Original research

A genome-wide association study of facial morphology identifies novel genetic loci in Han Chinese

Yin Huang [a,1], Dan Li [a,b,1], Lu Qiao [a], Yu Liu [a], Qianqian Peng [a], Sijie Wu [a,c], Manfei Zhang [a,c], Yajun Yang [c,d], Jingze Tan [c,d], Shuhua Xu [a,e,f,g], Li Jin [a,c,d,e], Sijia Wang [a,e,f,*], Kun Tang [a,b,*], Stefan Grünewald [a,*]

[a] Chinese Academy of Sciences Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, CAS, Shanghai 200031, China
[b] DeepBlue Technology (Shanghai) Co., Ltd, Shanghai 200336, China
[c] State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China
[d] Fudan-Taizhou Institute of Health Sciences, Taizhou 225300, China
[e] Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China
[f] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[g] School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

ARTICLE INFO

ABSTRACT

The human face is a heritable surface with many complex sensory organs. In recent years, many genetic loci associated with facial features have been reported in different populations, yet there is a lack of studies on the Han Chinese population. Here, we report a genome-wide association study of 3D normal human faces of 2,659 Han Chinese with autosegment phenotypes of facial morphology. We identify single-nucleotide polymorphisms (SNPs) encompassing four genomic regions showing significant associations with different facial regions, including SNPs in *DENND1B* associated with the chin, SNPs among *PISRT1* associated with eyes, SNPs between *DCHS2* and *SFRP2* associated with the nose, and SNPs in *VPS13B* associated with the nose. We replicate 24 SNPs from previously reported genetic loci in different populations, whose candidate genes are *DCHS2*, *SUPT3H*, *HOXD1*, *SOX9*, *PAX3*, and *EDAR*. These results provide a more comprehensive understanding of the genetic basis of variation in human facial morphology.
Copyright © 2020, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, and Genetics Society of China. Published by Elsevier Limited and Science Press. All rights reserved.

Introduction

The human face encompasses complex sensory organs and shows remarkable variations in facial traits. These complex and delicate traits undergo precise genetic regulation during facial development to produce our highly inherited, unique, and recognizable faces, which play an essential role in communication and physical identification. Although the genetic basis of human facial morphology is still limited, many genetic variants associated with facial features have been reported in recent years.

In 2012, two studies (Liu et al., 2012; Paternoster et al., 2012) respectively found an association between a genetic variant in *PAX3* with nose shape in European participants, which was the start of the genome-wide association study (GWAS) of normal human facial morphology. Furthermore, Liu et al. (2012) identified four novel genetic variants close to the genes *PRDM16*, *TP63*, *C5orf50*, and *COL17A1* affecting face morphology. More recently, a GWAS on Latin Americans reported five more genetic variants close to *DCHS2*, *RUNX2*, *GLI3*, *PAX1*, and *EDAR* related to nose shape (Adhikari et al., 2016b). Another recent GWAS (Shaffer et al., 2016) identified seven additional genetic variants for face traits such as facial width and depth and nose shape in a European-derived cohort. Interestingly, this study observed the same association between soft-tissue nasal width and *PAX1* identified in Latin Americans (Adhikari et al., 2016b). Lee et al. (2017) identified two genetic variants of *FREM1* and *PARK2* associated with face shape in individuals of European ancestry. In

2018, two studies (Claes et al., 2018; Crouch et al., 2018) found a number of newly associated genetic loci by applying novel approaches to a European-derived cohort and replicated many loci in previous studies (Liu et al., 2012; Paternoster et al., 2012; Shaffer et al., 2016; Adhikari et al., 2016b). In African participants, *SCHIP1* and *PDE8A* were identified to be associated with measures of human facial size (Cole et al., 2017). In Asians, a series of genome-wide association analyses of normal facial morphology in Uyghurs identified six significant loci (Qiao et al., 2018), and another study on Koreans identified five novel face morphology loci that were associated with facial frontal contour, nose shape, and eye shape (Cha et al., 2018). Recently, integrated analyses of multiple populations have been conducted to identify novel genetic loci affecting human facial shapes (Xiong et al., 2019).

So far, there have been some GWASs carried out on normal human facial morphology in different populations, including Europeans (Liu et al., 2012; Paternoster et al., 2012; Shaffer et al., 2016; Lee et al., 2017; Claes et al., 2018; Crouch et al., 2018; Xiong et al., 2019), Latin Americans (Adhikari et al., 2016b; Xiong et al., 2019), Africans (Cole et al., 2016), and Asians (Cha et al., 2018; Qiao et al., 2018; Xiong et al., 2019). However, studies on genome-wide associations of three-dimensional (3D) normal human facial morphology in Han Chinese are still lacking. This work is aimed to fill this gap and to find genetic variants associated with facial traits in the Han Chinese population. In addition, we try to innovate methods to define and visualize complex phenotypes. Here, we conduct a genome-wide association analysis on 10 different facial regions, which are automatically and systematically divided from 3D high-density faces of 2,659 unrelated participants of Han Chinese from Taizhou, Jiangsu province. We identify four genomic regions (one reported locus and three novel loci) associated with different facial phenotypes and visualize the effect on the associated facial region. This study expands the sample diversity in the field of facial genetics, provides a novel method of data-driven phenotyping inspired by a recent study (Claes et al., 2018), and replicates previously reported variants to further increase our understanding of the genetic basis of human facial morphology.

Results

Sample and phenotyping features

The study included 2,659 unrelated participants of Han Chinese with basic demographic descriptions and physiological characteristics (such as gender, age, and body mass index [BMI]), who were recruited from Taizhou, Jiangsu province, on the east coast of China. These participants had a mean age of 56.7 years, with an average BMI of 24.4 (Fig. S1), 36% of whom were men and 64% of whom were women. A high-resolution 3D facial image was collected for each participant and automatically registered into a tri-mesh with 32,251 points via an in-house landmark recognition method.

The facial morphologic phenotypes were represented by the principal components (PCs) of facial vertices, which were extracted from high-density grid meshes with 32,251 vertices. The high-density face was automatically segmented into different regions (such as eyes, nose, and mouth) by using a data-driven approach (Fig. 1). In brief, we used partial least squares regression to remove confounders, such as gender, age, and BMI, from facial morphology and calculated the 3D correlation between each pair of vertices on high-density faces. Then, the unsupervised clustering method K-means was applied to the top 10 eigenvectors of the affinity matrix to partition the vertices into 10 facial regions (see Method). The vertices within each region are clustered together, which is consistent with our facial anatomical structure. Finally, for each region, we used principal component analysis (PCA) with parallel analysis to

determine the number of significant components contributing to the variation within that region (Table S1).

Genetic association analysis and follow-up study

We performed a multivariable analysis of genotype-phenotype association for the PCs of each facial region on the whole sample to keep the maximal statistical power because the participants were all from the same region with the same genetic background. To control the false-positive results, we used the two-step iterative resampling (TSIR) method (Kang et al., 2015) to verify the one-stage GWAS results. A total of 4,757,938 SNPs were tested using the multivariate statistical model, canonical correlation analysis (CCA), after removing gender, age, and BMI as covariates under an additive genetic model. We found 798 SNPs among 7 genetic loci that reached the threshold of $P < 5 \times 10^{-8}$ (Table S2). The associated facial regions included eyes, upper cheeks, the chin, and the nose. The results of the GWAS for each facial region are shown in QQ plots and Manhattan plots, with $-\log_{10}(P\text{-value})$ against the chromosome position (Fig. S2). To correct for the multiple test burden associated with the separated genome-wide association analysis of 10 facial regions, we used Bonferroni correction as a more conservative significance threshold (5×10^{-9}). Four loci with 406 SNPs remained after the Bonferroni correction and TSIR verification. The most significant SNP in each locus is shown in Table 1, which includes rs4915551 at 1q31 associated with the chin, rs12633011 at 3q23 associated with eyes, and rs2045323 at 4q31.3 and rs11988731 at 8q22.2 associated with the nose (Fig. 2).

The lead SNP rs4915551 ($P = 1.37 \times 10^{-12}$) was located in the intron region of the *DENND1B* gene. The SNPs at 1q31 associated with the chin show extensive linkage disequilibrium (LD) and cover the whole *DENND1B* gene (Fig. 2A). To show the effect on the face by different alleles, we compared the changes in the volume under each triangle of average faces, which were separated by rs4915551 under GG and AA genotypes. The result showed that participants with GG genotype at rs4915551 tended to have a more protruding chin than those with AA genotype (Fig. 3). The lead SNP rs12633011 ($P = 1.77 \times 10^{-14}$) associated with eyes was located within an LD block including the LncRNA gene *PISRT1* on chromosome 3q23 (Fig. 2B). It seemed that participants with the minor allele of rs12633011 were more likely to show sunken eyes with high brow bones (Fig. 3). Another two loci were associated with the nose. The lead SNP rs2045323 ($P = 1.77 \times 10^{-17}$) is an intergenic SNP between *DCHS2* and *SFRP2* on chromosome 4 q31.3 (Fig. 2C). The lead SNP rs11988731 ($P = 6.44 \times 10^{-11}$) was located in the intron region of the *VPS13B* gene, and SNPs at 8q22.2 show extensive LD and overlap with the *VPS13B* gene (Fig. 2D). Further analysis revealed that both of them appeared to affect the shape of nostrils (narrowness and length) by visualizing the effect on the heat plot. We observed that participants with AA genotype at rs2045323 tended to have narrower but longer nostrils than those with GG genotype, while participants with CC genotype at rs11988731 tended to have narrower and shorter nostrils than those with TT genotype (Fig. 3).

Replication analysis of reported genetic loci associated with facial morphology

We selected 56 SNPs (Table S3) previously reported in the GWAS of facial morphology and performed replication analysis. Sixteen SNPs were not included in our genomic data. We used LDlink (Machiela and Chanock, 2015) to extract the proxy SNPs of these 16 SNPs, with R^2 higher than 0.5. By screening out low-frequency SNPs (minor allele frequency [MAF] < 0.01) and unqualified SNPs for GWASs, we found 11 proxy SNPs corresponding to the highest LD of the original SNPs. We then examined the association between the
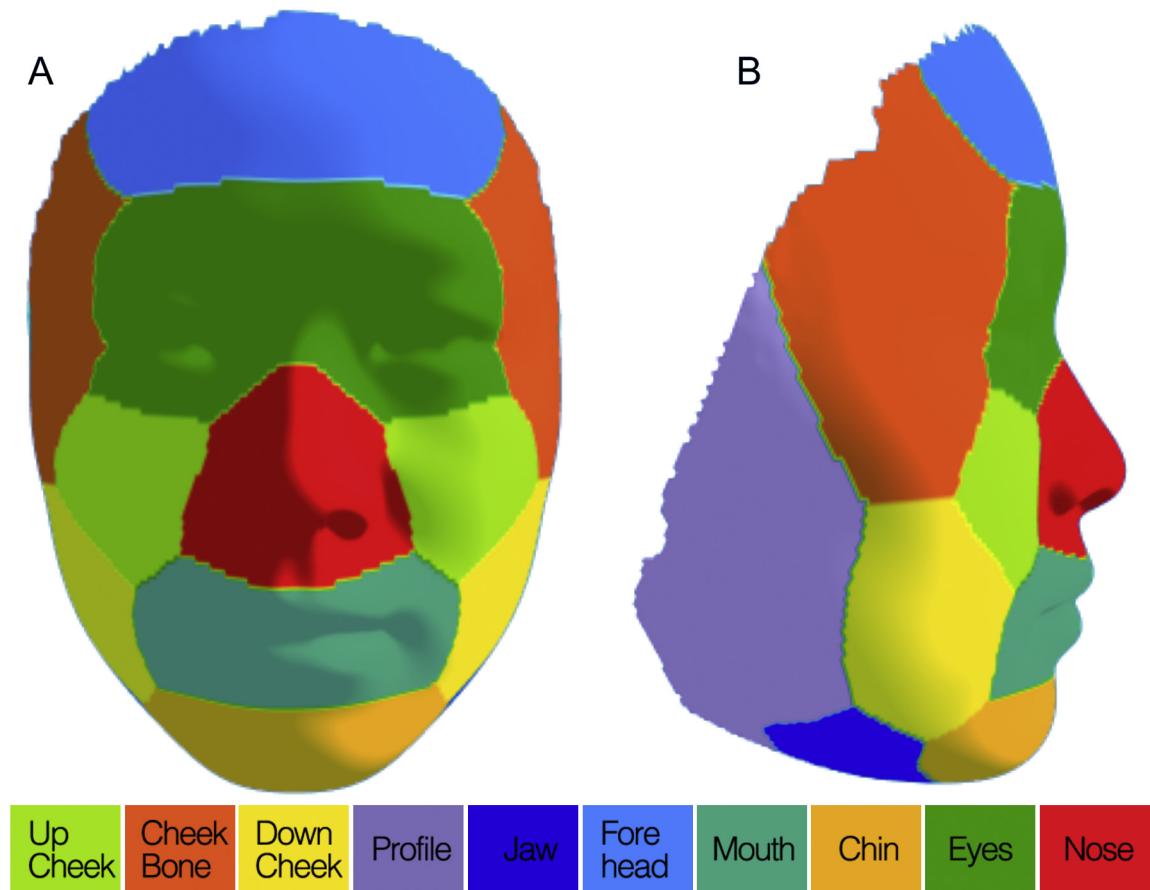
Fig. 1. Automatically segmented phenotypes of facial morphology. **A**: The front of the reference face. **B**: The side of the reference face. 3D high-density faces are automatically and systematically divided into 10 different facial regions. Each vertex of the reference face is labeled and given a different color. The color keys below correspond to the different facial regions above.

Table 1
Lead SNPs associated with automatically segmented phenotypes of facial morphology in the GWAS.

Region[a]	Lead SNP[b]	CHR	Position	Allele	Candidate genes[c]	MAF	CC[d]	Df[e]	P-value[f]	N[g]	R[h]
Chin	rs4915551	1	197508901	G<A	*DENND1B*	0.15	0.20	24	1.37E-12	100	73
Eyes	rs12633011	3	138938995	A<G	*PISRT1*	0.46	0.24	39	1.77E-14	100	88
Nose	rs2045323	4	154831899	A<G	*DCHS2/SFRP2*	0.19	0.23	28	1.95E-17	100	100
Nose	rs11988731	8	100078926	C<T	*VPS13B*	0.40	0.20	28	6.44E-11	100	45

GWAS, genome-wide association study; SNP, single-nucleotide polymorphism; MAF, minor allele frequency.
[a] the facial regions defined in Fig. 1; [b] the SNP with smallest P-value in the genetic loci; [c] the gene or genes closest to the lead SNP within a 500 kb window; [d] canonical correlation; [e] the degree of freedom, which is the number of PCs of each facial region $-$ 1; [f] the right-tail P-value for an F test of each coefficient by employing Rao's statistic; [g] the number of cross-validations; [h] the number of successful replications with a significant threshold at levels 10^{-7} and 0.05 in the discovery and replication steps, respectively.

final 51 SNPs and the phenotypes of different facial regions. The results (Fig. 4; Table S4) showed that 24 SNPs were associated with different facial regions under the P-value threshold of 0.05, whose primary associated facial traits were various, but mostly nose related. They were originally reported in European, Latin American, and Asian populations. Some important genes (such as *DCHS2*, *SUPT3H*, *HOXD1*, *SOX9*, *PAX3*, *EDAR*, etc.) were further replicated in Han samples. Three SNPs (rs9995821, rs2045323, and rs12644248) near the candidate gene *DCHS2* were replicated in this study. The SNP rs9995821 was reported to be associated with module 45 in the nose region (Claes et al., 2018), and the other two SNPs rs2045323 and rs12644248 were reported to be associated with nose protrusion and columella inclination, respectively (Adhikari et al., 2016b). The SNPs rs9995821 and rs2045323 were replicated for its association with the nose (rs9995821, $P = 9.22 \times 10^{-17}$; rs2045323, $P = 1.95 \times 10^{-17}$)

and down cheek (rs9995821, $P = 0.02$; rs2045323, $P = 0.04$), and the SNP rs126442248 was replicated for its association with eyes ($P = 0.02$). Two SNPs rs1852985 and rs227833 near the *SUPT3H* and *RUNX* genes were, respectively, reported to be associated with nose bridge breadth (Adhikari et al., 2016b) and module 10 (Claes et al., 2018), while they were associated with multiple facial regions, such as eyes (rs1852985, $P = 1.28 \times 10^{-4}$; rs227833, $P = 0.01$), up cheek (rs1852985, $P = 3.93 \times 10^{-3}$; rs227833, $P = 0.03$), and nose (rs1852985, $P = 3.16 \times 10^{-7}$; rs227833, $P = 1.18 \times 10^{-4}$), in this study. The SNP rs970797 between the *HOXD1* gene and *MTX2* gene also showed association with multiple facial regions, which was previously reported to be associated with module 38 in the mouth region (Claes et al., 2018) and the tangent line angle of el3 (left and right curvature of the upper eyelid) (Cha et al., 2018). Two SNPs near *SOX9*, rs5821892 associated with module 2 (Claes et al., 2018) and
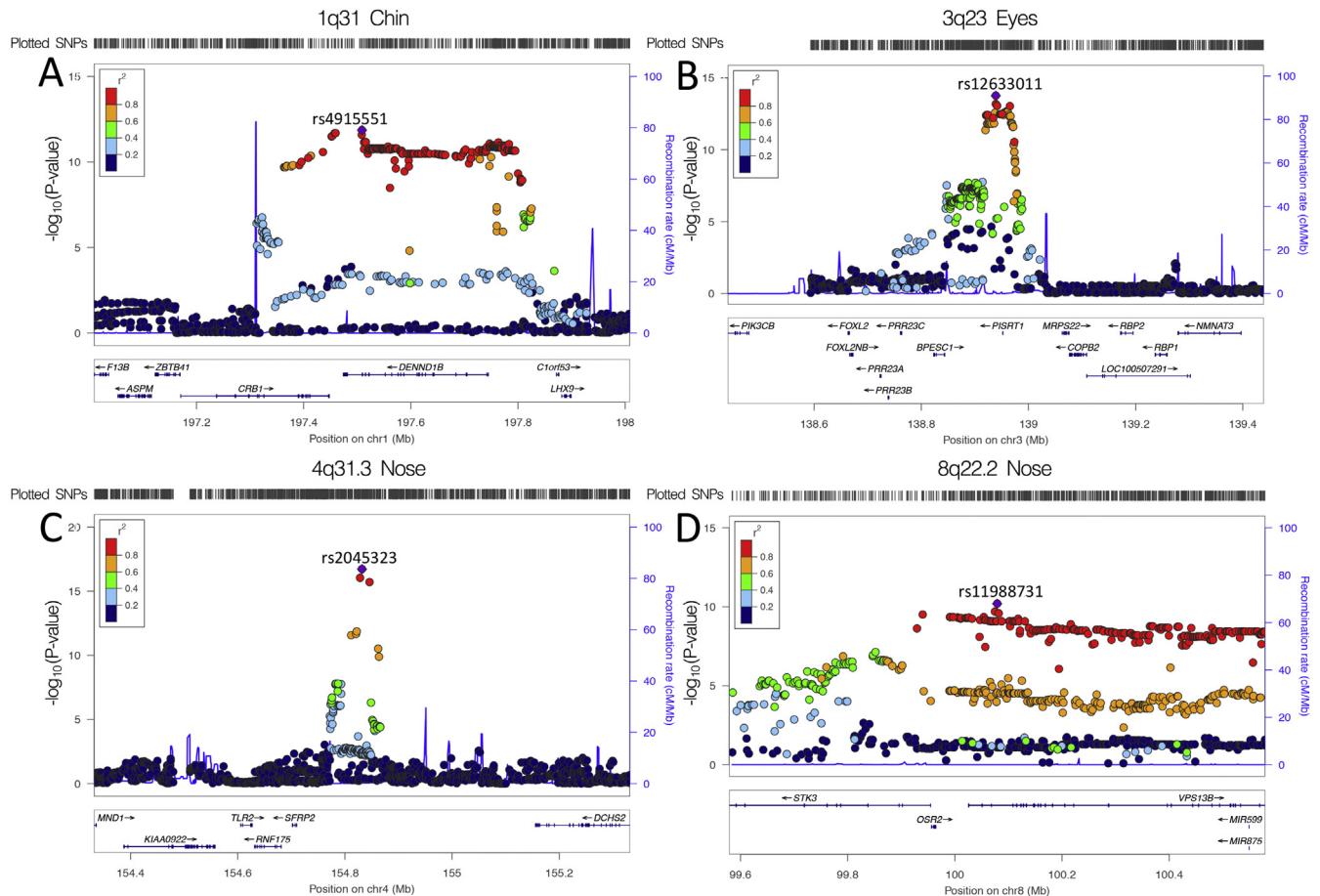
Fig. 2. Regional association plots of different facial regions across a 1-Mb window centered on the lead SNP. **A**: rs4915551 in the 1q31 region associated with the chin. **B**: rs12633011 in the 3q23 region associated with eyes. **C**: rs2045323 in the 4q31.3 region associated with the nose. **D**: rs11988731 in the 8q22.2 region associated with the nose. Association of SNPs in the one-stage GWAS plotted as -log$_{10}$(P-value) against the chromosomal base-pair position in the 500-kb flanking region centered on the lead SNP with the purple asterisk. The y-axis on the left shows -log$_{10}$(P-value), and the y-axis on the right shows the recombination rate, which is estimated from the HapMap China Beijing (CHB) and Japan Tokyo (JPT) populations. The gradient color of each SNP shows the linkage disequilibrium (R^2) between each SNP and the lead SNP. GWAS, genome-wide association study; SNP, single-nucleotide polymorphism.

rs2193054 associated with the profile nasal angle (Cha et al., 2018), were replicated for their association with the nose (rs5821892, $P = 1.49 \times 10^{-3}$; rs2193054, $P = 8.47 \times 10^{-4}$). The *PAX3* gene was originally reported to be associated with a nose-related trait (n-men) (Paternoster et al., 2012) and was replicated in several other studies (Liu et al., 2012; Adhikari et al., 2016b; Claes et al., 2018; Qiao et al., 2018). Two SNPs near the *PAX3* gene were also revalidated in this study. One SNP rs7559271 was associated with the nose ($P = 0.036$), while another SNP rs10176525 was associated with some other facial regions (Table S3). However, the MAF of rs10176525 is 0.01, so its result should be treated with caution. The SNP rs3827760 near *EDAR* was previously associated with chin protrusion (Adhikari et al., 2016b), and its proxy SNP (PSNP) rs4676213 ($R^2 = 0.87$) showed the association with the chin ($P = 0.028$), cheekbone ($P = 0.044$), and nose ($P = 0.042$) in our analysis.

Discussion

Recently, advanced high-resolution 3D imaging technology enables rapid and accurate acquisition of 3D human facial landscapes, which brings new possibilities for the genetic association studies of human facial morphological variations. GWASs combined with high-density 3D faces provide very detailed data of the association between millions of SNPs and normal facial variations. Phenotyping is a

crucial step in the association analysis, especially for complex traits such as faces. To understand how facial morphology is genetically defined, different facial measures have been used in previous studies. A recent study (Claes et al., 2018) argued that the traditional phenotype-first measures were not sufficient to detect associations between genetic variants and complex traits. They presented a data-driven hierarchical segmentation method. The nested structure of the obtained regions allows finding the global and local facial patterns. On the other hand, the segmentation results of the previous layer have a decisive impact on the subsequent layers, and the mandatory bifurcation may ignore the anatomical structure of the facial segments. Following the idea of data-driven segmentation, we propose a one-step approach to partition the face into several facial regions. The process of segmentation is fully automatic without pre-determining the number of regions. The result exhibits the unique anatomical features of the face that is composed of multiple regions.

Our findings show that three novel loci in *DENND1B*, *PISRT1*, and *VPS13B* are associated with the chin, eyes, and nose, respectively, and one reported locus between *DCHS2* and *SFRP2* was associated with the nose. *DENND1B* is a gene located in the chromosome 1q31 region. The region has been reported to be associated with chin morphology, but was linked to another candidate gene, *ASPM* (Claes et al., 2018). Our results show that the significant SNPs at 1q31 associated with the chin cover the whole *DENND1B* gene, whereas
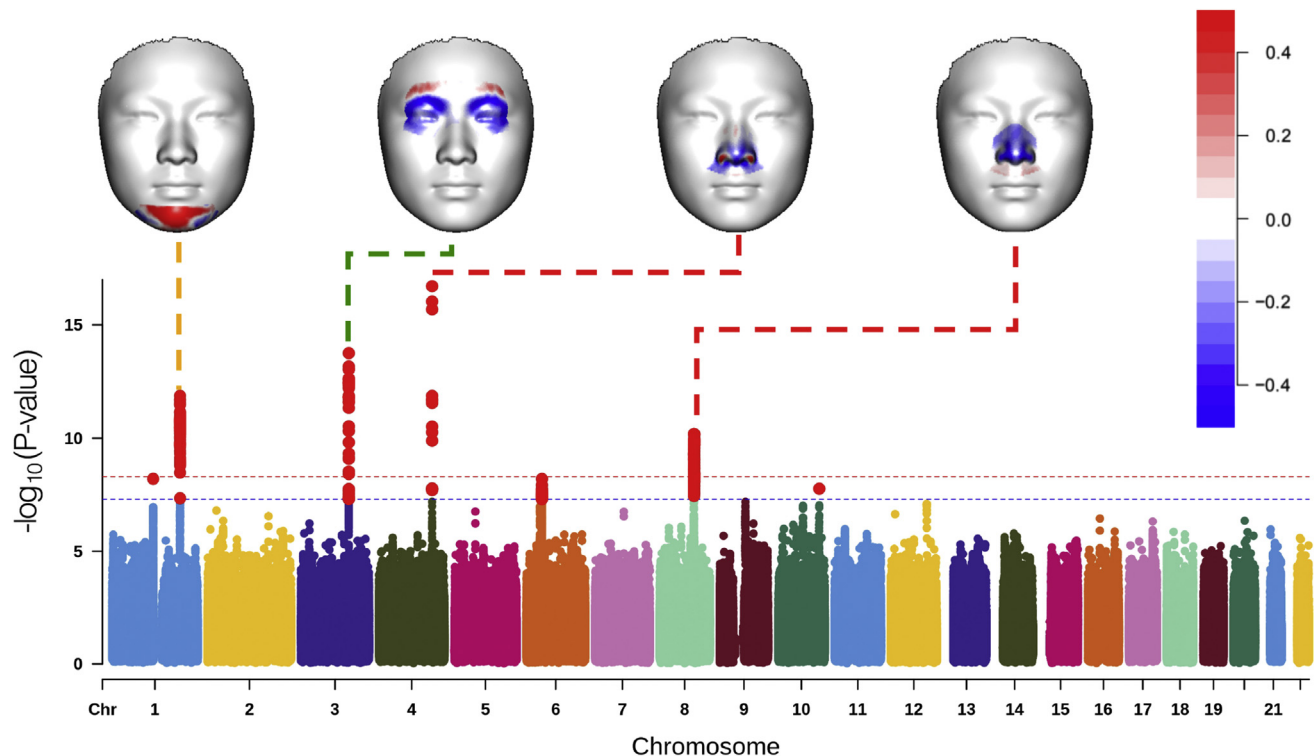
Fig. 3. The summary of the GWAS for different facial regions. The Manhattan plot of the combination of facial phenotypes with significant loci illustrates the chromosomal position of the association SNPs with -log₁₀(P-value) and highlights the SNPs with P values exceeding thresholds for every facial region. The blue horizontal dashed line represents the traditional genome-wide significance threshold ($P = 5 \times 10^{-8}$), and the red horizontal dash line represents the Bonferroni-corrected significance threshold ($P = 5 \times 10^{-9}$). Each heat plot of the face shows the effect on the representative facial region by its associated locus (linked by dash line), which is the external expansion or adduction by using the tetrahedral volume change of average faces with two homozygous genotypes at the lead SNP, with red indicating expansion and blue indicating adduction by the minor allele at the lead SNP, respectively. GWAS, genome-wide association study; SNP, single-nucleotide polymorphism.

we find no significant SNPs close to *ASPM* (Fig. 2A). *DENND1B* is a member of the connecdenn family with a DENN domain, an evolutionarily conserved protein module in eukaryotes (Marat et al., 2011), which interacts with the early endosomal small GTPase RAB35 as a guanine nucleotide exchange factor and binds to clathrin and clathrin adaptor protein-2 (Levivier et al., 2001; Yoshimura et al., 2010). It is involved in a variety of cellular functions, such as autophagy, receptor recycling, and synaptic vesicle endocytosis (Marat et al., 2011; Xu et al., 2015). Variants of *DENND1B* have been reported to be associated with asthma in children (Sleiman et al., 2010) and also with BMI in children with asthma (Melén et al., 2013). Asthma is a chronic disease characterized by the interaction among genetic and environmental factors, which may have an influence on face shape (Wenzel et al., 1985; Melén et al., 2013; Al Ali et al., 2014). However, the function of *DENND1B* in craniofacial development is not known. *PISRT1* is a nontranslated gene located 300 kb upstream of *FOXL2*, and the region around *PISRT1* can interact with the promoter of *FOXL2* (D`Haene et al., 2009; Verdin et al., 2013). In goats, *PISRT1* and *FOXL2* have been shown to be involved in polled intersex syndrome, in which the absence of their expression in the ovary leads to early testis differentiation and, as a result, a disorder of sexual development (Pailhoux et al., 2001; Pannetier et al., 2012). Loss of *FOXL2* expression leads to activation of the testis differentiation program and reduces the transcription of the *CYP19* gene, which converts androgen to estrogen, while *PISRT1* RNA is considered an inhibitor of androgenic differentiation genes, such as *SOX9* (Pailhoux et al., 2005). In addition, GWASs have recently linked SNPs at 3q23 between *FOXL2* and *PISRT1* to eyebrow thickness (Adhikari et al., 2016a), male-pattern baldness (Heilmann-Heimbach et al., 2017), and a Cornelia de Lange-like phenotype including ear and skin

dysmorphic features (Shaffer et al., 2017). The heterogeneous effects at 3q23 indicate complex associations with diverse craniofacial traits that deserve further study. *VPS13B* translates a potential transmembrane protein that may play a role in vesicle-mediated transport and sorting of proteins within the cell (Velayos-Baeza et al., 2004; Seifert et al., 2011). Mutations in the *VPS13B* gene underlie Cohen syndrome, a rare autosomal recessive disorder with characteristic craniofacial dysmorphism, such as micrognathia, short philtrum, and high-vaulted palate (Balikova et al., 2009; Seifert et al., 2009). *VPS13B* was also found in the discovery cohort of the Korean population, although it was not successfully verified in the replication cohort (Cha et al., 2018). *DCHS2* and *SFRP2* are the candidate genes of rs2045323 and rs9995821 at 4q31, the SNP rs2045323 was found to be associated with columella inclination and nose tip angle in admixed Latin Americans (Adhikari et al., 2016b), and the other SNP rs9995821 has been reported to be associated with the ala and nostrils (Claes et al., 2018), which is consistent with our results. *DCHS2* was first discovered in *Drosophila* and plays an important role in tissue morphogenesis and growth control (Clark et al., 1995; Zecca and Struhl, 2010). A recent study found that it is involved in the regulation network of cartilage differentiation during vertebrate craniofacial development (Le Pabic et al., 2014), which is initiated by *SOX9* (Bi et al., 1999). Another candidate gene *SFRP2* is a negative regulator of WNT signaling (Kurosaka et al., 2014). Disrupting hedgehog and WNT signaling interactions promotes cleft lip pathogenesis, and craniofacial deformities have been reported in *SFRP2*-mutant mice (Kurosaka et al., 2014).

To explore how these SNPs could potentially affect the early development of facial morphology, we compared the signal of histone modifications near these loci by using the data from a public
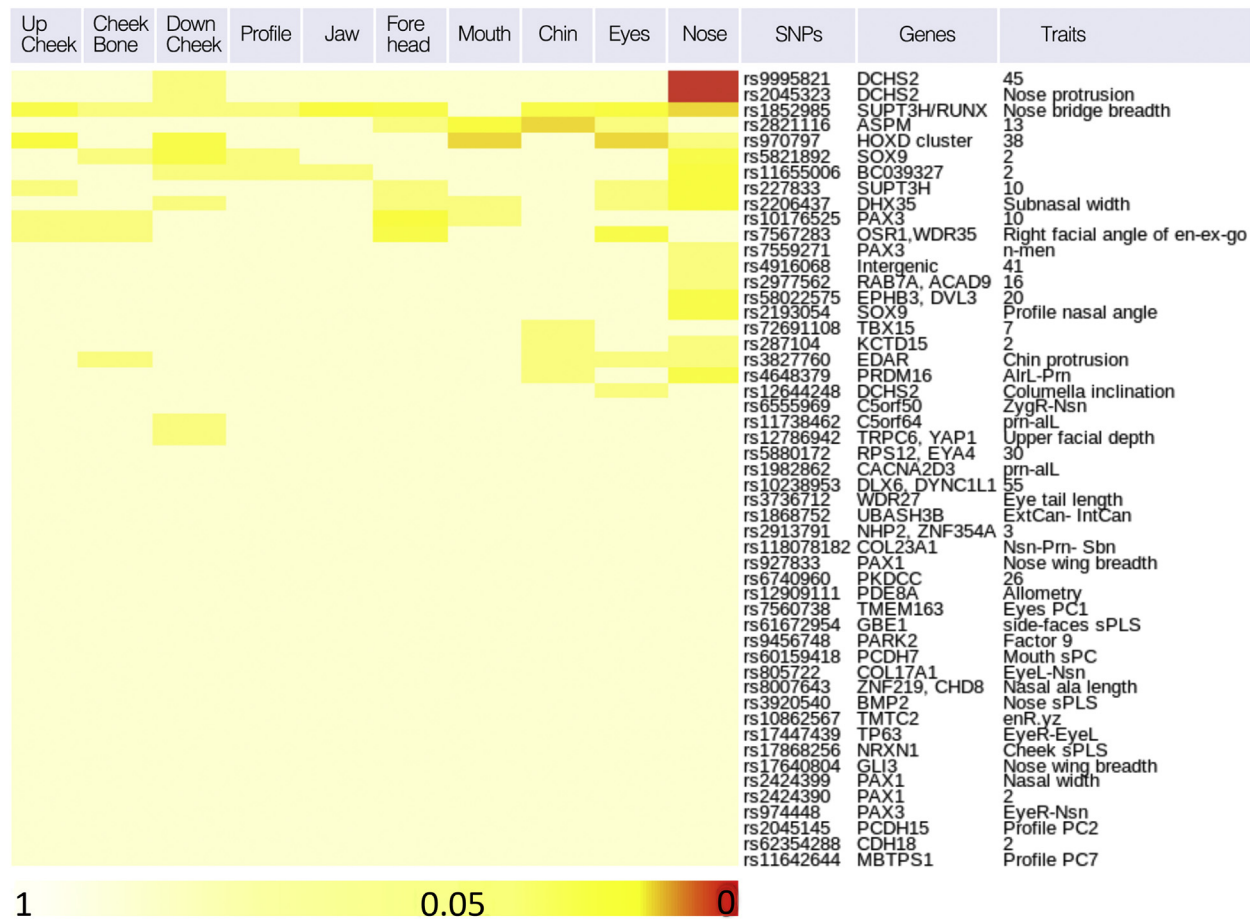
Fig. 4. The heatmap of the association results of the replication analysis. The heatmap shows the association between previously reported SNPs or its proxy SNPs with different facial regions, which were hierarchically clustered by its log$_{10}$(P-value). On the right are reported SNPs, candidate genes, and reported facial traits. The bottom is the P-value color key. Red is close to zero, yellow is less than or equal to 0.05, and white is close to 1.

study (Prescott et al., 2015) in human cranial neural crest cells (CNCCs), which plays a crucial role in early facial morphological formation generating individual differences and specific facial features (Bronner and Ledouarin, 2012; Green et al., 2015). We observed that there was a strong H3K27ac signal near rs11988731 in the human CNCCs (Fig. S5), suggesting that the SNP is located in a potential enhancer region and might affect the expression of the *VPS13B* gene through the potential regulatory element.

We performed a comprehensive replication analysis for previously reported variants associated with facial morphology (Table S3). About half of the SNPs (24/51) were replicated to varying degrees (Table S4), although the phenotypes were not exactly the same as the originally phenotypic measurements. Interestingly, we noticed that some loci were replicated in one facial region, while others were replicated in multiple facial regions. For example, the SNP rs7559271 near *PAX3* was associated with the nose, while the SNP rs1852985 near *SUPT3H* seemed to be associated with the whole face. The results indicate that the human face is affected by a number of genes with a range of effects, some of which affect only parts of the face and others affect the overall shape. The complex effects may be related to the precise gene regulation and complex signal pathways of face development and growth (Castaldo and Cerritelli, 2015). On the other hand, it also shows that some traditional facial measurements missed some associations. The segmentation approach to phenotyping can capture facial effects at different regions and scales.

A limitation of this study is the lack of a suitable independent replication panel. To avoid losing statistical power for discovery, we performed a one-stage GWAS on the whole panel, instead of using a two-phase design by randomly dividing the sample into the discovery panel and replication panel. However, the findings may increase false-positive results without the separate replication panel. To maintain the reliability and robustness of the results, the TSIR (Kang et al., 2015) approach as a cross-validation method was used to identify SNPs associated with the outcome of interest. TSIR is a powerful and efficient framework for identifying and internally validating SNPs for the GWAS when an independent replication panel for external validation is not available (Kang et al., 2015).

Another limitation is that the sample comes from the same city. Although the Han Chinese population shows a rather small genetic diversity compared with worldwide populations, it has a complex substructure. Our studied sample may not be fully representative of the Han Chinese population. However, under a limited sample size, the sample with a consistent genetic background can reduce the spurious associations caused by the population structure (Xu et al., 2009). In future, we would like to further expand the sample size and sample diversity of Han Chinese to get a more comprehensive understanding of genetic variation of facial traits and to help the reconstruction of facial features based on genotypes.

In conclusion, this is the first genome-wide association analysis of facial morphology in a Han Chinese population, which expands the

sample diversity for the genetic study of facial variations. For the complex multiregion structure of the human face, we proposed an autosegment phenotyping approach that can automatically divide 3D high-density faces into small anatomical regions. After a series of multivariable analyses of genotype-phenotype association, we identified three new genetic loci associated with the chin, eyes, and the nose and replicated many previously reported loci such as *PAX3* and *EDAR*. This study provides comprehensive phenotypes to capture biological variations in human facial morphology and expands our understanding of its genetic basis.

Materials and methods

Sample information

The 3,054 samples used in this study were collected in Taizhou, Jiangsu province, in 2014, among which 1,091 were men and 1,963 were women. The 3D facial images of the participants were collected using a 3dMDface 3d camera system, and blood samples from the participants were also collected. In addition, the age, height, and other information of the sample were collected through questionnaires. The participants were aged 31 through 87 years (mean: 56.7, standard deviation [SD]: 9.5), with men's age being 34—87 years (mean: 58, SD: 9.8) and women's age ranging between 31 and 86 years (mean: 56, SD: 9.3) (Fig. S1A). The BMI of samples was ranging between 14.7 and 50.3 (mean: 24.4, SD: 3.2), with men's BMI ranging between 14.7 and 35.2 (mean: 24.6, SD: 3) and women's BMI ranging between 15.6 and 50.3 (mean: 24.3, SD: 3.3) (Fig. S1B). All participants were apparently unrelated Han Chinese residents, with no history of facial trauma, facial plastic surgery, or any known disease affecting the face. In addition, the participants were excluded if they had any personal or family history of facial deformities or birth defects. After removing 395 participants, with missing personal information, 3D image mapping artifacts, or unqualified control of GWAS genotyping, a total of 2,659 participants were retained for analysis.

Ethics statement and consent to participate

All participants provided written informed consent to participate in the project. Sample collection for this study was conducted with the approval of the ethics committee of Shanghai Institutes for Biological Sciences and in accordance with the standards of the Helsinki Declaration.

High-density 3D facial image collection and registration

The workflow of the overall procedure is shown in Fig. S3. There are 4 steps: capturing 3D images using a 3dMDface camera system, landmarking critical points of the face using Facial Registration Analysis Software (FRAS), mapping with reference facial mesh, and adjusting the coordinate system by generalized Procrustes analysis (GPA).

Capturing 3D images using a 3dMDface camera system

The 3D facial image of the participants in this study was all captured using the 3dMDface system (www.3dmd.com/3dMDface), which captures a 3D facial image at a speed of ~1.5 ms with a geometric accuracy of 0.2 mm RMS (Weinberg et al., 2016). As per standard facial image acquisition protocols (Heike et al., 2010), the participants were required to follow the following conditions: having a natural facial expression without wearing any cosmetic makeup, pulling back the hair covering the forehead, opening eyes without wearing glasses, and closing the mouth naturally. After the shooting process, the quality of the 3D images was double checked on the computer through the software provided by 3dMDface. The 3D images were saved in TSB format by the 3dMDface system and converted to OGJ format for later processing.

Landmarking critical points of the face by FRAS

Seventeen critical points were automatically annotated on the meshes using an in-house 3D face landmarking algorithm (FRAS) based on PCA projection of texture and shape information to achieve the high-density point-to-point registration (Guo et al., 2013). In brief, the OBJ files were imported into the FRAS. First, the nose tip was automatically identified by using a sphere fitting approach. Then, the pose was normalized into a uniform frontal view. The postural standardized 3D facial images were projected into two-dimensional space through PCA, and the remaining 16 points were annotated automatically as per the geometric relationship and texture information (Guo et al., 2013). These landmarks were checked in FRAS and corrected using 3dMD Patient if the landmark was an outlier. In this study, two landmarks at two earlobes were removed because many female participants blocked their ears with loose hair.

Mapping with reference facial mesh

A facial mesh with high-image quality and a smooth surface was selected from the annotated faces and resampled uniformly to achieve an even density of one vertex per 1 mm-by-1 mm grid area. The reference face was a high-density grid mesh with 32,251 vertices. Then, the tool thin plate spline was used to twist and distort each face in the sample in turn, and the 15 landmarks on each surface were aligned with those on the reference face in a nonrigid way. Each vertex in the reference face was mapped to the closest point on the sample faces, and the sample faces form a high-density mesh that matches the reference face. Each surface is represented by 32,251 points of the 3D mesh.

Adjusting the coordinate system by GPA

To eliminate differences in position and orientation of surfaces, all surfaces of the participants were shifted and rotated to the same coordinate system by GPA.

During the processing of the aforementioned steps, the sample images were screened to remove the samples that had a serious defect in the facial images and did not pass the aforementioned steps owing to other reasons. Finally, the remaining 2,963 samples with high-quality images obtained the coordinate information in the same coordinate system through the aforementioned workflow for subsequent GWAS analysis. Each face is represented by a tri-mesh through a point matrix of dimensions 32,251 by 3 and the edge information of these points.

Phenotyping quantitative features of 3D high-density faces

The phenotypes of facial morphology in this study were based on the idea of data-driven segmentation (Claes et al., 2018) and the anatomical features of the human face (Siemionow and Sonmez, 2011), which mainly includes three processes: confounder correction on facial morphology, facial segmentation based on the affinity matrix, and extraction of the variation of facial regions.

Confounder correction on facial morphology

Facial morphology is influenced by both genetic factors and environmental factors. Correction of confounders, such as age and BMI, is performed to remove the variation of the facial structure from confounders and to retain the association signals from genetic

effects on the face. We used partial least squares regression to correct facial morphology for the confounders of sex, age, BMI, and facial size.

Facial segmentation based on the affinity matrix

We used the RV coefficient (Robert and Escoufier, 1976) to calculate the 3D correlation between each pair of vertices and construct a square affinity matrix that stores the correlation between every vertex pair, after confounder correction. Then, we transformed the affinity matrix into a graph Laplacian matrix (Lutzeyer and Walden, 2017) to distinguish the clusters by placing the vertices into a more separable space. We selected the top 10 PCs of the Laplacian matrix and used k-means to cluster vertices into k (from 2 to 20) clusters and determined the number of clusters ($k = 10$) with silhouette analysis (Rousseeuw, 1987) on k-means clustering.

Extraction of the variation of facial regions

We performed PCA on different facial regions to extract the variation of high-density segments, and used parallel analysis (Dinno, 2009) to determine the number of significant components contributing to the facial shape (Table S1).

Genotyping, quality control, and imputation

The participants including 2,978 Han Chinese from Taizhou, Jiangsu Province, and 2 Uyghurs from Urumchi, XinJiang Province (only for checking the population structure), were genotyped on an Illumina (San Diego, CA) OmniZhongHua-8 kit, and a total of 887,270 SNPs sites were obtained. Quality control filtering of genome-wide genotype data was performed by using PLINK version 1.07 (Purcell et al., 2007). The quality control process was mainly divided into two steps.

(1) Quality control of chip data: SNP loci were required to satisfy the following conditions:
    (a) The marker missing rate is less than 0.02.
    (b) The MAF is higher than 0.05.
    (c) The Hardy-Weinberg equilibrium test with a *P*-value of less than 0.001.
    (d) Exclusion of SNPs without chromosome information and duplicated SNPs: after removing the unqualified SNPs, there were 670,273 remaining SNPs.
(2) Quality control of samples: the samples were required to meet the following conditions:
    (a) Individual missingness: four were removed owing to the low calling rate after SNP pruning.
    (b) Duplicates and cryptic relatedness: two pairs of participants had extremely high identity by descent (IBD), which means that the two participants of each pair may be identical
    (c) Population structure: we performed a PCA on 2,978 Han Chinese and two Uyghurs using LD-pruned common markers genome wide to assess the genetic structure of our population. The result shows that two Uyghur participants were far away from the center of Han participants, which supports the high level of genetic homogeneity among our study participants (Fig. S4).
    (d) Heterozygosity and inbreeding: none of the participants were removed at threshold value > 0.2 or < -0.2).
    (e) Gender check: for four individuals, the gender did not fit the information given (*F* value less than 0.2 for women and more than 0.8 for men).

To capture information on unobserved SNPs and sporadically missing genotypes among genotyped SNPs, we carried out imputation to the high-resolution reference sample, the 1000 Genomes Project (Altshuler et al., 2012) Phase 3 reference panel. Specifically, the chip data were prephased using SHAPEIT2 (Delaneau et al., 2013) and then imputed for probabilistic genotypes using Impute2 (Howie et al., 2009, 2011). SNPs were removed with a low imputation quality score (< 0.3), followed by converting the genotype probabilities to the corresponding allele pair when it was over the threshold > 0.9 using GTOOL software (Delaneau et al., 2014). SNPs with a MAF < 0.05 or a call rate $< 90\%$ were omitted. There are a total of 4,757,938 genotyped SNPs after imputation and quality control.

Statistical analysis

The GWAS was performed on 2959 participants (952 men and 1,707 women) with 4,757,938 SNPs and 10 different facial regions. We used CCA to discover variants associated with each facial region. CCA is a multivariate statistical method that reflects the overall correlation between two groups of variables by using the correlation between the pairs of comprehensive indicators, which is implemented in PLINK and has demonstrated its advantages in multivariable analysis of genotype-phenotype association. Here, in brief, X is the sample phenotypic matrix, the PCs of facial variations from each segment, and Y is the genotype of the sample. $\rho(X,Y)$ is the canonical correlation (Formula 1).

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \tag{1}$$

The target function of CCA is to maximize the $\rho(X',Y')$ by optimizing the corresponding projection vector a and b, called the canonical correlation coefficients between X and Y, respectively (Formula 2).

$$\arg\max_{a,b} \frac{cov(X', Y')}{\sqrt{D(X')}\sqrt{D(Y')}} \tag{2}$$

where $X' = a^T X, Y' = b^T Y$.

We used the function 'cca' to carry out the CCA and the function 'F.test.cca' to test the statistical significance by using Rao's statistic from the R package yacca in this study.

The process of TSIR (Kang et al., 2015) was conducted to validate the SNPs, which were discovered in the one-stage GWAS with the significance threshold $P < 5 \times 10^{-8}$. In brief, we randomly divided the sample into the discovery group and the replication group as per the ratio of 7:3. SNPs were individually tested for the association with the outcome using CCA in both the discovery group and the replication group. If the significance level of an SNP exceeded $\alpha 1$ (10^{-7}) in the discovery group and $\alpha 2$ (0.05) in the replication group, respectively, the SNP was successfully discovered and replicated in this discovery-replication process. The process was repeated n times ($n = 100$) and if an SNP successfully appeared at least r times ($r = 20$) in these n tests, we assumed that the SNP is associated with the phenotype.

Quantile-quantile plots for facial phenotypes are shown with the distribution of observed *P*-values against the theoretical distribution of expected *P*-values through qqman (D. Turner, 2018).

The Manhattan plots for facial phenotypes with significant association loci were displayed with $-\log_{10}$ (*P*-value), the SNPs with significant signal were highlighted, and two threshold lines at 5×10^{-8} (Blue) and 5×10^{-9} (Red) were drawn using CMplot (https://github.com/YinLiLin/R-CMplot).

The regional association plot for a genomic region of 1 Mb centered on the significant loci was constructed using LocusZoom (Pruim et al., 2011).

Visualization of the effects on the associated facial region

In 3D facial registration, each 3D human face was represented by a tri-mesh surface after elastic registration to the reference face. All surfaces were under the same coordinate system by GPA. To catch the effects on the face by different alleles, the average faces of the genetic loci of interest were calculated as per different genotypes, such as TT, TG, and GG. Each of the average faces under different genotypes was composed of a tri-mesh surface with 32,251 vertices and 63,770 triangles. The advantage of using the average faces was that it could counteract the influence of random noise on the face and showed the influence of genetic factors. Because the GWAS was based on an additive model in the study, it was considered that these genetic variants also had an additive effect on the phenotype. The differences between homozygous genotypes were used to evaluate the genetic influence on the face.

To measure the effects on the associated facial region, the changes of the local face were obtained by correspondingly subtracting the volumes of the average face of homozygous genotypes at the lead SNP and smoothing the volume residuals of local facial regions. In detail, we first calculated the volumes of the tetrahedra formed by the origin (the center of the head) and each triangle of the average faces, respectively. The volume of the tetrahedron was determined by the height from the origin to the triangle and the area of the triangle. The distance between the center of the head and the surface of the face did not vary much. The changes of facial surface were usually due to the external expansion or adduction of small local regions or due to the variation of the angle between the triangles and the lines to the origin. Both kinds of differences could be well represented by the volume of the tetrahedra. Then, we compared the tetrahedral volume of average faces with two homozygous genotypes to evaluate the difference of the local facial region caused by genetic variants. The effect on the representative facial region by the lead SNP was smoothed by averaging the volume residuals of adjacent regions and shown on the heat plot of the reference face, with red indicating expansion and blue indicating adduction by the minor allele at the lead SNP, respectively. Source code can be accessed in GitHub (https://github.com/changebio/MAGPA).

CRediT authorship contribution statement

Yin Huang: Methodology, Investigation, Formal analysis, Software, Visualization, Writing - Original draft, Writing - Review & Editing. **Dan Li**: Methodology, Investigation, Formal analysis, Software, Writing - Review & Editing. **Lu Qiao**: Software, Data curation, Investigation. **Yu Liu**: Software, Data curation, Investigation. **Qianqian Peng**: Software, Data curation, Investigation. **Sijie Wu**: Software, Data curation, Investigation. **Manfei Zhang**: Software, Data curation, Investigation. **Yajun Yang**: Software, Data curation, Investigation. **Jingze Tan**: Software, Data curation, Investigation **Shuhua Xu**: Conceptualization, Resources, Funding acquisition, Writing - Review & Editing. **Li Jin**: Conceptualization, Resources, Funding acquisition, Writing - Review & Editing. **Sijia Wang**: Conceptualization, Methodology, Resources, Funding acquisition, Supervision, Writing - Review & Editing. **Kun Tang**: Conceptualization, Methodology, Resources, Funding acquisition, Supervision, Writing - Review & Editing. **Stefan Grünewald**: Conceptualization, Methodology, Resources, Funding acquisition, Supervision, Writing - Review & Editing.

Conflict of interest

The authors declare no conflict of interest.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2020.10.004.

References

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J.C., Al-Saadi, F., Johansson, J.A., Quinto-Sanchez, M., Acuña-Alonzo, V., et al., 2016a. A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. Nat. Commun. 7, 10815.

Adhikari, K., Fuentes-Guajardo, M., Quinto-Sánchez, M., Mendoza-Revilla, J., Camilo Chacón-Duque, J., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R.B., Pérez, G.M., et al., 2016b. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat. Commun. 7, 1—11.

Al Ali, A., Richmond, S., Popat, H., Toma, A.M., Playle, R., Zhurov, A.I., Marshall, D., Rosin, P.L., Henderson, J., 2014. The influence of asthma on face shape: a three-dimensional study. Eur. J. Orthod. 36, 373—380.

Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B., et al., 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56—65.

Balikova, I., Lehesjoki, A.E., De Ravel, T.J.L., Thienpont, B., Chandler, K.E., Clayton-Smith, J., Träskelin, A.L., Fryns, J.P., Vermeesch, J.R., 2009. Deletions in the VPS13B (COH1) gene as a cause of Cohen syndrome. Hum. Mutat. 30, 845—854.

Bi, W., Deng, J.M., Zhang, Z., Behringer, R.R., De Crombrugghe, B., 1999. Sox9 is required for cartilage formation. Nat. Genet. 22, 85—89.

Bronner, M.E., Ledouarin, N.M., 2012. Development and evolution of the neural crest : an overview. Dev. Biol. 366, 2—9.

Castaldo, G., Cerritelli, F., 2015. Craniofacial growth: evolving paradigms. Cranio - J. Craniomandib. Pract. 33, 23—31.

Cha, S., Lim, J.E., Park, A.Y., Do, J.-H., Lee, S.W., Shin, C., Cho, N.H., Kang, J.-O., Nam, J.M., Kim, J.-S., Woo, K.-M., Lee, S.-H., Kim, J.Y., Oh, B., 2018. Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genom. 19, 481.

Claes, P., Roosenboom, J., White, J.D., Swigut, T., Sero, D., Li, J., Lee, M.K., Zaidi, A., Mattern, B.C., Liebowitz, C., et al., 2018. Genome-wide mapping of global-to-local genetic effects on human facial shape. Nat. Genet. 50, 414—423.

Clark, H.F., Brentrup, D., Schneitz, K., Bieber, A., Goodman, C., Noll, M., 1995. Dachsous encodes a member of the cadherin superfamily that controls imaginal disc morphogenesis in Drosophila. Genes Dev. 9, 1530—1542.

Cole, J.B., Manyama, M., Kimwaga, E., Mathayo, J., Larson, J.R., Liberton, D.K., Lukowiak, K., Ferrara, T.M., Riccardi, S.L., Li, M., et al., 2016. Genomewide association study of african children identifies association of SCHIP1 and PDE8A with facial size and shape. PLoS Genet. 12, 1—19.

Cole, J.B., Manyama, M., Larson, J.R., Liberton, D.K., Ferrara, T.M., Riccardi, S.L., Li, M., Mio, W., Klein, O.D., Santorico, S.A., et al., 2017. Human facial shape and size heritability and genetic correlations. Genetics 205, 967—978.

Crouch, D.J.M., Winney, B., Koppen, W.P., Christmas, W.J., Hutnik, K., Day, T., Meena, D., Boumertit, A., Hysi, P., Nessa, A., et al., 2018. Genetics of the human face: identification of large-effect single gene variants. Proc. Natl. Acad. Sci. U. S. A. 115, E676—E685.

D. Turner, S., 2018. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J. Open Source Softw. 3, 731.

Delaneau, O., Marchini, J., McVeanh, G.A., Donnelly, P., Lunter, G., Marchini, J.L., Myers, S., Gupta-Hinch, A., Iqbal, Z., Mathieson, I., et al., 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat. Commun. 5, 1—9.

Delaneau, O., Zagury, J.F., Marchini, J., 2013. Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods 10, 5—6.

D'Haene, B., Attanasio, C., Beysen, D., Dostie, J., Lemire, E., Bouchard, P., Field, M., Jones, K., Lorenz, B., Menten, B., et al., 2009. Disease-causing 7.4 kb cis-regulatory deletion disrupting conserved non-coding sequences and their interaction with the FOXL2 promotor: implications for mutation screening. PLoS Genet. 5, e1000522.

Dinno, A., 2009. Implementing horn's parallel analysis for principal component analysis and factor analysis. STATA J. 9, 291—298.

Green, S.A., Simoes-costa, M., Bronner, M.E., 2015. Evolution of vertebrates as viewed from the crest. Nature 520, 474—482.

Guo, J., Mei, X., Tang, K., 2013. Automatic landmark annotation and dense correspondence registration for 3D human facial images. BMC Bioinf. 14, 1.

Heike, C.L., Upson, K., Stuhaug, E., Weinberg, S.M., 2010. 3D digital stereo-photogrammetry: a practical guide to facial image acquisition. Head Face Med. 6, 1—11.

Howie, B.N., Donnelly, P., Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529.

Howie, B., Marchini, J., Stephens, M., 2011. Genotype imputation with thousands of genomes. G3 Genes, Genomes, Genet. 1, 457—470.

Heilmann-Heimbach, S., Herold, C., Hochfeld, L.M., Hillmer, A.M., Nyholt, D.R., Hecker, J., Javed, A., Chew, E.G.Y., Pechlivanis, S., Drichel, D., et al., 2017. Meta-analysis identifies novel risk loci and yields systematic insights into the biology of male-pattern baldness. Nat. Commun. 8, 1—8.

Kang, G., Liu, W., Cheng, C., Wilson, C.L., Neale, G., Yang, J.J., Ness, K.K., Robison, L.L., Hudson, M.M., Srivastava, D.K., 2015. Evaluation of a two-step iterative resampling procedure for internal validation of genome-wide association studies. J. Hum. Genet. 60, 729—738.

Kurosaka, H., Iulianella, A., Williams, T., Trainor, P.A., 2014. Disrupting hedgehog and WNT signaling interactions promotes cleft lip pathogenesis. J. Clin. Invest. 124, 1660—1671.

Le Pabic, P., Ng, C., Schilling, T.F., 2014. Fat-dachsous signaling coordinates cartilage differentiation and polarity during craniofacial development. PLoS Genet. 10, e1004726.

Lee, M.K., Shaffer, J.R., Leslie, E.J., Orlova, E., Carlson, J.C., Feingold, E., Marazita, M.L., Weinberg, S.M., 2017. Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. PloS One 12, 1—13.

Levivier, E., Goud, B., Souchet, M., Calmels, T.P.G., Mornon, J.P., Callebaut, I., 2001. uDENN, DENN, and dDENN: indissociable domains in Rab and MAP kinases signaling pathways. Biochem. Biophys. Res. Commun. 287, 688—695.

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M.M., Hysi, P.G., Wollstein, A., Lao, O., de Bruijne, M., Ikram, M.A., et al., 2012. A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genet. 8, e1002932.

Lutzeyer, J.F., Walden, A.T., 2017. Comparing graph spectra of adjacency and laplacian matrices. Arxiv 1712, 03769.

Machiela, M.J., Chanock, S.J., 2015. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31, 3555—3557.

Marat, A.L., Dokainish, H., McPherson, P.S., 2011. DENN domain proteins: regulators of Rab GTPases. J. Biol. Chem. 286, 13791—13800.

Melén, E., Granell, R., Kogevinas, M., Strachan, D., Gonzalez, J.R., Wjst, M., Jarvis, D., Ege, M., Braun-Fahrländer, C., Genuneit, J., et al., 2013. Genome-wide association study of body mass index in 23,000 individuals with and without asthma. Clin. Exp. Allergy 43, 463—474.

Pailhoux, E., Vigier, B., Chaffaux, S., Servel, N., Taourit, S., Furet, J.P., Fellous, M., Grosclaude, F., Cribiu, E.P., Cotinot, C., et al., 2001. A 11.7-kb deletion triggers intersexuality and polledness in goats. Nat. Genet. 29, 453—458.

Pailhoux, E., Vigier, B., Schibler, L., Cribiu, E.P., Cotinot, C., Vaiman, D., 2005. Positional cloning of the PIS mutation in goats and its impact on understanding mammalian sex-differentiation. Genet. Sel. Evol. 37, 55—64.

Pannetier, M., Elzaiat, M., Thépot, D., Pailhoux, E., 2012. Telling the story of XX sex reversal in the goat: highlighting the sex-crossroad in domestic mammals. Sex. Dev. 6, 33—45.

Paternoster, L., Zhurov, A.I., Toma, A.M., Kemp, J.P., St Pourcain, B., Timpson, N.J., McMahon, G., McArdle, W., Ring, S.M., Smith, G.D., et al., 2012. 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.

Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., Wysocka, J., 2015. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell 163, 68—84.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., Frishman, D., 2011. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 27, 2336—2337.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559—575.

Qiao, L., Yang, Y., Fu, P., Hu, S., Zhou, H., Peng, S., Tan, J., Lu, Y., Lou, H., Lu, D., et al., 2018. Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. J. Genet. Genomics 45, 419—432.

Robert, P., Escoufier, Y., 1976. A unifying tool for linear multivariate statistical methods: the RV- coefficient. Appl. Stat. 25, 257.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53—65.

Seifert, W., Holder-Espinasse, M., Kühnisch, J., Kahrizi, K., Tzschach, A., Garshasbi, M., Najmabadi, H., Walter Kuss, A., Kress, W., Laureys, G., et al., 2009. Expanded mutational spectrum in Cohen syndrome, tissue expression, and transcript variants of COH1. Hum. Mutat. 30, 404—420.

Seifert, W., Kühnisch, J., Maritzen, T., Horn, D., Haucke, V., Hennies, H.C., 2011. Cohen syndrome-associated protein, COH1 is a novel, giant Golgi matrix protein required for Golgi integrity. J. Biol. Chem. 286, 37665—37675.

Shaffer, J.R., Li, J., Lee, M.K., Roosenboom, J., Orlova, E., Adhikari, K., Agee, M., Alipanahi, B., Auton, A., Bell, R.K., et al., 2017. Multiethnic GWAS reveals polygenic architecture of earlobe attachment. Am. J. Hum. Genet. 101, 913—924.

Shaffer, J.R., Orlova, E., Lee, M.K., Leslie, E.J., Raffensperger, Z.D., Heike, C.L., Cunningham, M.L., Hecht, J.T., Kau, C.H., Nidey, N.L., et al., 2016. Genome-wide association study reveals multiple loci influencing normal human facial morphology. PLoS Genet. 12, 1—21.

Siemionow, M.Z., Sonmez, E., 2011. Face as an organ: the functional anatomy of the face. In: Siemionow, M.Z. (Ed.), The know-how of face transplantation. Springer London, London, pp. 3—10.

Sleiman, P.M.A., Flory, J., Imielinski, M., Bradfield, J.P., Annaiah, K., Willis-Owen, S.A.G., Wang, K., Rafaels, N.M., Michel, S., Bonnelykke, K., et al., 2010. Variants of DENND1B associated with asthma in children. N. Engl. J. Med. 362, 36—44.

Velayos-Baeza, A., Vettori, A., Copley, R.R., Dobson-Stone, C., Monaco, A.P., 2004. Analysis of the human VPS13 gene family. Genomics 84, 536—549.

Verdin, H., D'haene, B., Beysen, D., Novikova, Y., Menten, B., Sante, T., Lapunzina, P., Nevado, J., Carvalho, C.M.B., Lupski, J.R., et al., 2013. Micro-homology-mediated mechanisms underlie non-recurrent disease-causing microdeletions of the FOXL2 gene or its regulatory domain. PLoS Genet. 9, e1003358.

Weinberg, S.M., Raffensperger, Z.D., Kesterke, M.J., Heike, C.L., Cunningham, M.L., Hecht, J.T., Kau, C.H., Murray, J.C., Wehby, G.L., Moreno, L.M., et al., 2016. 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.

Wenzel, A., Höjensgaard, E., Henriksen, J.M., 1985. Craniofacial morphology and head posture in children with asthma and perennial rhinitis. Eur. J. Orthod. 7, 83—92.

Xiong, Z., Dankova, G., Howe, L.J., Lee, M.K., 2019. Novel genetic loci affecting fecial shape variation in humans. Elife 8, e49898.

Xu, J., Fotouhi, M., McPherson, P.S., 2015. Phosphorylation of the exchange factor DENND 3 by ULK in response to starvation activates Rab12 and induces autophagy. EMBO Rep. 16, 709—718.

Xu, S., Yin, X., Li, S., Jin, W., Lou, H., Yang, L., Gong, X., Wang, H., Shen, Y., Pan, X., et al., 2009. Genomic dissection of population substructure of han Chinese and its implication in association studies. Am. J. Hum. Genet. 85, 762—774.

Yoshimura, S.I., Gerondopoulos, A., Linford, A., Rigden, D.J., Barr, F.A., 2010. Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors. J. Cell Biol. 191, 367—381.

Zecca, M., Struhl, G., 2010. A feed-forward circuit linking wingless, Fat-Dachsous signaling, and the warts-hippo pathway to drosophila wing growth. PLoS Biol. 8, e1000386.