is not to be combined with the first meta-analysis outcome, because the two are not independent. Do this for each of the 63 facial segments.
6. Repeat steps 1-5 for a total of 10,000 permutation cycles for each of the 203 lead SNPs.

Under the null hypothesis for random genotypes, we expect to see a uniform distribution [0,1] of *P* values for each of the three statistical stages: identification, verification, and meta-analysis combination. If any stage of the analysis were artificially inflating the *P* values through non-independence of the results, we would expect severe deviations from this distribution even with noisy simulated data. In Figure SN1-1 below, we show with QQ plots that we maintain a uniform distribution at all analysis stages for random genotypes, indicating that the overall approach of identification > verification > meta-analysis is not causing spurious results. In performing this simulation for the 203 lead SNPs, illustrated below using two SNPs with extreme

MAF values, we demonstrate that our approach is well-calibrated per SNP, and it is likely to be so genome-wide. The true (non-permuted) QQ plots for each segment are available in the FigShare repository for this manuscript.

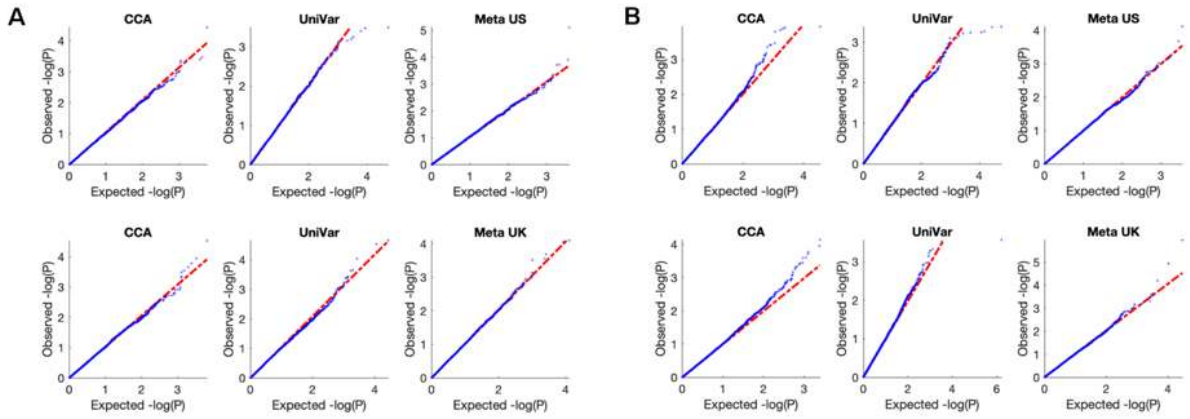


Figure SN1-1. QQ plots of the permutation *P* values. For two example lead SNPs tested in segment 1: (A) rs76244841 (MAF = 0.42) and (B) rs16834081 (MAF = 0.09), this figure contains the QQ plots from the entire meta-analysis pipeline after randomly permuting the SNPs 10,000 times. Each QQ plot represents the distribution of *P* values at that stage of the analysis, with the CCA plots being the identification stage of the analysis (one-sided right-tailed Rao's F test), the 'UniVar' plots being the verification stage (one-sided right-tailed linear regression), and the 'META US' and 'META UK' columns being the meta-analysis stage, following the pipeline described in the text and Extended Data Fig. 2. As the purpose of this permutation was to illustrate the distribution of *P* values expected for permuted results, significance of these *P* values was not assessed.

We can also illustrate via simulation that the combination of *P* values via Stouffer's meta-analysis is also viable, and that the two *P* values from the identification step and the verification step are independent. To do so, we have simulated the following three scenarios, and show the results in Figure SN1-2:

A. One randomly generated vector of 10^6 *P* values, illustrating a uniform distribution of *P* values expected under noisy conditions that reflect the null hypothesis of chance associations.
B. The randomly generated vector of *P* values in A was used to combine each *P* value with itself using Stouffer's method. This scenario creates false-positives and illustrates dependent *P* values and their effect on the Stouffer meta-analysis.
C. The randomly generated vector of 10^6 *P* values in A was used to combine each *P* value with another randomly chosen (and thus independent) *P* value using Stouffer's method. This scenario illustrates the combination of independent *P* values in a Stouffer meta-analysis.

From these results (Figure SN1-2), one can appreciate the difference in QQ plots using dependent *P* values (B, the 'incorrect' scenario) and independent *P* values (C, the 'correct' scenario) in a Stouffer *P* value combination. One can visually compare the QQ plots presented

in Figure SN1-2 with those presented above in Figure SN1-1 and appreciate that those presented in Figure SN1-1 are most similar to the QQ plots illustrating independent *P* values, in which the blue line follows the dotted red line more closely for a long range of $-\log_{10}(P)$ values.

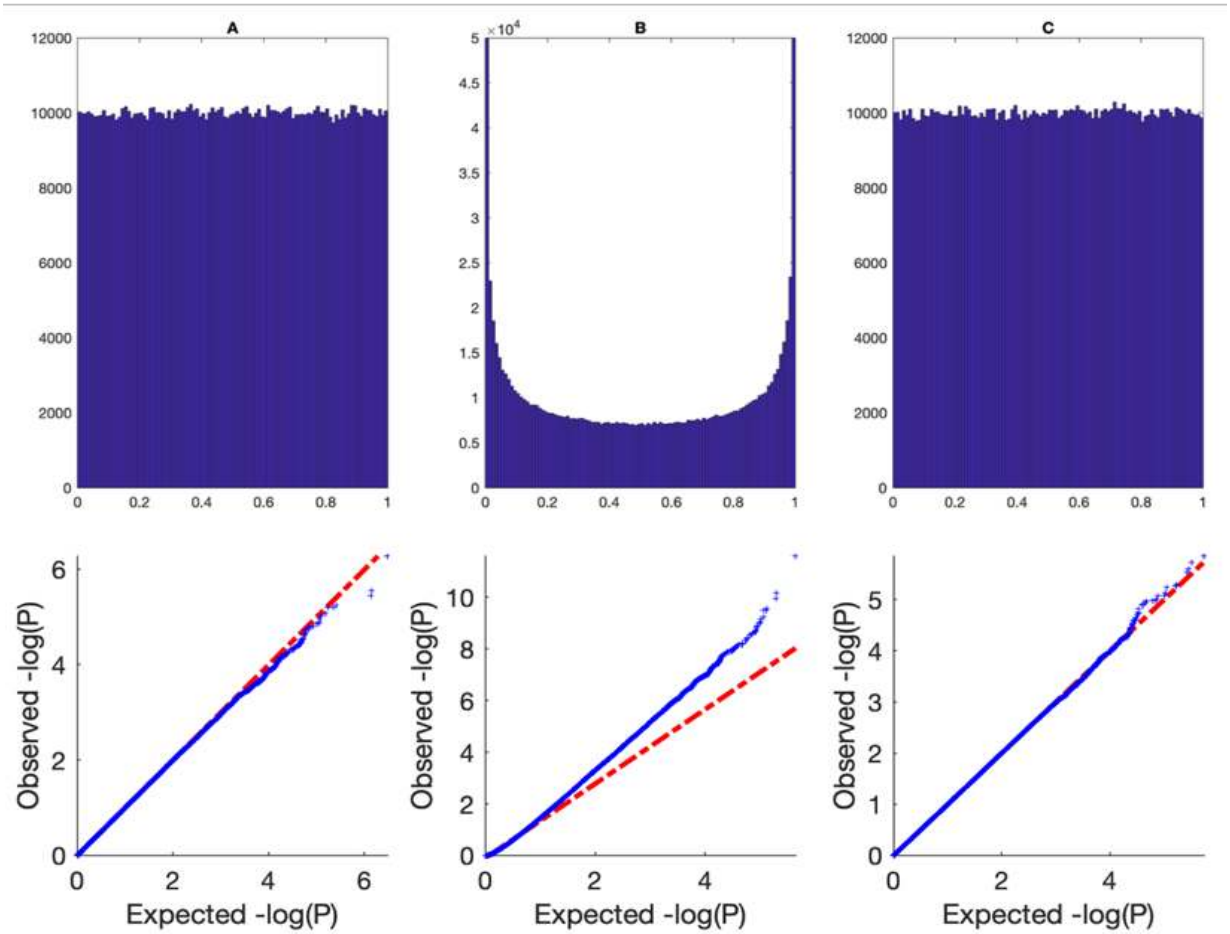


Figure SN1-2. Simulated *P* values under enforced independence and non-independence. Histogram distribution of *P* values (*top*) and QQ plots (*bottom*) from (A) a random vector of *P* values, (B) a completely dependent meta-analysis, and (C) a completely independent meta-analysis. The QQ-plots in scenarios A and C are similar to our permutation QQ plots, shown in Figure SN1-1, demonstrating that when we combine the one-sided, right-tailed identification (CCA) *P* values and verification (UniVar) *P* values using Stouffer's method, we are combining independent *P* values.

Supplementary Note 2

Empirical calculation of study-wide significance threshold

To calculate the study-wide significance threshold, we performed an empirical calculation of the number of independent tests performed in our meta-analysis framework, following the procedure in Kanai et al. (2016)[2]. Using the permutation framework described in Supplementary Note 1, we calculated the number of independent tests performed for each of the 203 lead SNPs as well as 1,000 randomly selected SNPs from throughout the genome. This resulted in a total of 10,000 permutations × 63 segments × 2 meta-analysis tracks = 1,260,000 *P* values and latent traits for each SNP. In the FigShare repository for this manuscript[3], we have provided the permutation outcomes for both the 203 lead SNPs (Permutation203LeadSNPs.mat) and the 1,000 randomly selected SNPs (Permutation1000RandomSNPs.mat) alongside a MatLab 2017b script (IndependentTests.m) in which the empirical calculations of the number of independent tests, as discussed below and in Table SN2-1, can be replicated. Confirming expectations, permutation results based on the 203 lead SNPs or 1,000 randomly selected SNPs are close to the same, with a slightly lower need to correct for multiple testing for the former (study-wide *P* value = 7.09×10^{-10}) compared to the latter (study-wide *P* value = 6.96×10^{-10}). Therefore, we opted to incorporate the latter, more strict, threshold as the study-wide *P* value threshold.

In Calculation A, we determined the number of independent CCA tests per segment, SNP, and dataset used as identification (Step 2 from Supplementary Note 1). For each segment, in each dataset, and for each SNP, we divided 0.05 by the 5th percentile of the 10,000 permuted CCA *P* values. For the 1,000 random SNPs, this was done 126,000 times (1000 SNPs × 63 segments × 2 datasets) and the mean number of tests performed was 1.0094 ± 0.0485 (median = 1.0070). For the 203 lead SNPs, this was done 25,578 times (203 SNPs × 63 segments × 2 datasets) and the mean number of tests performed was 1.0052 ± 0.0445 (median = 1.0036). Here, we can observe that performing the CCA for one SNP in one segment is practically equivalent to a single test.

In Calculation B, we determined the number of tests performed using our meta-analysis pipeline per SNP and per segment, using one dataset for identification (Step 2), followed by the other for verification (Step 3), then meta-analyzing the *P* values (Step 4). For each segment, using both datasets, and for each SNP, we divided 0.05 by the 5th percentile of the 10,000 permuted meta-analysis *P* values. For the 1,000 random SNPs, this was done 126,000 times (1,000 SNPs × 63 segments × 2 meta-analysis tracks). The mean number of tests was 1.0058 ± 0.0448 (median = 1.0042). For the 203 lead SNPs, this was done 25,578 times (203 SNPs × 63 segments × 2 meta-analysis tracks) and the mean number of tests was 1.0043 ± 0.0435 (median = 1.0028). Here, similar to Calculation A, we can again observe that performing the meta-analysis pipeline for one SNP in one segment is practically equivalent to a single test.

In Calculation C, we determined the number of independent meta-analysis tracks performed per segment and SNP, after using each dataset in turn for identification and verification (Step 5). To do this, we retained for each of the 10,000 permutation cycles the

minimum (i.e. best) meta-analysis *P* value from either the US-driven meta-analysis or the UK-driven meta-analysis. Subsequently, we divided 0.05 by the 5th percentile of the 10,000 best *P* values. For the 1,000 random SNPs, this was done 63,000 times (1,000 SNPs × 63 segments) and the mean number of tests was 1.6723 ± 0.1145 (median = 1.6631). For the 203 lead SNPs, this was done 12,789 times (203 lead SNPs × 63 segments), and the mean was 1.6679 ± 0.1087 (median = 1.6600). According to this empirical estimation, a factor of 1.66 should be included in the calculation of the study-wide threshold to account for the two meta-analysis tracks. Given that the two meta-analysis tracks use the same two datasets, just with their roles reversed, full independence, which would have required a correction factor of 2, was not expected.

In Calculation D, we determined the number of tests per SNP in all 63 multivariate facial segments, using one dataset for identification (Step 2), followed by the other dataset for verification (Step 3), and a meta-analysis of the *P* values (Step 4). This time, for each of the 10,000 permutation cycles we retained the minimum (i.e. best) meta-analysis *P* value across the 63 segments for each SNP from both the US-driven and UK-driven meta-analyses. Subsequently, we divided 0.05 by the 5th percentile of the 10,000 best *P* values. For the 1,000 random SNPs, this was done 2,000 times (1,000 SNPs × 2 meta-analysis tracks). The mean number of tests was 43.0933 ± 2.3187 (median = 43.0145). For the 203 lead SNPs, this was done 406 times (203 SNPs × 2 meta-analysis tracks), and the mean number of tests was 42.4920 ± 2.0819 (median = 42.3652). This illustrates that the 63 multivariate facial segments are not independent, as expected, so multiple-testing correction by a factor of 63 would be too conservative.

Based upon the empirical tests performed here (summarized in Table SN2-1), the number of tests per SNP across 63 segments and using our two meta-analysis tracks, and thus the appropriate study-wide correction factor, was estimated by multiplying the outcomes of Calculations B, C, and D. Using the median estimators from 1,000 random SNPs, the study-wide correction factor becomes 1.0042 × 1.6631 × 43.0145 = 71.8379. Therefore, the study-wide significance threshold for this work is $\frac{5\ x\ 10^{-8}}{71.8379} = 6.96\ x\ 10^{-10}$.

Table SN2-1. The number of independent tests according to empirical calculations, estimated from permutations of the 203 lead SNPs and 1,000 randomly selected SNPs.

	203 Lead SNPs			**1000 Random SNPs**		
	Mean	***Std***	***Median***	***Mean***	***Std***	***Median***

Calculation A	1.0052	0.0445	1.0036	1.0094	0.0485	1.0070
Calculation B	1.0043	0.0435	1.0028	1.0058	0.0448	1.0042
Calculation C	1.6679	0.1087	1.6600	1.6723	0.1145	1.6631
Calculation D	42.4920	2.0819	42.3652	43.0933	2.3187	43.0145
K = Calculation B × Calculation C × Calculation D	71.1772		70.5231	72.4829		71.8379
Study-wide threshold = 5×10^{-8} / K	7.02×10^{-10}		7.09×10^{-10}	6.90×10^{-10}		6.96×10^{-10}

194
195

Supplementary Note 3

Literature evidence of epistatic interactions (Table 1)

Follow up data mining on the four epistatic SNP pairs was performed using VarElect[4], StringDB[5], and Encode[6,7], with the results detailed here and in the Discussion section of the manuscript.

The strongest SNP × SNP epistatic interaction was found between rs10838269 and rs11175967 (stat = 23.9422; $P = 9.94 \times 10^{-7}$) within segment 6, which covers the area of the face from the zygoma to the mandible. Rs11175967 is an intronic variant mapped to the *HMGA2* transcription factor (12q14.3), which has been associated with Silver-Russell syndrome, symptoms of which include a triangular face shape and broad foreheads[8,9]. Rs10838269, its epistatic partner, is an intergenic variant whose nearest protein coding gene is the transcription factor *ALX4* (11p11.2), which is expressed in the mesenchyme of developing bones and has been shown to play a vital role in craniofacial development[10]. Previous morphology studies identifying both *HMGA2* and *ALX4* show that the genes respectively contribute to ear morphology and stature in sheep[11]. In addition, genomic analyses on finches have shown that alterations of *ALX1*, the protein product of which is functionally redundant to Alx4 in mice[12], and *HMGA2* have been associated with beak shape[13] and size[14].

Rs76244841 (candidate gene *PRDM16*) and rs62443772 (candidate gene *GLI3* associated) were found to have a significant interaction in facial segment 9 (stat = 16.5745; $P = 4.68 \times 10^{-6}$), which covers the premaxillary soft tissue from the base of the columella to the oral commissure. *PRDM16* encodes a zinc finger transcription factor[15,16] and has been shown to affect palatal shelf elevation through repression of TGFβ signaling[17,18]. *GLI3* encodes a transcriptional activator and a repressor of the sonic hedgehog (Shh) pathway, which has been shown to play a role in limb development[19–21]. In addition, there is evidence that mouse null *Gli3* mutants have a broad nose phenotype[22] and genome-wide scans have previously implicated *GLI3* in nose morphology[23]. The connection between *PRDM16* and *GLI3* can be seen by their interaction through the *SUFU* intermediary. Multiple studies conducted on *Drosophila melanogaster* have identified evidence for a tetrameric Hedgehog signaling complex comprising Fu, Ci (an ortholog of PRDM16), Cos2, and Su(fu) (an ortholog of SUFU), including evidence that Su(fu) binds directly to Ci[24–26]. SUFU has also been shown to mediate the phosphorylation of GLI3 via GSK3[27] and has also been shown to interact with the GLI1-3 zinc-finger DNA-binding proteins[28,29].

Rs6740960 (candidate gene *PKDCC*) and rs6795164 (candidate gene *SLCO2A1*) (stat = 16.3707; $P = 5.21 \times 10^{-5}$), and rs7373685 (candidate gene *GATA2*) and rs7843236 (candidate gene *SNTB1*) (stat = 15.7837; $P = 7.10 \times 10^{-5}$) were significant pairs in facial segments 11 and 22, respectively, which are hierarchical segments that include areas surrounding the base of the nose. Due to the overlapping nature of the segments, these variants were analyzed as a collective group. The nature of the relationship between these four variants is less clear, however some trends are evident. The first is that there appears to be a connection between

GATA2 and SLCO2A1 through AKT1. AKT1 is one of 3 related serine/threonine-protein kinases first characterized in mouse models[30], which regulate multiple processes, such as metabolism, cell survival, growth, proliferation, and angiogenesis. It has also been implicated in Proteus Syndrome, whose symptoms include bone development abnormalities[31,32]. SLCO2A1 is a solute carrier involved in the release and transport of prostaglandin[33,34] and has also been shown to be involved in hypertrophic osteoarthropathy[35–37]. SLCO2A1 regulates AKT1 and the Akt pathway through prostaglandin[38,39]. Furthermore, the PI3K/Akt signal pathway has been shown to negatively regulate the transcriptional activator GATA2[40]. There were not any connections found between PKDCC and SNTB1, however, there was an interesting connection between SNTB1 and GATA2 via Dystrophin (DMD). DMD serves as a key component of the dystrophin-associated glycoprotein complex, which helps stabilize the sarcolemma[41]. SNTB1 is an adapter protein that has been suggested to link receptors to the dystrophin glycoprotein complex[42,43]. GATA2 has also been shown to be a transcriptional factor of DMD[44]. Finally, there is evidence in mouse models that supports a connection between the Akt signaling pathway and DMD[45], which serves as another underlying link between three of the four epistatic hits. While there were no evident links between PKDCC and the other epistatic hits, it may be worth noting that this tyrosine-protein kinase has been previously shown to be involved in bone growth[46–48].

Supplementary Methods

Sample and recruitment

For the 3DFN sample, 3D images and genotype data were obtained from the 3D Facial Norms repository[49]. 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.

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.

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 unrelated individuals of European descent 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[50,51]. 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 unrelated participants of European ancestry with quality-controlled images, covariates, and genotype data included 3,566 individuals. 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).

Genotyping and imputation

Genotyping of the 3DFN sample was performed at the Center for Inherited Disease Research at Johns Hopkins University. Participants and control samples were genotyped on the Illumina OmniExpress + Exome v1.2 array, plus 4,322 investigator-chosen SNPs included to capture regions of interest. PSU participants were genotyped by 23andMe on the v3 and v4 arrays (Mountain View, CA). Participants sampled at IUPUI were genotyped using Illumina's Infinium Multi-Ethnic Global-8 v1 array, performed by the University of Chicago's DNA Sequencing & Genotyping Facility (Chicago, IL). ALSPAC participants were genotyped using the Illumina Human Hap550 quad array, performed by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute (Cambridge, UK) and the Laboratory Corporation of America (Burlington, NC), supported by 23andMe.

For all datasets, samples were evaluated for concordance of genetic and reported sex (--check-sex), evidence of chromosomal aberrations, genotype call rate (--mind 0.1), and batch effects using PLINK 1.9[52]. SNPs were evaluated for call rate (--geno 0.1), Mendelian errors (--set-me-missing), deviation from Hardy-Weinberg genotype proportions (--hwe 0.01), and sex differences in allele frequency and heterozygosity, also using PLINK 1.9. The genotypes were "harmonized" with 1000 Genomes Project (1000G) Phase 3[53] using Genotype Harmonizer (v1.4.20)[54] with a window size of 200 SNPs, a minimum of 10 variants, and alignment based on minor allele frequency (--mafAlign 0.1). This program was also used to filter out ambiguous SNPs, update the SNP id, and update the reference allele as needed, all in reference to the 1000G Phase 3 genotypes. After genotype harmonization, the study datasets were merged (n = 44,383 SNPs in common) and explored using principal components analysis (PCA) to assure that there were no batch effects by genotyping platform. Relatedness across the entire US sample was also assessed using this intersection and the KING (v2.1.3) software[55]. Relatives were noted in the per-platform subsets, and the imputation process proceeded for the full number of quality-controlled SNPs from each platform.

Prior to phasing, special quality control steps were performed on each platform. First, the allele frequency of each SNP was compared to the allele frequency of that SNP in the 1000G Phase 3 dataset. SNPs were removed if the allele frequency in the study dataset was not within |0.2| of any one of the 1000G super populations. We also removed SNPs with duplicate

positions, any remaining insertions/deletions, copy number variants, and haploid genotypes. Individuals were removed if they had heterozygosity values ±3 standard deviations from the mean. Haplotypes were estimated using SHAPEIT2 (v2.r900)[56]. The samples were then imputed to the 1000G Phase 3 reference panel using the Sanger Imputation Server (v0.0.6)[57] with the positional Burrows-Wheeler Transform (PBWT) pipeline (v.3.1)[58], resulting in nearly 40 million variants for each dataset. SNP-level (INFO score >0.8) and genotype-per-participant-level (genotype probability >0.9) filters were used to omit poorly imputed variants. The datasets were then merged and filtered by SNP missingness (--geno 0.5), minor allele frequency (--maf 0.01), and Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$) to produce a single merged dataset of all US participants with 7,417,619 SNPs for analysis.

The raw genotype data from ALSPAC was not available and restrictions are in place against merging the ALSPAC genotypes with any other genotypes. For this reason, imputed ALSPAC genotypes were obtained directly from the ALSPAC database and held separately during the analysis. Prior to phasing and imputation, the ALSPAC genotypes were subjected to standard quality control methods. Individuals 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). Only individuals of European descent, compared to the HapMap II dataset by way of multidimensional scaling analysis, were kept for imputation. SNPs were removed if they had a minor allele frequency of <1%, a call rate of <95%, or evidence for violations of Hardy-Weinberg equilibrium ($P < 5 \times 10^{-7}$). Related individuals were excluded based on pedigree information provided by ALSPAC and double-checked using the IBD command in PLINK 1.9.

Ancestry axes and selection of European participants

From the post-imputation merged dataset of the US participants, we identified the European participants by projecting them into a principal component (PC) space constructed using the 1000G Phase 3 dataset. To do this, we first excluded all indels, multi-allelic SNPs, and SNPs with MAF ≤ 0.1 in both the 1000G dataset and the US dataset and identified the SNPs common to both datasets. On this list (n = 1,940,221 SNPs), we iteratively performed linkage disequilibrium pruning (50 bp window, 5 bp step size, 0.2 correlation threshold) on the 1000G dataset until no variants were excluded. We then used this LD-pruned list (n = 461,372 SNPs) in a principal component analysis to construct a population structure space based upon the 1000G project and projected the US dataset onto that space to obtain the ancestry axes of our dataset.

3D image registration and quality control

After cleaning each image to remove hair, ears, clothing, and other imaging artifacts, five positioning landmarks are roughly indicated to establish image orientation. The MeshMonk registration framework (v0.0.6) is then used to map a symmetric (relative to the sagittal plane) anthropometric mask of 7,160 landmarks onto the images and their reflections, constructed by changing the sign of the x coordinate[59]. Outlier images, likely caused by image mapping errors, were identified using two approaches. First, as described in prior work[60,61], 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. Although variation in asymmetric facial features is of interest, in this work we sought to only study variation in symmetric facial shape. Therefore, when discussing facial shape, we always refer to the symmetric quasi-landmark configuration, which is calculated as the average of the original and reflected quasi-landmark configurations after aligning the two using Generalized Procrustes analysis to eliminate differences in position, orientation, and centroid size[62].

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[63,64] was applied to the meta-analysis, after converting the meta *P* values to z-scores (Equation 1), 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, by design, CCA is a one-sided test with canonical correlations always in the positive range of [0 1]. As a result, any of the resulting genetic correlations reported here are restricted to be positive as well.

(Equation 1) $$z_i = \phi^{-1}\left(1 - \frac{p_i}{2}\right)$$

Based on LDSC v1.0.1 (https://github.com/bulik/ldsc), we first used the munge_sumstats.py function to convert *P* values into z-scores, restricting ourselves to SNPs that overlap with the HapMap3 project and excluding those within the major histocompatibility complex region (--merge-alleles w_hm3.noMHC.snplist), with a minimum of *n* = 4,155 individuals (--n-min 4155). As input to the munging step, an additional column SIGN was added coding 1 for each SNP, reflecting a positive direction of effect (--signed-sumstats SIGN,1). Subsequently, using ldsc.py the genetic correlation between pairs of facial segments were generated based on the precomputed reference LD scores and weights for a European population (--ref-ld-chr eur_w_ld_chr/ and --w-ld-chr eur_w_ld_chr/, respectively).

GWAS peak selection

Within a locus, Adjacent SNPs were considered suggestively similar and belonging to the same signal as the lead SNP if the slope of the adjacent SNP trait and the lead SNP trait had a regression *P* value lower than 0.2. On the other hand, a *P* value higher than 0.2 was

considered to indicate different traits by the two SNPs and led to a definition of a new lead SNP within the 1 Mb window.

The consistency of effect across datasets was determined with a regression of the CCA traits found in the identification stage for each dataset, with results considered sufficiently similar if they were below a Benjamini and Hochberg[65] false discovery rate of $P \leq 3.66 \times 10^{-2}$.

Gene annotation

Genes 500 kb up- and downstream of the genome-wide significant lead SNPs were identified using the Table Browser of the UCSC Genome Browser[66]. The most likely candidate gene per lead SNP was identified based on a three-step system. First, we investigated whether any gene in the window was previously associated with craniofacial development or morphology through normal-range facial association studies, genetic disorders with facial dysmorphology as a symptom, or animal models. This search was performed by using the HGNC gene name as the search term in the PubMed, OMIM[67], GWAS Catalog[68], and PubTator Central[69] repositories. The resulting literature was scanned for terms like “facial,” “craniofacial,” “craniofacial (dys)morphology,” and “facial development,” and we subsequently determined whether the relevant literature was a GWAS of normal-range facial morphology, a study focused on human craniofacial dysmorphology, or a study focused on animal craniofacial morphology. If we did not find the gene included in previous literature covering these topics, we checked whether the gene was a likely contributor to facial development based on the paper of Hooper and colleagues, who used transcriptome data from critical periods of mouse face formation to assess gene activity across facial development[70]. If both methods did not deliver a suitable candidate gene, the most likely candidate gene was selected based on the FUMA gene prioritization algorithm (v1.3.3)[71].

Cell-type-specific enhancer enrichment

Raw read (fastq) ChIP-seq files of signals of acetylation of histone H3 on lysine K27 (H3K27ac) from Prescott et al.[72] (GSE70751; CNCCs), Najafova et al.[73] (GSE82295; fetal osteoblast cell line, undifferentiated and differentiated), Baumgart et al.[74] (GSE89179; mesenchymal stem cell-derived osteoblasts), Nott et al.[75] (UCSC genome browser; various brain cell types), and Pattison et al.[76] (GSE119997; surface ectoderm) were downloaded and aligned to the GRCh37 human genome using bowtie2 (v2.3.4.2) with default parameters. Aligned read (tagAlign) files of H3K27ac ChIP-seq from Wilderman et al.[77] (GSE97752; embryonic craniofacial tissue), and the Roadmap Epigenomics Project[78] (https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/; various fetal and adult tissues and cell-types) were also downloaded. Sample type, ID, and URL for each cell type/tissue is available in Supplementary Table 4.

To compare H3K27ac signal in the vicinity of the genome-wide significant lead SNPs between cell-types in an unbiased manner, we divided the genome into 20 kb windows, and calculated H3K27ac reads per million (RPM) from each aligned read (bam or tagAlign) file in

each window using bedtools coverage (v2.27.1). We then performed quantile normalization (using the normalize.quantiles function from the preprocessCore package, v3.7) on the matrix of 154,613 windows × 133 ChIP-seq datasets. We then selected the windows containing each of the 203 genome-wide significant lead SNPs, 203 random SNPs matched for minor allele frequency and distance to the nearest gene using SNPsnap[79], or 619 Crohn's disease-associated SNPs from the NCBI-EBI GWAS catalog[68].

Structural Equation Modeling

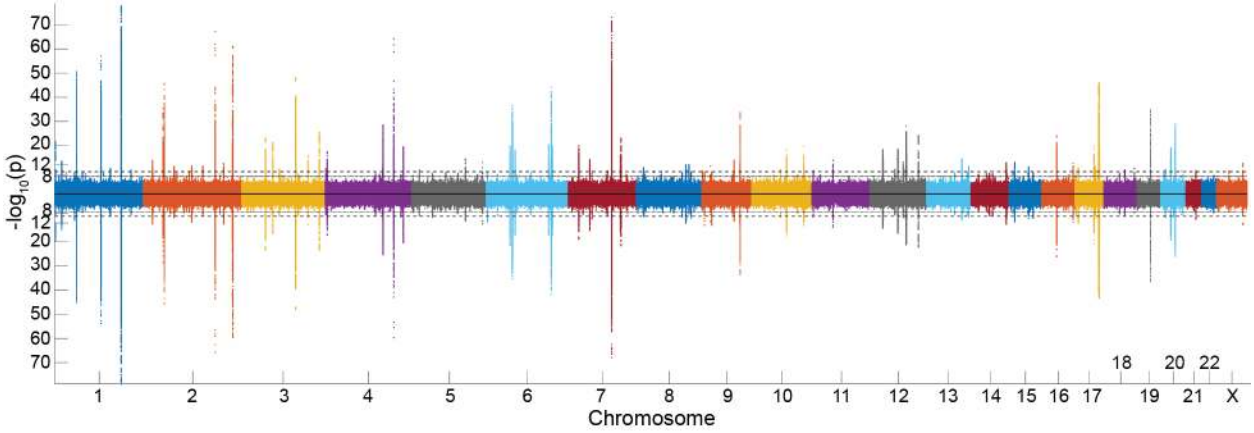
Mathematically, SEM analyses are a combination of a measurement model, which is constructed via confirmatory factor analysis, and a structural model, which is constructed using path analysis. In general, Lavaan outputs a best fit model that summarizes all genotype, phenotype, and covariate interactions, as well as a latent variable (aka "mask"), which is produced by a built-in dimension reduction that condenses the multidimensional facial phenotype from many principal components down to a single univariate phenotype. Parameters, which represent the interactions between the input variables, are generated by comparing the real covariance matrix between input variables and the estimated matrix created by numerical maximization, in our case carried out via maximum likelihood estimation. To maximize statistical power, Schreiber et al. recommend having at least 10 participants per parameter[80].

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 8,246 participants. Missing genotypes were substituted with the most common genotype based on frequency. Covariates of age, sex, height, weight, and face size (i.e. centroid size) were also included as model input. Prior to analysis, the distributions of these covariates were plotted and transformed, if necessary, to display near normal distributions. As genotypes are trichotomous, normality was not assessed.

Since analyzing all variants and all principal components for each segment via a single SEM would require the modeling of thousands of interactions and require extensive computational resources, separate SEM models were initially constructed. First, for each segment, we separated the 203 genome-wide significant lead SNPs into three groups and ran three SEM models on each of these groups, plus all covariates. If any of the three subset SEMs did not converge, we then re-grouped the SNPs into four or more groupings and re-ran the subset SEM models on these groupings. This process was repeated until all subset SEMs converged and we had parameter estimates for all 203 SNPs. Next, for each segment, SNPs with *P* values lower than 0.2 in the initial subset SEMs were collected and a unified SEM model for each segment was created and subsequently refined. If the unified SEM model did not converge, then this segment was discarded and no further analysis was performed. If all of the SNPs included in the unified model had *P* values lower than 0.2, a cutoff selected to maintain model stability, no further changes were made, and we reported the model fit indices and parameter estimates. For segments where the unified SEM model produced SNP *P* values greater than 0.2, the SNPs included in the SEM model were pruned by selecting SNPs with $P < 0.05$ and the model was re-run with this reduced set of SNPs. This process was repeated until

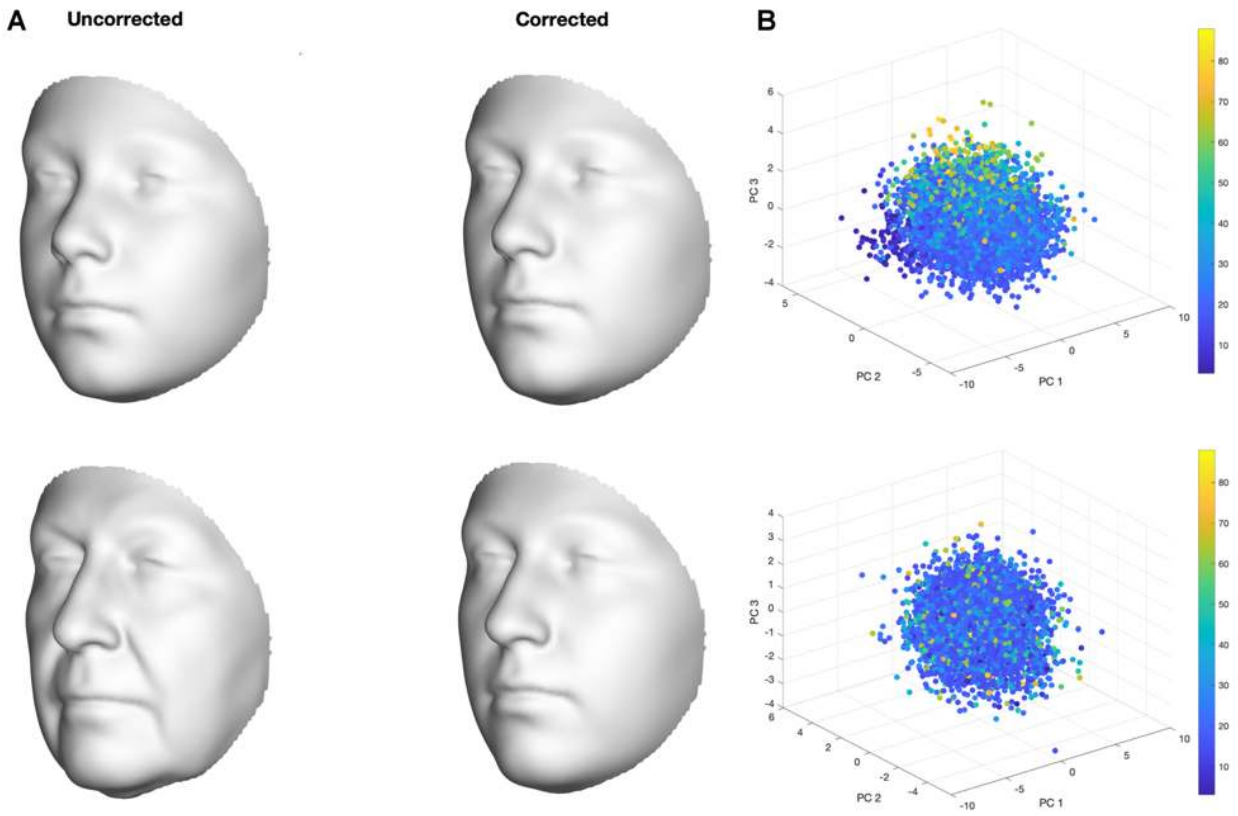
all SNPs had *P* values lower than 0.2. In the case of segments 7, 16, and 25, this iterative pruning process led to a rapidly declining model, so we elevated the SNP pruning *P* value from 0.05 to 0.1 to account for instability in these models. Once the model refinement was complete (i.e. all SNPs had $P < 0.2$), we designated the SNPs with $P < 0.05$ as significantly contributing to variance within the segment.

In general, the number of model parameters generated by the final refined SEM model for each segment ranged between 92 and 217, depending on the number of shape PCs and SNPs included in each model. As 8,246 participants were used, this led to a range of 38-90 participants per parameter, which is well above recommendations[80]. Additional statistical power was lent to our models by having a large number of samples and input variables per latent factor[81]. Of the 63 segments, the SEM models for 13 segments were discarded because they did not converge on a solution, which normally occurred when variants were non-informative for that particular segment or the variance of the segment was low. For each of the 50 SEM models where the refinement process was successful, we evaluated the fit of each model by instituting cutoffs on the following indices: Chi-square (*P* value < 0.05), comparative fit index (CFI > 0.90), root mean square error of approximation (RMSEA < 0.08), and standardized root mean square residual (SRMR < 0.08)[82,83], which generally indicate the strength of how well the SEM models the data. Eighteen models passed all recommended model fit parameters and 32 models passed all but one of the fit indices, leading to the conclusion that the refined SEM models fit our data well.



Supplementary Figure 1. Miami plot of all results.

Miami plot of all 63 facial segments combined, illustrating the chromosomal position of the associated loci from the meta-analysis track with the US dataset as identification (*top*) and UK as identification (*bottom*). Values plotted are the result of Stouffer's meta-analysis of one-sided right-tailed identification and verification *P* values, detailed in the Methods. The dotted horizontal line represents the genome-wide significance threshold ($P = 5 \times 10^{-8}$) and the dashed horizontal line represents the study-wide threshold ($P = 6.69 \times 10^{-10}$), calculated in Supplementary Note 2.



Supplementary Figure 2. Illustration of age correction using PLSR. (A) Average faces of 10 individuals with an age range of 9-10 in the top row, and 70-75 in the bottom row. The right column shows the same averages after our correction using PLSR. **(B)** The top plot shows the first three principal components, calculated from segment 1, before PLSR correction and color-coded according to age. In this plot, the images seem to separate along the axes of PC1 and PC3 based on age. The bottom plot shows the first three principal components after PLSR correction, also color-coded according to age. In this plot, the separation along the PCs is no longer according to age, indicating the success of the correction.

Supplementary Data 1 (Separate PDF File). Overview figures of 24 multi-peak loci

For each of the twenty-four multi-peak loci (listed in Supplementary Table 5): (**A**) $-\log_{10}(P)$ of the meta-analysis one-sided, right-tailed *P* value per facial segment in $META_{US}$ and $META_{UK}$ tracks. Black-encircled facial segments have reached genome-wide significance ($P = 5 \times 10^{-8}$). (**B**) The normal displacement (displacement in the direction locally normal to the facial surface) in each quasi landmark of the facial segment reaching the lowest *P* value in $META_{US}$ and $META_{UK}$, going from the minor to the major allele. Blue indicates inward depression; red indicates outward protrusion. (**C**) LocusZoom plots in $META_{US}$ (top) and $META_{UK}$ (bottom), for the segment in which the SNP had its lowest *P* value (one-sided). Points are colored based on linkage disequilibrium (r^2) in the 1000 Genomes Phase 3 EUR population. Asterisks represent genotyped SNPs and circles represent imputed SNPs.

Supplementary Data 2 (Separate Excel File). Structural Equation Modeling results

For each of the 50 segments with well-fitting SEM models, in this table we provide the number of principal components included to represent shape variation in that segment, the number of SNPs that survived the model refinement process (see Methods), the *P* value cutoff used to perform the model refinement and determine the SNPs to be used for epistasis, the number of SNPs used in the epistasis analysis for this segment, and values for the χ^2, CFI, RMSE, SRMR model fit indices, which were used to evaluate the models for our analysis. We also include the TLI and GFI model fit indices for completeness. This table also contains internal links to separate tabs where, for each surviving model, we have listed the parameters used and the estimate, standard error, z-score, two-sided *P* value, and 95% confidence intervals. SNPs which were selected for epistasis testing are highlighted in green.

References for Supplementary Material

1. Ferreira, M. A. R. & Purcell, S. M. A multivariate test of association. *Bioinformatics* **25**, 132–133 (2009).
2. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.* **61**, 861–866 (2016).
3. White, J. & Indencleef, K. Insights into the genetic architecture of the human face. (2020) doi:10.6084/m9.figshare.c.4667261.
4. Stelzer, G. *et al.* VarElect: the phenotype-based variation prioritizer of the GeneCards Suite. *BMC Genomics* **17**, 444 (2016).
5. Jensen, L. J. *et al.* STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412-416 (2009).
6. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
7. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
8. Silver, H. K., Kiyasu, W., George, J. & Deamer, W. C. Syndrome of congenital hemihypertrophy, shortness of stature, and elevated urinary gonadotropins. *Pediatrics* **12**, 368–376 (1953).
9. Russell, A. A syndrome of intra-uterine dwarfism recognizable at birth with cranio-facial dysostosis, disproportionately short arms, and other anomalies (5 examples). *Proc. R. Soc. Med.* **47**, 1040–1044 (1954).
10. Kayserili, H. *et al.* ALX4 dysfunction disrupts craniofacial and epidermal development. *Hum. Mol. Genet.* **18**, 4357–4366 (2009).

11. Rochus, C. M. *et al.* Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. *BMC Genomics* **19**, 71 (2018).
12. Qu, S., Tucker, S. C., Zhao, Q., deCrombrugghe, B. & Wisdom, R. Physical and genetic interactions between Alx4 and Cart1. *Development* **126**, 359–369 (1999).
13. Lamichhaney, S. *et al.* Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
14. Lamichhaney, S. *et al.* A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352**, 470–474 (2016).
15. Nishikata, I. *et al.* A novel EVI1 gene family, MEL1, lacking a PR domain (MEL1S) is expressed mainly in t(1;3)(p36;q21)-positive AML and blocks G-CSF–induced myeloid differentiation. *Blood* **102**, 3323–3332 (2003).
16. Mochizuki, N. *et al.* A novel gene, MEL1, mapped to 1p36.3 is highly homologous to the MDS1/EVI1 gene and is transcriptionally activated in t(1;3)(p36;q21)-positive leukemia cells. *Blood* **96**, 3209–3214 (2000).
17. Takahata, M. *et al.* SKI and MEL1 cooperate to inhibit transforming growth factor-beta signal in gastric cancer cells. *J. Biol. Chem.* **284**, 3334–3344 (2009).
18. Bjork, B., Turbe-Doan, A., Prysak, M., Herron, B. & Beier, D. Prdm16 is required for normal palatogenesis in mice. *Hum Mol Genet* **19**, 774–789 (2009).
19. Stamataki, D., Ulloa, F., Tsoni, S. V., Mynett, A. & Briscoe, J. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes Dev.* **19**, 626–641 (2005).
20. Koyabu, Y., Nakata, K., Mizugishi, K., Aruga, J. & Mikoshiba, K. Physical and functional interactions between Zic and Gli proteins. *J. Biol. Chem.* **276**, 6889–6892 (2001).
21. Wang, B., Fallon, J. F. & Beachy, P. A. Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell* **100**, 423–434 (2000).

631 22. Hui, C. C. & Joyner, A. L. A mouse model of greig cephalopolysyndactyly syndrome: the
632 extra-toesJ mutation contains an intragenic deletion of the Gli3 gene. *Nat. Genet.* **3**, 241–
633 246 (1993).
634 23. Adhikari, K. *et al.* A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1
635 and EDAR in human facial variation. *Nat. Commun.* **7**, 1–11 (2016).
636 24. Stegman, M. A. *et al.* Identification of a tetrameric hedgehog signaling complex. *J. Biol.*
637 *Chem.* **275**, 21809–21812 (2000).
638 25. Monnier, V., Dussillol, F., Alves, G., Lamour-Isnard, C. & Plessis, A. Suppressor of fused
639 links fused and Cubitus interruptus on the hedgehog signalling pathway. *Curr. Biol. CB* **8**,
640 583–586 (1998).
641 26. Méthot, N. & Basler, K. Suppressor of fused opposes hedgehog signal transduction by
642 impeding nuclear accumulation of the activator form of Cubitus interruptus. *Development*
643 **127**, 4001–4010 (2000).
644 27. Kise, Y., Morinaka, A., Teglund, S. & Miki, H. Sufu recruits GSK3beta for efficient
645 processing of Gli3. *Biochem. Biophys. Res. Commun.* **387**, 569–574 (2009).
646 28. Kogerman, P. *et al.* Mammalian suppressor-of-fused modulates nuclear-cytoplasmic
647 shuttling of Gli-1. *Nat. Cell Biol.* **1**, 312–319 (1999).
648 29. Humke, E. W., Dorn, K. V., Milenkovic, L., Scott, M. P. & Rohatgi, R. The output of
649 Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused
650 and the Gli proteins. *Genes Dev.* **24**, 670–682 (2010).
651 30. Staal, S. P., Hartley, J. W. & Rowe, W. P. Isolation of transforming murine leukemia viruses
652 from mice with a high incidence of spontaneous lymphoma. *Proc. Natl. Acad. Sci. U. S. A.*
653 **74**, 3065–3067 (1977).
654 31. Biesecker, L. G. & Sapp, J. C. Proteus Syndrome. in *GeneReviews*® (eds. Adam, M. P. et
655 al.) (University of Washington, Seattle, 1993).

32. Lindhurst, M. J. *et al.* A Mosaic Activating Mutation in AKT1 Associated with the Proteus Syndrome. *N. Engl. J. Med.* **365**, 611–619 (2011).
33. Bujok, K. *et al.* Interplay between the prostaglandin transporter OATP2A1 and prostaglandin E2-mediated cellular effects. *Cell. Signal.* **27**, 663–672 (2015).
34. Schuster, V. L. Prostaglandin transport. *Prostaglandins Other Lipid Mediat.* **68–69**, 633–647 (2002).
35. Seifert, W. *et al.* Mutations in the prostaglandin transporter encoding gene SLCO2A1 cause primary hypertrophic osteoarthropathy and isolated digital clubbing. *Hum. Mutat.* **33**, 660–664 (2012).
36. Zhang, Z. *et al.* Exome sequencing identifies SLCO2A1 mutations as a cause of primary hypertrophic osteoarthropathy. *Am. J. Hum. Genet.* **90**, 125–132 (2012).
37. Busch, J. *et al.* Mutations in the prostaglandin transporter SLCO2A1 cause primary hypertrophic osteoarthropathy with digital clubbing. *J. Invest. Dermatol.* **132**, 2473–2476 (2012).
38. Rovani, M. T. *et al.* Prostaglandin F2α-induced luteolysis involves activation of Signal transducer and activator of transcription 3 and inhibition of AKT signaling in cattle. *Mol. Reprod. Dev.* **84**, 486–494 (2017).
39. Sales, K. J., Battersby, S., Williams, A. R. W., Anderson, R. A. & Jabbour, H. N. Prostaglandin E2 mediates phosphorylation and down-regulation of the tuberous sclerosis-2 tumor suppressor (tuberin) in human endometrial adenocarcinoma cells via the Akt signaling pathway. *J. Clin. Endocrinol. Metab.* **89**, 6112–6118 (2004).
40. Mammoto, A. *et al.* A mechanosensitive transcriptional mechanism that controls angiogenesis. *Nature* **457**, 1103–1108 (2009).
41. Haenggi, T. & Fritschy, J.-M. Role of dystrophin and utrophin for assembly and function of the dystrophin glycoprotein complex in non-muscle tissue. *Cell. Mol. Life Sci. CMLS* **63**, 1614–1631 (2006).

682 42. Ahn, A. H. *et al.* Cloning of human basic A1, a distinct 59-kDa dystrophin-associated protein
683 encoded on chromosome 8q23-24. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4446–4450 (1994).
684 43. Erriquez, D., Perini, G. & Ferlini, A. Non-Coding RNAs in Muscle Dystrophies. *Int. J. Mol.*
685 *Sci.* **14**, 19681–19704 (2013).
686 44. Tian, L. *et al.* Unveiling transcription factor regulation and differential co-expression genes in
687 Duchenne muscular dystrophy. *Diagn. Pathol.* **9**, 210 (2014).
688 45. Peter, A. K. & Crosbie, R. H. Hypertrophic response of Duchenne and limb-girdle muscular
689 dystrophies is associated with activation of Akt pathway. *Exp. Cell Res.* **312**, 2580–2591
690 (2006).
691 46. Imuta, Y., Nishioka, N., Kiyonari, H. & Sasaki, H. Short limbs, cleft palate, and delayed
692 formation of flat proliferative chondrocytes in mice with targeted disruption of a putative
693 protein kinase gene, Pkdcc (AW548124). *Dev. Dyn.* **238**, 210–222 (2009).
694 47. Kinoshita, M., Era, T., Jakt, L. M. & Nishikawa, S.-I. The novel protein kinase Vlk is essential
695 for stromal function of mesenchymal cells. *Development* **136**, 2069–2079 (2009).
696 48. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and
697 reveals 14 loci associated with risk of fracture. *Nat. Genet.* **44**, 491–501 (2012).
698 49. Weinberg, S. M. *et al.* The 3D Facial Norms Database: Part 1. A Web-Based Craniofacial
699 Anthropometric and Image Repository for the Clinical and Research Community. *Cleft*
700 *Palate. Craniofac. J.* **53**, e185–e197 (2016).
701 50. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon
702 Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
703 51. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children:
704 ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
705 52. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
706 datasets. *GigaScience* **4**, 7 (2015).

53. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
54. Deelen, P. *et al.* Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. *BMC Res. Notes* **7**, 901 (2014).
55. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
56. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
57. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
58. Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinforma. Oxf. Engl.* **30**, 1266–1272 (2014).
59. Claes, P., Walters, M., Vandermeulen, D. & Clement, J. G. Spatially-dense 3D facial asymmetry assessment in both typical and disordered growth. *J. Anat.* **219**, 444–55 (2011).
60. Claes, P. *et al.* Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat. Genet.* **50**, 414–423 (2018).
61. White, J. D. *et al.* MeshMonk: Open-source large-scale intensive 3D phenotyping. *Sci. Rep.* **9**, 6085 (2019).
62. Rohlf, F. J. & Slice, D. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Syst. Biol.* **39**, 40–59 (1990).
63. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
64. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

65. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
66. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493-496 (2004).
67. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). Online Mendelian Inheritance in Man, OMIM®. https://omim.org/ (2019).
68. 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).
69. Wei, C.-H., Allot, A., Leaman, R. & Lu, Z. PubTator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Res.* **47**, W587–W593 (2019).
70. Hooper, J. E. *et al.* Systems biology of facial development: contributions of ectoderm and mesenchyme. *Dev. Biol.* **426**, 97–114 (2017).
71. Watanabe, K., Taskesen, E., Bochoven, A. van & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
72. Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
73. 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).
74. Baumgart, S. J. *et al.* CHD1 regulates cell fate determination by activation of differentiation-induced genes. *Nucleic Acids Res.* **45**, 7722–7735 (2017).
75. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease risk association. *Science* **366**, 1134–1139 (2019).
76. 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).

757 77. Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J. P. & Cotney, J. High-Resolution
758 Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep.* **23**, 1581–1597
759 (2018).
760 78. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human
761 epigenomes. *Nature* **518**, 317–330 (2015).
762 79. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a Web-based tool for identification
763 and annotation of matched SNPs. *Bioinformatics* **31**, 418–420 (2015).
764 80. Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. & King, J. Reporting Structural
765 Equation Modeling and Confirmatory Factor Analysis Results: A Review. *J. Educ. Res.* **99**,
766 323–338 (2006).
767 81. Wolf, E. J., Harrington, K. M., Clark, S. L. & Miller, M. W. Sample Size Requirements for
768 Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educ.*
769 *Psychol. Meas.* **76**, 913–934 (2013).
770 82. Kline, R. B. *Principles and Practice of Structural Equation Modeling*. (The Guilford Press,
771 2015).
772 83. Hooper, D., Coughlan, J. & Mullen, M. R. Evaluating Model Fit: a Synthesis of the Structural
773 Equation Modelling Literature. in 1–10 (2008). doi:10.21427/D79B73.
774
775