HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2018 September 01.

Published in final edited form as:

Nat Genet. 2018 March ; 50(3): 414–423. doi:10.1038/s41588-018-0057-4.

Genome-wide mapping of global-to-local genetic effects on human facial shape

Peter Claes[1,2,3,*], **Jasmien Roosenboom**[4], **Julie D. White**[5], **Tomek Swigut**[6], **Dzemila Sero**[1,2], **Jiarui Li**[1,2], **Myoung Keun Lee**[4], **Arslan Zaidi**[5], **Brooke C. Mattern**[5], **Corey Liebowitz**[5], **Laurel Pearson**[5], **Tomás González**[5], **Elizabeth J. Leslie**[4], **Jenna C. Carlson**[7], **Ekaterina Orlova**[8], **Paul Suetens**[1,2], **Dirk Vandermeulen**[1,2], **Eleanor Feingold**[7,8], **Mary L. Marazita**[4,8], **John R. Shaffer**[8], **Joanna Wysocka**[6,9,*], **Mark D. Shriver**[5,*], and **Seth M. Weinberg**[4,10,*]

[1]Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium

[2]Medical Imaging Research Center, MIRC, UZ Leuven, Leuven, Belgium

[3]Murdoch Childrens Research Institute, Melbourne, Victoria, Australia

[4]Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

[5]Department of Anthropology, Penn State University, University Park, Pennsylvania, United States of America

[6]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California, United States of America

[7]Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

[8]Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

[9]Department of Developmental Biology, Stanford University School of Medicine, Stanford, California, United States of America

[10]Department of Anthropology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

[*]Corresponding authors: Peter Claes (peter.claes@kuleuven.be) ORCID: 0000-0001-9489-9819; Joanna Wysocka (wysocka@stanford.edu) ORCID: 0000-0002-6909-6544; Mark Shriver (mds17@psu.edu) 0000-0003-0006-0479; Seth M. Weinberg (smwst46@pitt.edu) ORCID: 0000-0001-9467-4556.

Author Contributions

This work is the results of a four center collaboration between the KU Leuven, University of Pittsburgh, Penn State University and Stanford University School of Medicine, led by P.C., S.M.W, M.D.S, and J.W. respectively. All four centers contributed equally to this work. P.C, and D.S. with input from P.S and D.V. conceptualized and implemented the global-to-local facial segmentations. J.L provided genomic data support and analysis at the KU Leuven site and together with P.C ran the GWAS. M.K.L, J.R., E.J.L, J.C.C., E.O. E.F., M.L.M., J.R.S. and S.M.W organized the PITT cohort and co-analyzed the GWAS results, generated locus zoom plots and organized the results into 38 loci. J.D.W, A.Z., B.C.M., C.L., L.P., T.G., and M.D.S organized the PSU cohort, imputed the PSU genetic data and co-analyzed the GWAS results. T.S., and J.W. performed the GREAT analysis and the association with CNCCs. P.C. wrote the manuscript with extensive input from all co-authors.

Competing financial interest

The authors have no financial conflict of interest to report.

Abstract

Genome-wide association scans of complex multipartite traits like the human face typically use preselected phenotypic measures. Here we report a data-driven approach to phenotyping facial shape at multiple levels of organization, allowing for an open-ended description of facial variation, while preserving statistical power. In a sample of 2,329 persons of European ancestry we identified 38 loci, 15 of which replicated in an independent European sample (n=1,719). Four loci were completely novel. For the others, additional support (n=9) or pleiotropic effects (n=2) were found in the literature, but the results reported here were further refined. All 15 replicated loci revealed distinctive patterns of global-to-local genetic effects on facial shape and showed enrichment for active chromatin elements in human cranial neural crest cells, suggesting an early developmental origin of the facial variation captured. These results have implications for studies of facial genetics and other complex morphological traits.

The human face is a multipartite trait composed of distinct features (i.e., eyes, nose, chin, and mouth), whose size, shape and composition are clearly heritable. However, our knowledge of which genetic variants are responsible for human facial variation is still lacking[1]. Several genome-wide association studies (GWAS) have each identified a handful of loci associated with a small number of facial traits, few of which have been replicated[2–7]. While a GWAS benefits from being an unbiased approach to gene mapping, the phenotypic descriptions used in such studies are typically preselected and used to classify individuals in a “phenotype-first” way of thinking. This approach may be appropriate in certain instances (e.g., affected and unaffected disease status), but is less so for complex multipartite traits like the human face. The result is that facial shape has been reduced to a limited series of measurements (e.g., linear distances) that are analyzed individually, resulting in a loss of information.

In this study, we present a data-driven approach to facial phenotyping that exploits both the partable and integrated information contained in 3D facial images, allowing for the identification of genetic effects on facial shape at multiple levels of organization – from global-to-local. This approach generates a nested series of multivariate GWAS, with a low computational burden and, more importantly, controlled multiple testing burden. We applied this new approach to a European-derived discovery cohort and then tested significant variants for replication in an independent European-derived cohort. In an effort to provide additional validation, we integrated our work with previously published human facial GWAS. We show a number of novel genetic loci supported by strong statistical evidence, revealing unseen patterns in global-to-local genetic effects on facial shape. Furthermore, these loci are preferentially marked by active chromatin signatures in human cranial neural crest cells, an embryonic cell type that gives rise to most of the craniofacial structures. This suggests a developmental origin of much of the facial variation uncovered by our study. These results offer novel insights on the genetic basis of human facial shape with potentially far reaching implications. More generally, the results present an alternative to the prevailing phenotype-first mindset and is widely applicable to any GWAS on complex, quantitative and multipartite traits, especially those captured thoroughly using images.

Results

Study Samples

A study sample of 2,329 unrelated participants of European ancestry comprised the discovery cohort for our analysis (the Pittsburgh sample, PITT). These participants had a median age of 23 years and were recruited from several US sites. An additional, independently collected and genotyped, sample of 1,719 adult participants of European ancestry comprised our replication cohort (the Penn State sample, PSU). These participants had a median age of 22 years and were recruited at several sites in the US and Europe. For both cohorts, imputation of unobserved genetic variants and sporadic missing genotype calls for assayed SNPs was performed using the 1000 Genomes Project[8] Phase 3 reference panel. Basic demographic descriptors and general physical characteristics (e.g., sex, age, height, and weight) were available from participants in both cohorts.

Global-to-local facial segmentations

We used digital stereophotogrammetry to obtain 3D facial images from all participants. After trimming and cleaning, 3D facial images were aligned in dense correspondence[9], ensuring that homology was established among roughly 10,000 3D points (considered quasi-landmarks[10]) comprising an individual's facial shape. Facial shape, in both cohorts separately, was adjusted for potential confounding variables (age, sex, height, weight, facial size, and cohort-specific population structure), and then partitioned unsupervised into a series of global-to-local facial segments. To accomplish this, we applied hierarchical spectral clustering[11] to the facial quasi-landmark configurations of the PITT sample. First, the quasi-landmarks were listed in a squared (~10,000 × ~10,000) similarity matrix using pairwise 3D correlations. Second, a Laplacian transformation was applied to enhance similarities prior to an Eigen decomposition of the squared matrix. Finally, within the Eigen spectral map, K-means++ clustering was used to group highly correlated quasi-landmarks, that, when mapped back to the facial surface, result in a segmentation of the face. This was done in a bifurcating hierarchical manner using five levels, such that facial segments with closer relationships are located nearby one another and each facial segment is split in two towards the next level resulting in a total of 63 segments, as depicted in Figure 1 using a polar dendrogram layout. The hierarchical design gradually focused on more local shape aspects that were otherwise overlooked, without ignoring the integration of facial parts more globally at previous levels. The same pipeline was applied to the PSU cohort (Figure S1) and the resulting segmentation was compared to the PITT segmentation using the Normalized Mutual Information ($0 < NMI < 1$; 0 = no overlap, 1 = perfect overlap) at each of the five levels ($NMI_{L0} = 1$, $NMI_{L1} = 0.90$, $NMI_{L2} = 0.80$, $NMI_{L3} = 0.72$, $NMI_{L4} = 0.75$, $NMI_{L5} = 0.78$). These high NMI values indicate substantial overlap and, therefore, good reproducibility of the global-to-local facial segmentations across similar but independent samples. The hierarchical spectral clustering (Figure 1, using the PITT cohort as the reference for the remainder of the study), subdivided facial shape into meaningful and recognizable segments. Globally, the midface was first separated from the rest of the face, and further partitioned into the region of the mouth (quadrant 1, starting at segment 4) and nose (quadrant 2, starting at segment 5). The remainder of the face was further partitioned into the lower facial area (quadrant 3, starting at segment 6) and upper facial area (quadrant

4, starting at segment 7). Each quadrant was repeatedly partitioned, increasingly focusing on smaller facial parts. This provided an efficient and objective means for subdividing facial shape into global-to-local parts with internally well correlated shape variations.

In order to generate variables that captured biological shape, we applied generalized Procrustes analysis (GPA) separately to the quasi-landmarks comprising each facial segment. As such, a shape-space for each facial segment was constructed independent from the other segments and its relative positioning in lower-level (larger) segments. Following GPA, we applied principal component analysis (PCA) in order to extract the major factors of shape variation characterizing each facial segment. We used parallel analysis[12] to determine the number of principal components (PCs) needed to adequately summarize shape variation for the given segment. The number of PCs retained for each facial segment after parallel analysis and the percentage of total variation they explain in the PITT cohort are depicted in Figure S2. As expected, we found that lower-level facial segments require more PCs, and that the retained PCs for all facial segments explain most of the total shape variance (median of 95%, min = 89%, max = 97%) present within the respective segments. The result was near complete coverage of all facial shape variation at five different levels of detail.

Genetic mapping of global-to-local shape

We performed a series of GWAS scans to test the genetic association of a total 9,478,608 SNPs with the shape information contained in each of the 63 facial segments using the PITT discovery cohort. Since each of the facial segments was represented by multiple dimensions of variation (PCs), we used a multivariate canonical correlation analysis (CCA)[13]. In brief, CCA extracts the linear combination of PCs from a facial segment that has maximal correlation with the SNP being tested, under the additive genetic model, after correcting for confounding variables. Therefore, CCA avoids the preselection of individual PCs and reveals the associated facial effect as a linear direction in a multi-dimensional shape-space.

We found a total of 1,932 SNPs among 38 separate loci that reached nominal genome-wide significance ($p \leq 5 \times 10^{-8}$). Quantile-quantile, regional association and in detail supplementary illustrations of all 38 loci are provided online (see URLs in the online methods). To accommodate the multiple testing burden present in performing separate genome scans for 63 facial segments, we determined the false discovery rate for any test dependency structure[14] (FDRd) p-value threshold at 2.82×10^{-8} and a more conservative study-wide Bonferroni corrected significance threshold at 1.28×10^{-9}. Following imputation in the PSU cohort, 1,931 of these 1,932 SNPs were available for replication testing. Because it is possible for a SNP to affect different aspects of variation in the same facial segment across different datasets, replicating an observed genetic association necessitated both replicating a significant effect on a given facial segment and a determination that the same aspects of variation were affected. Therefore, we first used the discovery cohort as a phenotyping reference for the replication cohort (see online Methods for additional details). Then, the multi-dimensional PC scores of the PSU cohort were projected onto the CCA loadings of the discovered effect. In doing so, the associated facial trait, once identified using CCA in the PITT cohort, was kept fixed and thus consistent between the PITT discovery cohort and the PSU replication cohort. This resulted in a univariate facial measure

that was modeled using linear regression, to test for replication. This was done separately for each SNP and facial segment combination that achieved genome-wide significance in the discovery phase. From all replication efforts combined, we computed the FDRd adjusted significance threshold at 4.20×10^{-3} and a Bonferroni corrected significance threshold at 3.28×10^{-4}. A total of 1,821 SNPs across 15 loci replicated with p-values well below the FDRd threshold (Table 1).

We focused on the 15 replicated loci to investigate patterns of genome-to-facial shape associations. These 15 loci involved a variety of facial segments, and several loci affected segments in more than one facial region as illustrated in Figures 2 and 3. Several interesting patterns were observed. First, the majority of affected segments were located in the nose and the lower facial quadrant, primarily the chin. Second, some associations were involved in high level segments, often at the outer layer(s) of the polar dendrogram, suggesting a localized phenotypic effect (e.g., 1p32.1). In contrast, we observed other associations involving numerous linked facial segments from the edge ($Level_5$) of the dendrogram to the center ($Level_0$). Interestingly, the evidence of association (by tracking the −log10 p-values in Figure 3) in these instances may increase from center-to-edge (global-to-local, e.g., 7q21.3) or from edge-to-center (local-to-global, e.g., 1p12), reflecting a localized versus global phenotypic effect respectively. Still others, e.g., 1q31.3, reached a maximum association part-way up the dendrogram reflecting an intermediate phenotypic effect. Interestingly, both of the loci on chromosome 17 show the same pattern of association across nose segments, but with 17q24.3 (a, rs11655006:T>C) showing the stronger association at the more global levels, and 17q24.3 (b, rs5821892:C>CG) showing strongest association at the most local level. Third, most associations were limited to linked segments within one facial quadrant of the polar dendrogram, but in a few cases, genetic associations involved two distinct quadrants, e.g., 2q31.1 and 3q21.3. Finally, in the online illustrations, we observe, that facial effects propagate consistently in linked facial segments across different levels in the hierarchical design.

For completeness, we present an overview of all 38 loci identified in the discovery cohort with their peak SNP statistical details and imputation quality scores in Table S1 and S2, respectively. For 14 of the non-replicating loci, association was represented by a single SNP and for an additional five loci, the peak SNP had a minor allele frequency (MAF) less than 2%. Another locus that did not replicate showed a significant discrepancy in peak SNP MAF between the discovery and replication cohorts. For the remaining 18 loci (also given in Table 1), 16 loci replicated at a nominal level ($p \leq 0.05$), 15 of which replicated below the FDRd threshold and 12 of which replicated below the Bonferroni threshold. Two loci with robust associations in the discovery cohort failed to replicate. Additional illustrations on the discovery and replication results of all 38 loci are presented online.

Association between 15 replicated GWAS regions and signatures of active regulatory elements in Cranial Neural Crest Cells (CNCCs)

To explore biological processes and phenotypes associated with the identified GWAS loci, we performed Genomic Regions Enrichment of Annotations (GREAT) analysis[15]. Remarkably, even though only 15 loci were used in this analysis, we detected significant

associations with craniofacial development, including categories such as chondrocyte differentiation, 1st and 2nd branchial arch mesenchyme, development of facial bones, and cleft palate (Figure S3). Although craniofacial development is complex and involves interactions between multiple cell types, Cranial Neural Crest Cells (CNCCs), an embryonic cell group that arises at 3–6 weeks of gestation, plays the central role in the formation of the facial plan and in determining its species-specific and individual variation[16–18]. During embryogenesis CNCCs migrate away from the neural tube along stereotypical paths and form the majority of cranial mesenchyme, which then differentiates to the bone, cartilage, and connective tissue of the face and head[19,20]. We reasoned that if divergence of the facial shape within the human population captured by our GWAS indeed originates early during embryogenesis, then the 15 loci should be preferentially active in CNCCs. To test this, we took advantage of the epigenomic mapping datasets generated from human CNCCs, which were derived *in vitro* from embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs)[21,22]. Acetylation of histone H3 on lysine 27 (H3K27ac) has been associated with transcriptionally active gene promoters and with active distal enhancer regions, and thus is a useful mark to consider when exploring cell type-specific activity of both coding and non-coding regions of the genome[23,24]. We quantified H3K27ac ChIP-seq signals in the vicinity (e.g., within 10 kb) of the 15 peak SNPs in CNCCs of different genetic backgrounds and compared them to the H3K27ac signals over the same regions in over 30 other cell types, representing distinct adult, embryonic and *in vitro* derived cells. As shown in Figure 4, we observed higher H3K27ac signals in the vicinity of the peak SNPs in CNCCs than in any other examined cell type. These observations are consistent with the preferential activity of the identified loci in CNCCs and with an embryonic origin of the human facial variation captured by our study.

A large majority of genetic variants associated with human trait variation map to the non-coding parts of the genome. Many such variants are thought to reside within cis-regulatory elements[25]. However, cis-regulatory elements, especially distal enhancers, are often characterized by highly cell-type specific activity patterns (which are reflected in their cell-type specific chromatin marking patterns), underscoring the importance of analyzing relevant cell types for GWAS interpretation (reviewed in Long et al.[26,27]). To explore the overlap between the 15 loci and regulatory regions active in CNCCs we used our epigenomic datasets (including p300, *TFAP2A*, *NR2F1*, H3K4me1, H3K4me3, H3K27me3 and H3K27ac ChIP-seq and nucleosomal depletion analysis by ATAC-seq) to identify and classify cis-regulatory elements based on the type of element and activity. Specifically, transcription factor binding and transposase hypersensitivity maps were used to identify genomic positions of candidate regulatory regions, whereas relative enrichments of H3K4me1 to H3K4me3 were used to distinguish enhancers from promoters, and H3K27ac vs H3K27me3 signals to infer active, pre-marked or poised/repressed activity states (depicted in different colors in Figure 5(A)). We observed that the peak SNPs are more likely to be located within or in the immediate vicinity (< 1kb) of a detectable regulatory chromatin feature in CNCCs ($p = 1.58 \times 10^{-3}$). Furthermore, these SNPs are enriched in enhancer elements, especially strong active CNCC enhancers ($p = 1.19 \times 10^{-2}$, odds ratio 7.7, Fisher's exact test). An example of this is shown in Figure 5(B): peak SNP rs6740960:A>T (affecting the lower facial quadrant) is located within an active CNCC

enhancer marked by p300, H3K27ac, and H3K4me1. Interestingly, the same variant has also been recently associated with an elevated risk of CL/P in a European cohort[28]. Furthermore, comparisons with our epigenomic datasets previously obtained from the chimpanzee CNCCs[21] revealed that the overlapping enhancer is chimp-biased in activity (as evidenced by the elevated levels of H3K27ac in chimp as compared to human CNCCs). This example, and other examples of loci (e.g., *PAX3*) that showed association with facial variation, disease and enrichment for regulatory elements divergent between human and chimp, raise an intriguing possibility that the genetic variation within an overlapping set of loci/regulatory elements may influence both species-specific and individual facial shape variation in humans as well as determine predisposition to craniofacial malformations.

Candidate genes and integration with facial GWAS literature

A literature-based annotation of genes was performed and their involvement in facial variation is reported in supplementary Note 1. Additionally, in supplementary Note 2, Tables S3–S6 and Figures S4–S5, we first report statistical replication for 10 out of 16 SNPs from six previous facial GWAS based on 2D or 3D images[2–7] (Table S3). Second, we consulted GWAS central and Phenoscanner[29] and report additional evidence for nine SNPs in Table 1 (Table S4). Third, we cross-referenced 10 of our loci with a recent study[30] based on self-reported information on chin dimple and nose size (Table S5). Table 1 indicates the involvement of each locus in any of these three integration efforts. In summary, for the 15 replicated loci in this work, four were completely novel, nine were previously discovered showing consistent facial effects and two were previously discovered with pleiotropic facial effects.

Discussion

A deeper understanding of the genetic basis of human facial traits may provide insights into the mechanisms of craniofacial morphogenesis, improve our knowledge of the complex relationship between genotype and phenotype in craniofacial syndromes and birth defects, and eventually provide a basis for predicting facial features in numerous applications, ranging from early diagnosis, to personalized medicine, treatment planning in craniofacial surgery and orthodontics, and biometrics and forensics. In mice, whole-genome QTL studies have identified numerous genetic regions associated with craniofacial traits[31–33]. In humans, both candidate gene[10,34,35] and GWAS[2–7] have implicated specific variants. While the highly complex nature of craniofacial morphogenesis suggests that many genes are likely to influence facial morphology, the number of associated loci uncovered so far using GWAS has been limited with few independent replications.

Although traditional anthropometric facial measures remain a valuable resource for anthropologists and clinicians, because of their widespread use and simplicity, limited GWAS results with such measurements have been obtained so far. One reason for this is the difficulty inherent in defining proper facial types or relevant facial measurements *a priori*. As noted by, Adhikari et al.[4], quantitative measures, based upon which all previous facial GWAS efforts from 2D or 3D images have been conducted, are expected to yield higher statistical power compared to categorical ones, but they have not yet resulted in many robust

findings. Using well-defined categorical scale ratings instead, they discovered and replicated several loci, which we also replicate here. Similarly, but using self-reported categorizations on chin dimple and nose size, Pickrell et al[30], identified a number of loci in an impressively-sized sample (N>70.000). However, the relatively simple categorical phenotypes used in these studies preclude a comprehensive and precise description of the effects of genetic variants on facial morphology. We argue that these approaches are perfectly in line with the traditional phenotype-first thinking common to GWAS and remain a good paradigm for case-control designs and simple quantitative/qualitative traits, but perhaps are less suitable for multipartite quantitative traits like the human face.

The craniofacial complex is initially modulated by precisely timed embryonic gene expression and molecular and cellular processes mediated through complex signaling pathways[36]. As humans grow, hormones, nutritional status, and biomechanical factors affect the face[37,38]. A natural consequence of these forces and constraints during facial morphogenesis and growth is that the face exhibits a modular organization, with suites of facial features at different scales that show internal integration but remain relatively independent from other features[39]. Therefore, we expect the human face to be influenced by many genes that exhibit a range of effects, with some influencing only localized parts of the face and others influencing more global aspects of morphology. As a result, the proposed facial segmentation is related to the concept of modularity and integration in morphological studies[40], with the main addition that the partitioning was done hierarchically, going from globally integrated to locally focused modules. This allows for an investigation of facial shape effects propagating at different scales. Based on structural correlations between a vast number of individual 3D points on facial surfaces, we were able to hierarchically cluster the human face into segments. These were learned from the data, thus unsupervised and data-derived instead of candidate-driven, and could be duplicated across independent cohorts. This allowed for an efficient divide and conquer strategy enabling the screening of a huge set of facial variations, without compromising statistical power.

Efficient use of phenotypic data from modest-sized cohorts is necessary for investigating 3D facial shape. In contrast to a recent GWAS on human body height[41], which, due to the widespread availability of the phenotype, was able to identify 423 loci in combined samples over two orders of magnitude larger than what was used here, the specialized phenotype data needed for investigating 3D facial shape is limited to a few, smaller cohorts. Using the PITT cohort from Shaffer et al.[6], we managed to substantially increase the number of loci identified. In fact, these results are unmatched by any imaging-based facial GWAS to date.

The global and local facial patterns demonstrated by the loci uncovered may help elucidate their roles during craniofacial development. It is clear that many of the genes at these loci are expressed in relevant tissues during embryonic development. Connecting those patterns of early expression and function to eventual morphology is a major challenge. The kind of highly refined facial effects presented here can help provide a roadmap to clarify the connections between molecules and morphology. Consider, for instance, two loci associated with different aspects of nasal shape and depicted in Figure S6. 19q13.11 (*KCTD15*) showed a highly focal effect limited to the nasal tip. In contrast, 6p21.1 (*SUPT3H*) affected primarily the nasal root and lateral parts of the nasal bridge, effectively sparing the nasal tip. A similar

nasal phenotype was observed by Adhikari et al[4] at 6p21.1. These highly specific phenotypic effects might provide important insights about the role of these two genes during human facial morphogenesis and growth. *KCTD15* has been shown to regulate *TFAP2A*, which plays a critical role in neural crest formation[42] and when mutated resulted in reduced snout length in mice among other defects[43]. Perhaps *KCTD15* affects nasal tip shape in humans by influencing cartilage proliferation in the nasal septum, whereas *SUPT3H* exerts its influence on nasal shape by affecting portions of the maxilla and nasal bones. The precision of the phenotypic effects revealed here can form the basis of such testable hypotheses in the lab. As another example, previous research extensively described the role of *SOX9* in the development of a broad range of tissues[44]. For facial development specifically, *SOX9* is expressed in cranial neural crest cells, populating the pharyngeal arches in the head region. Previous animal and developmental studies described that the first pharyngeal arch gives rise to the maxillary and mandibular prominences[45]. However, the results of our study shows that the locus associated with *SOX9* (17q24.3) and the one located 1 Mb away (*CASC17*, which is known to functionally interact with *SOX9* through transcription factors), both influence nasal shape, although the nose is known to be formed from the frontonasal prominence. This raises a fascinating question about the involvement of *SOX9* during the development of the nose, and is an interesting subject for functional follow-up studies.

In our search through web-based GWAS repositories, we did find associations ($p<1\times10^{-7}$) to non-facial traits for loci in Table 1, including aspects of body size and composition (waist-hip ratio and height, *TBX15/WARS2* and *SUPT3H*, respectively), developmental traits like age at menarche (*RAB7A*), and brain DNA methylation levels (*RPS12/EVA4*) and one associated with risk of biliary atresia (*PAX3*). That genes and more specifically, even relatively small gene regions, such as those identified in GWAS, have multiple functions in the body or pleiotropy, has been appreciated for some time[46]. However, the extent to which pleiotropy plays a role in trait variation and disease risk is perhaps much greater than what previously has been appreciated[47].

In conclusion, we proposed a data-driven global-to-local facial phenotyping approach, well-suited for a genome-wide association scan. Using this approach, we have substantially advanced the literature on facial genetics on several fronts. First, we identified and replicated a number of new genetic loci using modest-sized cohorts. Second, we provided additional support for numerous previously identified loci and showed a strong integration of our results with the facial GWAS literature. Third, we demonstrated the preferential activity of the replicated loci in CNCCs, consistent with an embryonic origin of the human facial variation captured by our study. Last, we revealed unseen patterns of global-to-local genetic effects on facial shape, supporting the genetic organization of facial features at different scales.

Online Methods

Sample and Recruitment Details

For the discovery cohort (the Pittsburgh (PITT) sample), data from 2,449 participants were obtained from the 3D Facial Norms repository[48]. The repository includes 3D facial surface

images and self-reported demographic descriptors (e.g., age, ancestry) as well as basic physical characteristics (e.g. height and weight) from individuals recruited at four U.S. sites: Pittsburgh, PA; Seattle, WA; Houston, TX; and Iowa City, IA. Recruitment was limited to individual's age three to 40 years 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. A total of 2,329 participants, were retained for analysis, after removing 120 participants, with missing information on sex, age, height, weight or with 3D image mapping artifacts (N = 22).

For the replication cohort (the Penn State (PSU) sample), participants were recruited through several studies at The Pennsylvania State University and sampled in the following locations: State College, PA; New York, NY; Urbana-Champaign, IL; Twinsburg, OH; Dublin, Ireland; Rome, Italy; Warsaw, Poland, and Porto, Portugal. Participants self-reported information on age, ethnicity, ancestry, and body characteristics, and data were gathered on height, weight, and pigmentation of the hair and skin. Individuals were excluded from the analysis if they were below 18 years of age (range 18 – 88) and if they reported a personal history of significant facial 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 (N = 2,059). Participants were removed because of missing sex, age, height, or weight information (N=81) or the presence of 3D image mapping artifacts (N = 33). A further reduction to N = 1,719 was done by excluding participants who were not included in the genotype imputation efforts.

Facial Phenotyping, 3D imaging quality control and shape variables

3D facial surface imaging—Digital facial stereophotogrammetry was used to capture 3D facial surfaces in both samples. This well-established approach uses digital photography to generate a dense 3D point cloud representing surface geometry of the face from multiple 2D images with overlapping fields of view. For the Pittsburgh sample, facial surfaces were acquired using 3dMDface camera systems (3dMD, Atlanta, GA). For the Penn State sample, 3D images were obtained with 3dMDface or Vectra H1 (Canfield Scientific, Parsippany, NJ) systems. Applying standard facial image acquisition protocols[49], participants were asked to close their mouths and hold their faces with a neutral expression.

Spatially-dense facial quasi-landmarking—3D images in wavefront .obj file format were imported into an in-house 3D image cleaning program for cropping and trimming, removing hair, ears, and any dissociated polygons. Five crude positioning landmarks were placed on the face to establish a rough facial orientation. An anthropometric mask (AM)[50] was non-rigidly mapped[9] onto all 3D images and their reflections, which were constructed by changing the sign of the x-coordinate[51]. The AM is essentially a surface template covering the facial area of interest, and the mapping thereof onto the facial images is a process equivalent to the indication of traditional landmarks. This establishes homologous spatially-dense (~10,000) quasi-landmark (QL) configurations for all 3D images and their reflections. In other words, image data from different individuals became standardized, enabling a spatially-dense analysis.

Facial size was obtained as the centroid size of the QL configurations. Facial shape was symmetrized using a generalized Procrustes analysis (GPA)[52] to eliminate differences in position, orientation, and size of both original and reflected QL configurations. The average of an original and its reflected QL configuration constitutes the symmetric component while the difference between the two configurations constitutes the asymmetric component. Although of interest, in this work we currently ignore variations in facial asymmetry. Therefore, unless otherwise mentioned, when discussing facial shape we always refer to the symmetric QL configuration.

Facial quality control—Outlier faces, due to QL mapping errors, were detected by measuring the Mahalanobis distance for each face to the overall average face in the symmetrized shape-space spanned by an orthogonal basis of principal components that captures 98% of the total variation. From the distribution of Mahalanobis distances, a z-score for each facial shape was established, and each face with a z-score equal to or larger than two was manually inspected for QL errors. Identified erroneous faces were removed, and the whole process starting from the generalized Procrustes superimposition of both original and reflected QL configurations was repeated again. The PSU and PITT datasets were processed separately with the same AM, resulting in error-free symmetric and compatible (homologous) QL configurations across both datasets.

Phenotyping the discovery panel—The PITT cohort served as discovery panel. First, the superimposed and symmetrized facial shapes were corrected using a partial least-squares regression (PLSR, function *plsregress* from Matlab™ 2016b) for the confounders of sex, age, age^2, weight, height, facial size and the first four genetic PCA axes to correct for population stratification. Of note, while correcting for genetic background and other confounders is a required step in genetic mapping efforts, it additionally ensures that structural facial variations due to the confounders do not influence the global-to-local segmentation of faces. Second, a 3D correlation, using the RV coefficient[53], between each pair of corrected QLs was computed to construct the squared similarity matrix (~10,000 × ~10,000). Subsequently, a hierarchical spectral clustering with 5 levels was performed, resulting in a total of 63 hierarchically linked facial segments, with 1,2,4,8,16,32 non-overlapping modules at levels 0,1,2,3,4,5 (Figure 1). The hierarchical design provides a step-wise focused shape analysis at different scales, by gradually zooming in to more local shape variations without ignoring the integration of facial parts at previous levels. Furthermore, specific shape effects should propagate consistently in facial segments linked across different levels, lending additional support that the discovered signals are robust. At each level, for each segment, all QLs in the segment are subjected to a new GPA. As such, a shape-space for each facial module is constructed independently of the other modules and its relative positioning within the full face. This is particularly interesting for smaller shape variations (e.g., nose tip) that when superimposed using the full facial surface are overlooked and become undetectable. Finally, after GPA, each modular shape space is spanned by an orthogonal basis using principal component (PC) analysis (PCA) combined with parallel analysis[12] to determine the number of significant components contributing to facial shape.

This application of PCA had several advantages. First, PCs form an orthogonal basis, such that shape variations could be described as linear combinations of PCs. Second, by selecting significant PCs, it was possible to eliminate (via parallel analysis[12]) noisy or meaningless shape variations that result from various sources of error, such as 3D image acquisition and/or QL mapping. Third, PCA is an excellent tool to reduce the dimensions of high dimensional data when strong correlations between the individual data elements are present. The construction of facial modules, as performed here, resulted in the grouping of highly correlated QLs, such that a substantial reduction of dimensions was obtained. This in turn enabled the use of multivariate association techniques, such as CCA, that are constrained in the number of variables tested as a function of the sample size. For almost all segments and especially for the larger segments containing up to ~10,000 quasi-landmarks, this constraint would otherwise have been violated. Furthermore, the multivariate association was more computationally tractable given the lower number of variables tested. Finally, an effect identified in PCA space could be transformed back into the QL shape-space.

Note that some prior studies[2,3,5], also analyzed facial shape data using PCs, but they did so by analyzing them separately. Basically, in these studies full facial multivariate sparse landmark configurations were projected onto each PC, and thus the phenotypic data was preselected along the PCs individually and one-by-one. Any preselection of measurements, either distances or PCs that are subsequently analyzed individually result in a loss of information, as combinations of these measurements are not considered. In contrast, our work uses all PCs together in a single multivariate association effort based on CCA. Therefore, any possible linear combination of PCs is investigated simultaneously, and information is not lost. In CCA, “canonical” is the statistical term for analyzing latent variables (which are not directly observed) that represent multiple variables (which are directly observed). Such an optimization and construction of latent variables is absent in a linear regression on an individual PC.

Phenotyping the replication panel—The PSU cohort was used as replication panel. First, independent of the PITT dataset, the superimposed and symmetrized facial shapes were corrected, using a PLSR, for the confounders of sex, age, age^2, weight, height, facial size and the first three PCA axes of genetic background. The reason for this separate and independent correction was twofold 1) both cohorts have independent axes of genetic background and 2) this workflow shows how future replication efforts can be done without the need to merge, and therefore have access to, all the genetic and covariate data from the discovery cohort. The establishment of spatially-dense quasi-landmarks (AM mapping) allowed for consistent phenotyping. After AM mapping, the PSU faces were segmented using the QL clustering from the discovery PITT cohort. Finally, for each facial segment, all PSU shape instances were superimposed onto the segmental average shape and PCs of the discovery panel. In this way, the replication effort only required the discovery QL cluster labels, average shapes, and PCs for each of the facial segments.

Genotyping, quality checks, imputation, population structure and annotation

The Pittsburgh Sample—Genotyping was performed at the Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Participants, including 70 duplicate samples

and 72 HapMap control samples, were genotyped on the Illumina (San Diego, CA) OminExpress+Exome v1.2 array plus 4,322 investigator-chosen SNPs included to capture variation in specific regions of interest based on previous studies of the genetics of facial variation. Standard data cleaning and quality assurance procedures[54] were performed in collaboration with the University of Washington Genetics Coordinating Center. Specifically, samples were evaluated for concordance of genetic and reported sex, evidence of chromosomal aberrations, biological relatedness across study participants, ancestry, genotype call rate, and batch effects. SNPs were evaluated for call rate, discordant genotype calls between duplicate samples, Mendelian errors in HapMap control parent-offspring trios, deviation from Hardy-Weinberg genotype proportions, and sex differences in allele frequency and heterozygosity.

Imputation of unobserved genetic variants and sporadic missing genotype calls for assayed SNPs was performed using the 1000 Genomes Project[8] Phase 3 reference panel. SHAPEIT2[55] was used for pre-phasing haplotypes and IMPUTE2[56,57] was used to impute nearly 35M variants. SNP-level (INFO score > 0.5) and genotype per participant-level (genotype probability > 0.9) filters were used to omit poorly-imputed variants. Masked variant analyses, in which genotyped SNPs were imputed as though they had not been assayed, indicated high concordance between imputed and observed genotypes: 0.982 for SNPs with MAF ≥ 0.05 and 0.998 for SNPs with MAF < 0.05.

Population structure was assessed using PCA of approximately 97K autosomal genotyped SNPs chosen for call rate (>95%), MAF (>0.05), and LD (pairwise $r^2 < 0.1$ across variants in a sliding window of 10Mb). Tests of genetic association between the first 20 PCs and all SNPs confirmed that PCs did not represent local variation at specific genetic loci. Based on the scree plot and joint distributions, four PCs are sufficient to capture population structure within the PITT sample.

The Penn State Sample—Participants sampled from 2006–2012 (IRB #32341) were genotyped on the Illumina Human Hp200c1 BeadChip (Illumina Inc., San Diego, CA). Participants sampled from 2013–2016 (IRB #44929, #13103, #2503, and #4320) were genotyped on the 23andMe v3 and v4 arrays (23andMe, Mountainview, CA). Samples were evaluated for concordance of genetic and reported sex, evidence of chromosomal aberrations, biological relatedness across study participants, ancestry, genotype call rate, and batch effects. SNPs were evaluated for call rate, discordant genotype calls between duplicate samples, Mendelian errors in HapMap control parent-offspring trios, deviation from Hardy-Weinberg genotype proportions, and sex differences in allele frequency and heterozygosity.

Using the 1000 Genomes Project[8] Phase 3 reference panel, samples with > 500K variants were imputed, as fewer variants result in uncertain imputation probabilities. SHAPEIT2[55] was used for pre-phasing haplotypes and IMPUTE2[56,57] was used to impute nearly 35M variants. SNP-level (INFO score > 0.5) and genotype per participant-level (genotype probability > 0.9) filters were used to omit poorly-imputed variants.

To select the Europeans used in this study, the HapMap 3 dataset (N = 998) was merged with the non-imputed genotypes, and genetic ancestry was estimated using the ADMIXTURE

program[58] assuming k=3 to 9 ancestral populations. Based on the cross-validation (CV) error for each k value, we selected a value of k=6 as appropriate for the dataset. These results are then used to provide a matrix of genetic ancestry axes, which were used to select samples most closely related to the CEU and TSI population references.

Additional population structure was assessed using PCA of approximately 100K autosomal genotyped SNPs chosen for call rate (>95%), MAF (>0.05), and LD (pairwise $r^2 < 0.1$ across variants in a sliding window of 10Mb). Tests of genetic association between the first 20 PCs and all SNPs confirmed that PCs did not represent local variation at specific genetic loci. Based on the scree plot and joint distributions, three PCs are sufficient to capture population structure within the European-only PSU sample.

Top SNPS and annotation—We observed 1932 genome-wide significant SNPs across 38 loci using a 500kb window. For each locus a top SNP was defined as the SNP generating the highest association (lowest p-value) in any of the 63 facial segments. Genes 500kb up- and downstream of the top SNPs were identified using the Table Browser of the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTables). Each gene was annotated based on the literature found in PubMed. As a search term, only the gene name was used, and relevant articles were selected based on their title and abstract. The OMIM database (https://www.omim.org) was searched for syndromes associated with the annotated genes.

Statistics and epigenomic analysis

SNP—statistical association—Our global-to-local facial phenotyping partitioned facial shape into 63 facial segments, each of which was represented by multiple dimensions of variation (PCs). Canonical correlation analysis (CCA)[13], as implemented in the function *canoncorr* from Matlab™ 2016b, was used as a straightforward multivariate testing framework (note that CCA is also implemented in PLINK 1.9[59]). CCA extracts the linear combination of PCs from a facial segment that has maximal correlation with the SNP being tested. The correlation is tested for significance based on Rao's F-test approximation[60] (right-tail, one-sided test). Using CCA, we tested each SNP (n=9,478,608) individually under the additive genetic model in the PITT cohort (n=2,329). Note that CCA does not accommodate adjustments for covariates, but effects of important variables such as sex, age, height, weight, facial size, and genetic ancestry were corrected for (using PLSR) at the phenotyping stage. Additionally, we applied a similar correction for the covariates on each SNP, again using PLSR. Therefore, the CCA analysis was performed under the reduced model, which was obtained after removing the effects of covariates on both the independent (SNP) as well as the dependent (facial shape) variables. The minimum MAF cut-off for SNPs to test was 1%. Quantile-quantile plots, provided online, indicate that population stratification under this reduced model was dealt with properly.

Multiple testing corrections—Given the burden of multiple comparisons, a strict significance threshold of $p \leq 5\times10^{-8}$ was used to declare "genome-wide significance", which corresponds to a Bonferroni correction for 1 million independent tests and mostly applicable in a European GWAS cohort[61]. Given that we tested facial variation as one of many facial segments separately, the multiple comparisons burden was magnified.

Therefore, we also determined a more stringent threshold for declaring "study-wide significance" corresponding to an additional adjustment for the effective number of independent tests[62]. The eigenvalues of pairwise multivariate correlations of all 63 modules determined a total of 39 effective independent tests. This reduction in effective tests was expected because of the hierarchical and overlapping construction of the facial segments. Therefore, the study-wide significance threshold was determined to be 1.2821×10^{-9} (i.e., $5\times10^{-8}/39$). Additionally, we computed a false discovery rate adjusted significance threshold of Benjamini & Yekutieli[14], that is accurate for any test dependency structure.

SNP—statistical replication—Based on previous work[10], the effect found in the discovery cohort was measured directly in the replication cohort (n=1,719), which was followed by a univariate analysis. First, the discovery cohort was used as a phenotyping reference for the replication cohort. Facial shapes of the replication cohort were projected onto the PCs of each of the 63 facial segments. Subsequently, they were further projected onto the SNP related CCA loadings, constructing specific genetic effect scores[10]. In doing so, the phenotypic trait, once discovered, was fixed and explicitly measured in the replication cohort. This resulted in a univariate phenotypic measure that was tested for association using a standard linear regression. We reported the regression coefficient, standard error and a p-value (two-sided) based on a t-statistic for regression coefficients using the function *regstats* from Matlab™2016b. This was done separately for each combination of SNP (n=1,931) and facial segment that achieved genome-wide significance. From all replication efforts combined (n=7,467), we computed a false discovery rate adjusted significance threshold[14].

SNP—tissue specific enhancer association—ChIP-seq datasets from human and chimpanzee CNCCs were obtained from Prescott et al.,[21] and compared to 72 publicly available H3K27ac ChIP-seq datasets corresponding to 31 cell types/tissues (as indicated in Figure 4A) and downloaded from the NIH SRA repository. Normalized H3K27ac ChIP coverage was calculated for 800,000 genomic regions corresponding to a superset of ENCODE DHS and transcription factor (TF) binding sites. Next, all intervals in +/−10kb window around 15 top replicated SNPs were selected (intersectBed) and a boxplot (using boxplot function in R with default interval settings) of the logarithm of normalized ChIP coverage was plotted. As control a 5× larger set of randomized SNP positions was used in the same procedure.

SNP—CNCC chromatin state association and statistics—Identification of different classes of CNCC regulatory regions based on chromatin signatures was performed as described in Prescott et al.[21]. In brief, 106,000 of Human CNCC candidate regulatory regions were called based on combined p300, *AP2α*, *NR2F1* ChIP-seq and ATAC-seq data by combining MACS2 called peaks into one set using mean-shift clustering. Normalized coverage around these regions was obtained for H3K4me1, H3K4me3, H3K27ac, and H3K27me3 histone modifications. To visualize chromatin states at the candidate regions the log of the H3K4me1 to H3K4me3 signal ratio was plotted vs. log of the ratio of H3K27ac to H3K27me3 as a scatterplot and colored by k-means (k=5) cluster identity, corresponding to promoters, active enhancers, poised regions and pre marked regions. Regulatory regions

closest to the top replicated GWAS (closestBed function) SNPs were identified and highlighted by circles. We observed 3 out of 15 lead SNPs to be closest to a region with active chromatin state (distance less than 9kb in each case). This represent significant enrichment for active chromatin-lead SNP association under Fisher's exact test (contingency table: 3 active regions proximal to SNP, 12 other regions proximal to SNP, 3622 active regions not proximal to SNP, 102694 other regions not proximal to SNP), exact $p = 1.19 \times 10^{-2}$. We also compared replicated SNPs to non-replicated SNPs from the screen in terms of direct overlap (1kb distance) with CNCC regulatory elements (contingency table: 26 non-replicated SNPs with no overlap, 0 non-replicated SNPs with overlap, 7 replicated SNPs with no overlap, 5 replicated SNPs with overlap), Fisher's exact test $p = 1.58 \times 10^{-3}$.

Chromatin modifications visualization in the genome browser—Kernel smoothed coverage files in wig format were generated with QuEST2.0[63] from datasets from Prescott et al.[21] and uploaded to UCSC browser. Representative tracks were selected to make figure 5B concise and instructive.

Gene Ontology term enrichment analysis—was performed with GREAT 3.0.0[15]. Genomic positions for replicated SNPs were uploaded to great.stanford.edu and results of the analysis in terms of FDR and enrichments for each significant category was summarized in supplementary figure 4.

Cell line sources WiCell (H9 ESC) and Fred Gage laboratory (iPSC) were authenticated by analyzing genomic sequences data from the lines. PCR tests were done to test for mycoplasma contamination.

URLs

https://www.esat.kuleuven.be/psi/research/global-to-local-facial-phenotyping (data)

www.great.stanford.edu (results for GREAT analysis)

http://www.ncbi.nlm.nih.gov/gap (genotype data PITT cohort)

www.facebase.org (facial images PITT cohort)

www.gwascentral.org (GWAS central)

http://genome.ucsc.edu/cgi-bin/hgTables (the Table Browser of the UCSC Genome Browser)

https://www.omim.org (Omim database)

https://github.com/TheWebMonks/meshmonk. (facial mapping software)

Ethics Statement

Institutional review board (IRB) approval was obtained at each recruitment site and all participants gave their written informed consent prior to participation; for children, written consent was obtained from a parent or legal guardian. For the Pittsburgh sample the following local ethics approvals were obtained: University of Pittsburgh IRB

#PRO09060553 and #RB0405013; UT Health Committee for the Protection of Human Subjects #HSC-DB-09-0508; Seattle Children's IRB #12107; University of Iowa Human Subjects Office/IRB #200912764 and #200710721. For the Penn State sample, the following local ethics approvals were obtained: State College, PA (IRB #44929 & #4320 New York, NY (IRB #45727); Urbana-Champaign, IL (IRB #13103); Dublin, Ireland; Rome, Italy; Warsaw, Poland; and Porto, Portugal (IRB #32341); and Twinsburg, OH (IRB #2503)

Data and Code availability statements

All of the genotypic markers for the Pittsburgh Dataset are available to the research community through the dbGaP controlled access repository (http://www.ncbi.nlm.nih.gov/gap) at accession number: phs000949.v1.p1. The raw source data for the phenotypes – the 3D facial surface models in .obj format – are available for through the FaceBase Consortium (www.facebase.org). 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 comprising the Penn State University dataset 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 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.

The KU Leuven provides the spatially-dense facial mapping software, free to use for academic purposes. The software referred to as MeshMonk is hosted at https://github.com/TheWebMonks/meshmonk. Matlab™ implementations of the hierarchical spectral clustering to obtain facial segmentations are available upon request from P.C. The statistical analyses in this work were based on functions of the statistical toolbox in Matlab™ 2016b as mentioned throughout the materials and methods.

All relevant data to run future replications and meta-analyses efforts is provided in Matlab™ format online at https://www.esat.kuleuven.be/psi/research/global-to-local-facial-phenotyping. This includes the anthropometric mask used, facial segmentation cluster-labels, PCA shape-spaces for all 63 facial segments in the PITT cohort, CCA loadings and all association statistics for the peak SNPs in Table 1. An example Matlab™ script to explore the data is provided as well.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This investigation was supported by the KU Leuven, BOF funds GOA, CREA, & C1. The collaborators at the University of Pittsburgh were supported by The National Institute for Dental and Craniofacial Research (http://www.nidcr.nih.gov/) through the following grants: U01-DE020078; U01-DE020057; R01-DE016148; K99-DE02560, 1-R01-DE027023. Funding for genotyping was provided by the National Human Genome Research

Institute (https://www.genome.gov/): X01-HG007821 and X01-HG007485. Funding for initial genomic data cleaning by the University of Washington was provided by contract #HHSN268201200008I from the National Institute for Dental and Craniofacial Research (http://www.nidcr.nih.gov/) awarded to the Center for Inherited Disease Research (CIDR). The collaborators at the Penn State University were supported in part by grants from the Center for Human Evolution and Development at Penn State, the Science Foundation of Ireland Walton Fellowship (04.W4/B643), the United States National Institute Justice (www.nij.gov; 2008-DN-BX-K125), and by the United States Department of Defense (www.defense.gov). The collaborators at Stanford University School of Medicine were supported by Howard Hughes Medical Institute, NIH U01 DE024430 and the March of Dimes Foundation 1-FY15-312 (J.W.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Roosenboom J, Hens G, Mattern BC, Shriver MD, Claes P. Exploring the Underlying Genetics of Craniofacial Morphology through Various Sources of Knowledge. BioMed Research International. 2016; 2016
2. Paternoster L, et al. Genome-wide Association Study of Three-Dimensional Facial Morphology Identifies a Variant in *PAX3* Associated with Nasion Position. The American Journal of Human Genetics. 2012; 90:478–485. [PubMed: 22341974]
3. Liu F, et al. A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS genetics. 2012; 8:e1002932. [PubMed: 23028347]
4. Adhikari K, et al. A genome-wide association scan implicates *DCHS2, RUNX2, GLI3, PAX1* and *EDAR* in human facial variation. Nature communications. 2016; 7
5. Cole JB, et al. Genomewide association study of African children identifies association of *SCHIP1* and *PDE8A* with facial size and shape. PLoS Genet. 2016; 12:e1006174. [PubMed: 27560698]
6. Shaffer JR, et al. Genome-wide association study reveals multiple loci influencing normal human facial morphology. PLoS Genet. 2016; 12:e1006149. [PubMed: 27560520]
7. Lee MK, et al. Genome-wide association study of facial morphology reveals novel associations with *FREM1* and *PARK2*. PLoS One. 2017; 12
8. Consortium, G.P. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
9. Snyders J, Claes P, Vandermeulen D, Suetens P. Development and comparison of non-rigid surface registration and extensions. Technical Report: KUL/ESAT/PSI/1401. 2014:1–55.
10. Claes P, et al. Modeling 3D facial shape from DNA. Plos Genetics. 2014; 10:e1004224. [PubMed: 24651127]
11. Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. NIPS. 2001; 14:849–856.
12. Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organizational research methods. 2004; 7:191–205.
13. Thompson B. Canonical correlation analysis. Encyclopedia of statistics in behavioral science. 2005
14. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Annals of statistics. 2001:1165–1188.
15. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology. 2010; 28:495–501.
16. Bronner ME, LeDouarin NM. Development and evolution of the neural crest: an overview. Developmental biology. 2012; 366:2–9. [PubMed: 22230617]
17. Green SA, Simoes-Costa M, Bronner ME. Evolution of vertebrates as viewed from the crest. Nature. 2015; 520:474–482. [PubMed: 25903629]
18. Schneider RA. Chapter Eleven-Regulation of Jaw Length During Development, Disease, and Evolution. Current topics in developmental biology. 2015; 115:271–298. [PubMed: 26589929]
19. Cordero DR, et al. Cranial neural crest cells on the move: their roles in craniofacial development. American journal of medical genetics Part A. 2011; 155:270–279.
20. Kulesa PM, McLennan R. Neural crest migration: trailblazing ahead. F1000prime reports. 2015; 7
21. Prescott SL, et al. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell. 2015; 163:68–83. [PubMed: 26365491]

22. Rada-Iglesias A, et al. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell stem cell. 2012; 11:633–648. [PubMed: 22981823]
23. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. [PubMed: 21160473]
24. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences. 2010; 107:21931–21936.
25. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics & chromatin. 2015; 8:57. [PubMed: 26719772]
26. Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell. 2016; 167:1170–1187. [PubMed: 27863239]
27. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. Genome medicine. 2014; 6:85. [PubMed: 25473424]
28. Ludwig KU, et al. Imputation of Orofacial Clefting Data Identifies Novel Risk Loci and Sheds Light on the Genetic Background of Cleft Lip±Cleft Palate and Cleft Palate Only. Human Molecular Genetics. 2017 ddx012.
29. Staley JR, et al. PhenoScanner: a database of human genotype–phenotype associations. Bioinformatics. 2016; 32:3207–3209. [PubMed: 27318201]
30. Pickrell JK, et al. Detection and interpretation of shared genetic influences on 42 human traits. Nature genetics. 2016
31. Klingenberg CP. Quantitative genetics of geometric shape: heritability and the pitfalls of the univariate approach. Evolution. 2003; 57:191–195. [PubMed: 12643582]
32. Pallares LF, et al. Mapping of craniofacial traits in outbred mice identifies major developmental genes involved in shape determination. PLoS Genet. 2015; 11:e1005607. [PubMed: 26523602]
33. Maga AM, Navarro N, Cunningham ML, Cox TC. Quantitative trait loci affecting the 3D skull shape and size in mouse and prioritization of candidate genes in-silico. Frontiers in physiology. 2015; 6:92. [PubMed: 25859222]
34. Coussens AK, Van Daal A. Linkage disequilibrium analysis identifies an FGFR1 haplotype-tag SNP associated with normal variation in craniofacial shape. Genomics. 2005; 85:563–573. [PubMed: 15820308]
35. Peng S, et al. Detecting genetic association of common human facial morphological variation using high density 3D image registration. PLoS Comput Biol. 2013; 9:e1003375. [PubMed: 24339768]
36. Carlson, DS. Seminars in Orthodontics. Vol. 11. Elsevier; 2005. Theories of craniofacial growth in the postgenomic era; p. 172-183.
37. Sperber, GH., Guttmann, GD., Sperber, SM. Craniofacial Development (Book for Windows & Macintosh). PMPH-USA; 2001.
38. Williams SE, Slice DE. Regional shape change in adult facial bone curvature with age. American journal of physical anthropology. 2010; 143:437–447. [PubMed: 20949614]
39. Mitteroecker P, Bookstein F. The evolutionary role of modularity and integration in the hominoid cranium. Evolution. 2008; 62:943–958. [PubMed: 18194472]
40. Klingenberg CP. Morphological integration and developmental modularity. Annual review of ecology, evolution, and systematics. 2008; 39:115–132.
41. Wood AR, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature genetics. 2014; 46:1173–1186. [PubMed: 25282103]
42. Zarelli VE, Dawid IB. Inhibition of neural crest formation by Kctd15 involves regulation of transcription factor AP-2. Proceedings of the National Academy of Sciences. 2013; 110:2870–2875.
43. Ahituv N, et al. An ENU-induced mutation in *AP-2α* leads to middle earand ocular defects in Doarad mice. Mammalian genome. 2004; 15:424–432. [PubMed: 15181535]
44. Lee YH, Saint-Jeannet JP. *SOX9* function in craniofacial development and disease. Genesis. 2011; 49:200–208. [PubMed: 21309066]
45. Francis-West, PH., Robson, L., Evans, DJ. Craniofacial Development The Tissue and Molecular Interactions That Control Development of the Head. Springer Science & Business Media; 2012.

46. Stearns FW. One hundred years of pleiotropy: a retrospective. Genetics. 2010; 186:767–773. [PubMed: 21062962]
47. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017; 169:1177–1186. [PubMed: 28622505]
48. Weinberg SM, et al. The 3D Facial Norms Database: Part 1. A web-based craniofacial anthropometric and image repository for the clinical and research community. The Cleft Palate-Craniofacial Journal. 2016; 53:e185–e197. [PubMed: 26492185]
49. Heike CL, Upson K, Stuhaug E, Weinberg SM. 3D digital stereophotogrammetry: a practical guide to facial image acquisition. Head & face medicine. 2010; 6:18. [PubMed: 20667081]
50. Claes P, Walters M, Clement JG. Improved Facial Outcome Assessment using a 3D Anthropometric Mask. International journal of oral and maxillofacial surgery. 2012; 41:324–330. [PubMed: 22103995]
51. Claes P, Walters M, Vandermeulen D, Clement JG. Spatially-dense 3D facial asymmetry assessment in both typical and disordered growth. Journal of Anatomy. 2011; 219:444–455. [PubMed: 21740426]
52. Rohlf F, Slice D. Extensions of the procrustus method for the optimal superimposition of landmarks. Systematic Zoology. 1990; 39:40–59.
53. Robert P, Escoufier Y. A unifying tool for linear multivariate statistical methods: the RV-coefficient. Applied statistics. 1976:257–265.
54. Laurie CC, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genetic epidemiology. 2010; 34:591–602. [PubMed: 20718045]
55. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nature methods. 2013; 10:5–6. [PubMed: 23269371]
56. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]
57. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3: Genes, Genomes, Genetics. 2011; 1:457–470. [PubMed: 22384356]
58. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome research. 2009; 19:1655–1664. [PubMed: 19648217]
59. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 2007; 81:559–575. [PubMed: 17701901]
60. Olson CL. On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin. 1976; 83:579.
61. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic epidemiology. 2008; 32:381–385. [PubMed: 18348202]
62. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 2005; 95:221–227. [PubMed: 16077740]
63. Valouev A, et al. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nature methods. 2008; 5:829–834. [PubMed: 19160518]

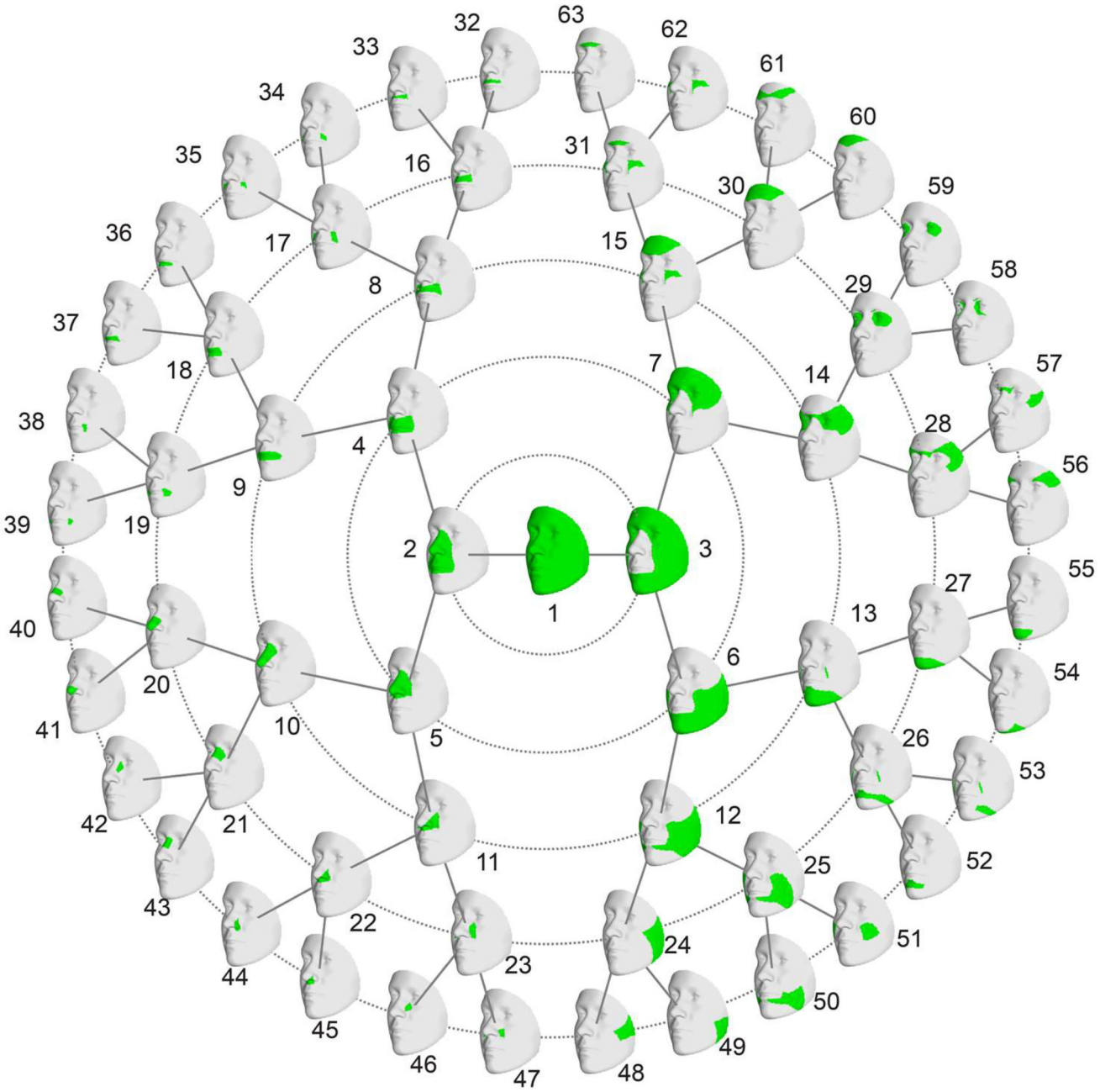

Figure 1.

Hierarchical spectral clustering of facial shape: Global-to-local facial segmentation of the PITT cohort (n=2,329) obtained using hierarchical spectral clustering. Segments are colored in green. Segment 1, representing the entire face and shown in the center, is partitioned into segments 2 and 3, shown in the inner-most concentric circle. Moving outward across the five concentric circles, corresponding to the five hierarchical levels, each facial segment is further partitioned into two more localized segments.

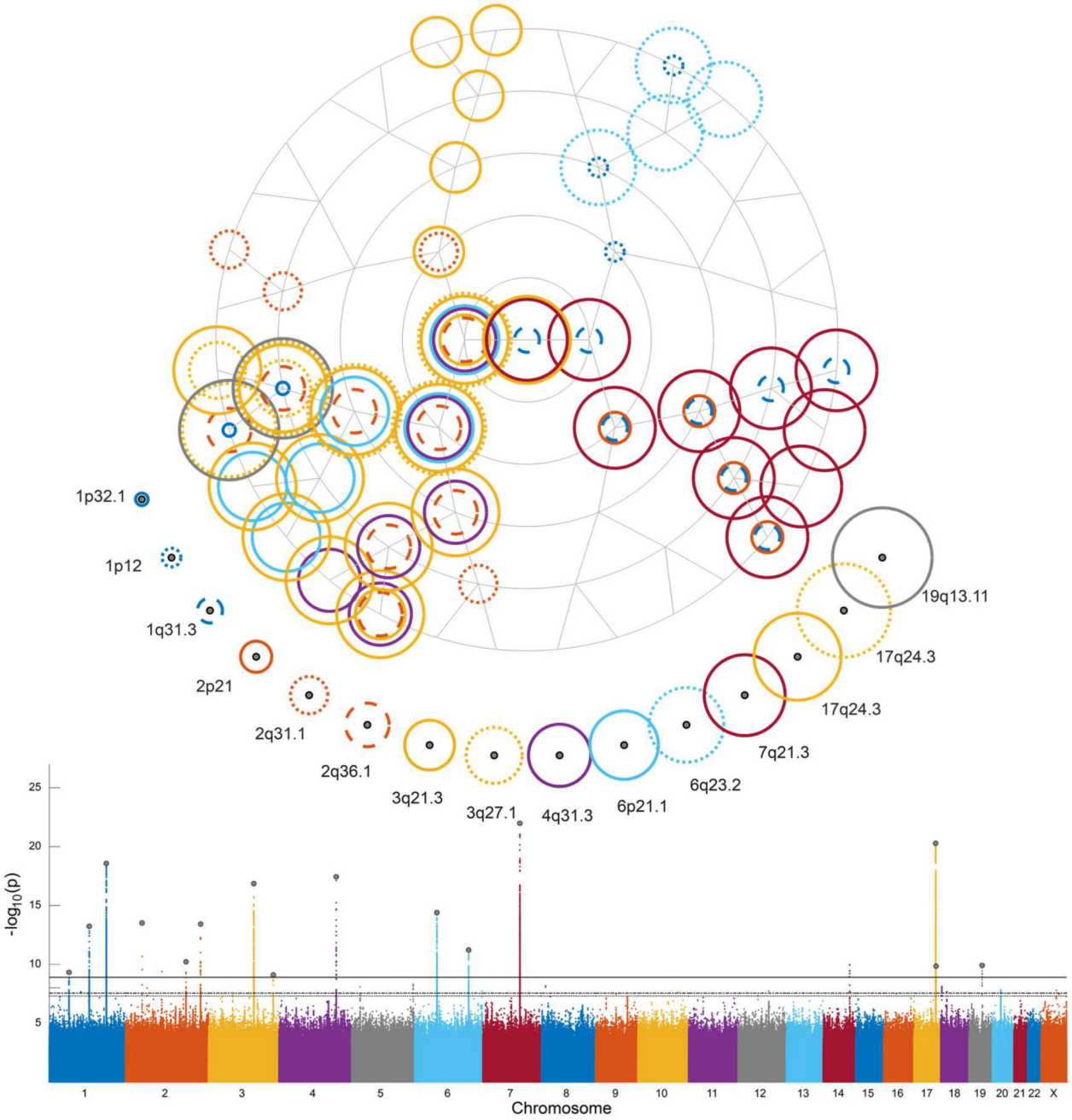

Figure 2.

15 replicating loci: Top, a polar dendrogram corresponding one-to-one with facial segments shown in Figure 1; concentric circles indicate loci reaching genome-wide significance for each segment. Bottom, Manhattan plot of all 63 facial segments combined illustrating the chromosomal position of the loci. The bottom (dotted) horizontal line represents the genome-wide threshold ($p = 5 \times 10^{-8}$), the middle (stripe dotted) horizontal line, represents the false discovery rate threshold ($p = 2.82 \times 10^{-8}$) and the top (full) horizontal line represents study-wide significance ($p = 1.28 \times 10^{-9}$). The correspondence of the 15 loci is indicated by color, ordered by chromosome number, left to right.

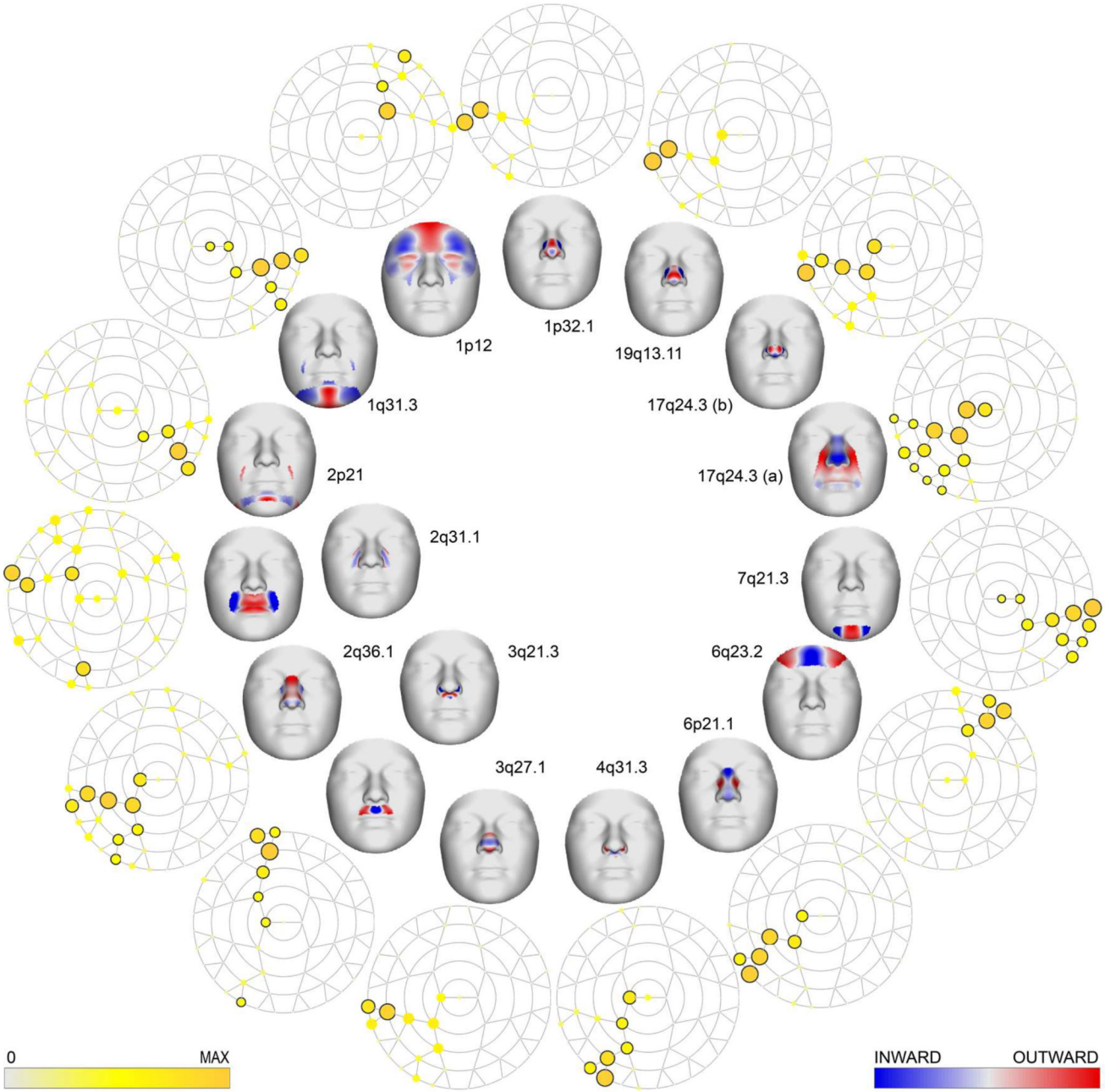

Figure 3.

Facial shape effects. Outer layer: For each locus, the $-\log_{10}$ p-value of the lead SNP is shown for all 63 facial segments plotted on a polar dendrogram (ranging from 0 to the maximum $-\log_{10}$ p-value as reported in Table 1). Facial segments with a p-value lower than the genome-wide threshold ($p = 5 \times 10^{-8}$) are encircled. Inner layer: For each locus, its effect on a representative segment is illustrated using the normal displacement (displacement in the direction locally perpendicular to the facial surface). Blue/red indicates a local shape depression/protrusion due to the minor allele SNP variant. Two loci (2q31.1 and 3q21.3) were associated with segments from separate quadrants of the polar dendrogram, hence,

normal displacement is illustrated for representative segments in two areas of the face. Loci are ordered counter-clockwise by chromosomal location.

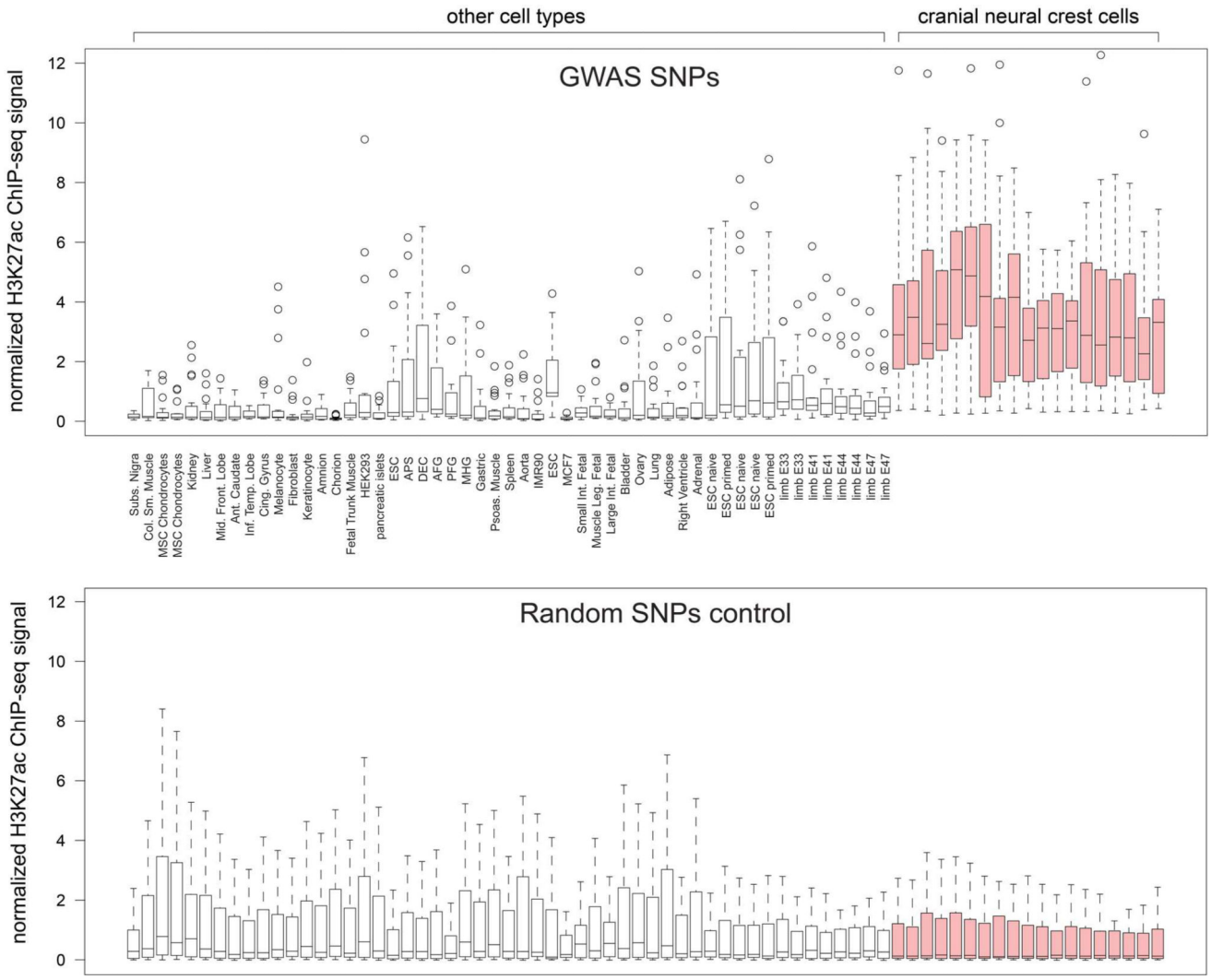

Figure 4.

Preferential activity in CNCCs: Facial variation SNPs are associated with signatures of active regulatory regions in human cranial neural crest cells. Peak replicated SNPs are located in regions preferentially marked by H3K27ac in CNCCs. Plotted is boxplot of normalized distribution of H3K27ac ChIP-seq signals in regions immediately surrounding (within 10kb) each of the 15 top replicated SNPs (or, as a control, randomized SNP positions, lower panel) in various adult, embryonic and *in vitro* derived cell types, as indicated. Samples corresponding to *in vitro* derived CNCCs are highlighted in red and separated from other cell types. Boxplot central line represents median, hinges represent 1st and 3rd quartile, whiskers are 1.5 interquartile distance.

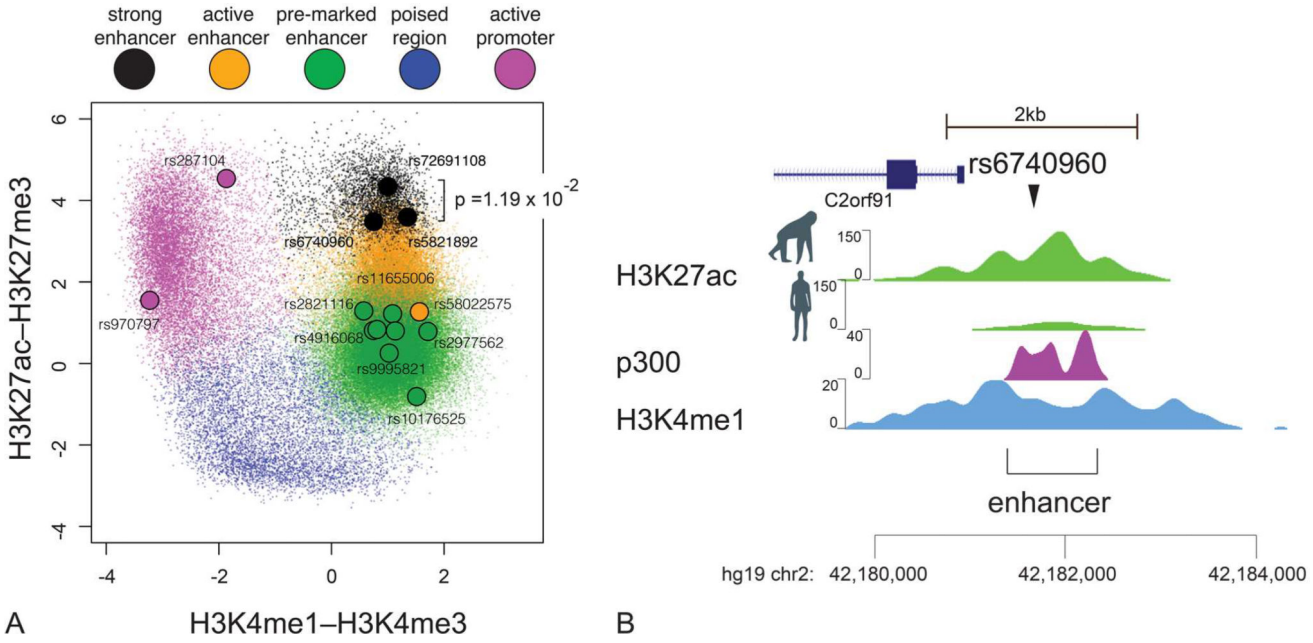

Figure 5.

Regulatory regions active in CNCCs. (A) Top replicated SNPs are enriched in vicinity of active CNCC enhancer regions. Plotted is promoter/enhancer score, corresponding to the ratio H3K4me1 to H3K4me3 ChIP-seq signal (ordinate) vs. activity score, corresponding to the ratio of H3K27ac to H3K27me3 ChIP-seq signal (abscissa). Small dots represent all putative regulatory regions identified in CNCCs, large dots highlight chromatin regions closest to each of the significant SNPs. Colors represent classes and activity states of regulatory elements, black: strong active enhancers, orange: weakly active enhancers, green: primed enhancers, magenta: promoters, blue: poised/repressed regulatory regions. Significant diagnostic SNPs are enriched in strong active enhancers, ($p = 1.19 \times 10^{-2}$, odds ratio 7.7, Fisher's exact test). In addition, SNPs that retested with combined FDR< 0.001 are more likely to be close (< 1kb) to a detectable regulatory chromatin feature in CNCCs ($p = 1.58 \times 10^{-3}$, Fisher's exact test). (B) A significant SNP rs6740960:A>T associated with lower face variation is located within an active CNCC enhancer. Plotted are ChIP-seq tracks for H3K27ac (green), p300 acetyltransferase (purple) and H3K4me1 (blue). Human and chimp H3K27ac tracks are indicated with silhouette icons. Position of rs6740960:A>T SNP is indicated with arrowhead and of the enhancer with a bracket. Track height are normalized enrichment scores from QuEST 2.0.

Nat Genet. Author manuscript; available in PMC 2018 September 01.

Table 1

Genome-wide significant loci and replication results: Properties of top SNPs in 18 chromosomal regions showing genome-wide significance.

Annotation					Effect		Discovery					Replication			
Chromosomal Position	**SNP**	**Candidate Gene**	**Alleles**	**Lit. Int.**	**Quadrant**	**Module**	**CC**	**CC2**	**p-value**	**MAF**	**Nr SNP**	**B**	**p-value**	**MAF**	**Nr SNP**
1p32.1 (60997570)	rs4916068	intergenic	C < T		2	41	0.16	0.03	4.81E-10	0.500	28	0.25	1.39E-12	0.494	28
1p12 (119762175)	rs72691108	*TBX15*	A < G	TS5	4	7	0.25	0.06	5.81E-14	0.236	224	0.34	1.01E-15	0.243	224
1q31.3 (197329041)	rs2821116	***ASPM***	A < T	TS5*	3	13	0.24	0.06	2.63E-19	0.210	342	0.32	2.29E-15	0.209	342
2p21 (42181679)	rs6740960	*PKDCC*	T < A	TS4*,TS5*	3	26	0.21	0.05	3.03E-14	0.497	2	0.17	3.44E-06	0.499	2
2q31.1 (177111819)	rs970797	*HOXD* cluster	T < G	TS4,TS5	2,3	38	0.17	0.03	6.17E-11	0.434	15	0.11	1.12E-03	0.428	4
2q36.1 (223039052)	rs10176525	*PAX3*	A < T	TS3*, TS4*,TS5*	2	10	0.21	0.05	3.76E-14	0.224	38	0.26	4.39E-11	0.233	38
3q21.3 (128106267)	rs2977562	***RAB7A, ACAD9***	G < A	TS4*	2,3	16	0.22	0.05	1.39E-17	0.244	389	0.20	8.63E-07	0.248	374
3q27.1 (184333169)	rs58022575	*EPHB3, DVL3*	G < GA	TS4*,TS5*	2	20	0.17	0.03	7.99E-10	0.473	16	0.21	2.39E-09	0.467	16
4q31.3 (154828366)	rs9995821	*DCHS2*	C < T	TS3*,TS4*	2	45	0.22	0.05	3.67E-18	0.200	10	0.39	7.33E-19	0.204	10
6p21.1 (44681840)	rs227833	*SUPT3H*	C < G	TS3*,TS4*, TS5*	2	10	0.21	0.05	3.38E-14	0.261	164	0.14	5.63E-04	0.254	150
6q23.2 (133615646)	rs5880172	***RPS12, EYA4***	A < ACT		4	30	0.21	0.04	6.10E-12	0.312	33	0.26	6.91E-13	0.297	33
7q21.3 (96124975)	rs10238953	***DLX6, DYNC1L1***	G < A	TS5*	3	55	0.24	0.06	1.06E-22	0.140	256	0.55	1.29E-26	0.137	256
17q24.3 (69139583)	rs11655006	***BC039327***	G < A	TS5*	2	2	0.27	0.07	5.24E-21	0.376	311	0.15	1.77E-05	0.364	308
17q24.3 (70036479)	rs5821892	*SOX9*	C < CG	TS5*	2	10	0.19	0.03	2.30E-09	0.427	31	0.13	3.01E-04	0.453	26
19q13.11 (34290995)	rs287104	***KCTD15***	G < A		2	41	0.17	0.03	1.26E-10	0.349	8	0.16	2.86E-05	0.329	8
5p14.3 (20558014)	rs62354288	***CDH18***	T < A		2	21	0.17	0.03	8.05E-09	0.355	4	0.02	7.04E-01	0.357	0
5q35.3 (177760576)	rs2913791	***NHP2, ZNF354A***	A < G		3	55	0.16	0.03	4.96E-09	0.429	5	0.00	9.14E-01	0.417	0
20p11.22 (21626627)	rs2424390	*PAX1*	A < T	TS3*,TS4*, TS4*	2	11	0.17	0.03	1.50E-08	0.233	15	0.09	2.47E-02	0.233	0

SNP, single nucleotide polymorphism, CC, canonical correlation, CC2 canonical correlation squared, B regression coefficient, MAF, minor allele frequency, Nr SNPs (Discovery), the number of SNPs reaching genome-wide significance within the same locus. Intragenic SNP candidate genes are indicated in bold. Nr SNPs (Replication), the number of SNPs reaching FDRd replication significance within the same locus. Lit. Int., Literature Integration providing a link to supplementary tables reported in the manuscript, * is when the associated trait as described in the literature involved the facial segments identified, e.g. nose protrusion – nose segments. Additional details, such as the degrees of freedom for CC, and the standard error for B, can be found in Table S1.