Journal of Genetics and Genomics 45 (2018) 419–432

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Original research

Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction

Lu Qiao [a], Yajun Yang [b, c], Pengcheng Fu [d], Sile Hu [a], Hang Zhou [a], Shouneng Peng [a], Jingze Tan [b, c], Yan Lu [a], Haiyi Lou [a], Dongsheng Lu [a], Sijie Wu [a], Jing Guo [a], Li Jin [a, b, c, e], Yaqun Guan [f], Sijia Wang [a, e, g, *], Shuhua Xu [a, e, g, h, *], Kun Tang [a, *]

[a] Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, CAS, Shanghai 200031, China
[b] State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China
[c] Fudan-Taizhou Institute of Health Sciences, Taizhou 225300, China
[d] Department of Neurology, The First People's Hospital of Chenzhou, Hunan 423000, China
[e] Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China
[f] Department of Biochemistry and Molecular Biology, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, China
[g] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China
[h] School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

ARTICLE INFO

Article history:
Received 17 May 2018
Received in revised form
1 July 2018
Accepted 3 July 2018
Available online 16 August 2018

Keywords:
Genome-wide association study
Dense 3D facial image
Ancestry-divergent phenotypes
Face prediction
Forensic scenario

ABSTRACT

It is a long-standing question as to which genes define the characteristic facial features among different ethnic groups. In this study, we use Uyghurs, an ancient admixed population to query the genetic bases why Europeans and Han Chinese look different. Facial traits were analyzed based on high-dense 3D facial images; numerous biometric spaces were examined for divergent facial features between European and Han Chinese, ranging from inter-landmark distances to dense shape geometrics. Genome-wide association studies (GWAS) were conducted on a discovery panel of Uyghurs. Six significant loci were identified, four of which, rs1868752, rs118078182, rs60159418 at or near *UBASH3B*, *COL23A1*, *PCDH7* and rs17868256 were replicated in independent cohorts of Uyghurs or Southern Han Chinese. A prospective model was also developed to predict 3D faces based on top GWAS signals and tested in hypothetic forensic scenarios.

1. Introduction

The human face plays a pivotal role in daily life. Communication, mutual identification, sexual attraction, etc. strongly depend on face. It has been long noted that faces bear characterized features that may surrogate one's ancestry, even in highly admixed populations (Loos and Janssens, 2017). Our recent investigation (Guo et al., 2014) also revealed strong morphological divergence on multiple facial features, including nose, brow ridges, cheeks and jaw, between Europeans and Han Chinese. It is therefore a fundamental and intriguing question to ask: Which genetic variants contribute to the substantial morphological differences among continental populations?

Normal facial shape is known to be highly heritable (Weinberg et al., 2013; Hair et al., 2014; Ng et al., 2017). However, until recently, very little was known about the genetic basis of common variation of facial morphology. In the last few years, several genome-wide association studies (GWAS) were carried out and multiple face shape associated loci were identified (Liu et al., 2012; Paternoster et al., 2012; Adhikari et al., 2016; Cole et al., 2016; Shaffer et al., 2016). These studies all based their phenotypes on conventional scalar measurements involving limited number of landmarks. On the other hand, efforts have been paid to use dense 3D face modeling (3dDFM), a novel high-dimensional data format, to fully represent the complex facial shape phenome (Guo et al., 2013, 2014; Liu et al., 2016). Peng et al. (2013) first applied

* Corresponding authors.
E-mail addresses: wangsijia@picb.ac.cn (S. Wang), xushua@picb.ac.cn (S. Xu), tangkun@picb.ac.cn (K. Tang).

https://doi.org/10.1016/j.jgg.2018.07.009

3dDFM to identify the association between common mouth shape variation and a cleft-lip related genetic locus. Claes et al. (2014) showed that numerous genes are associated with complex normal facial shape variation based on 3dDFM, and proposed the potential of modeling the 3D face based on genotype and its use in forensic practice.

In this study, we aimed at identifying loci on a genome-wide scale that contribute to the divergent facial morphological features between Europeans and Han Chinese. In brief, GWAS was conducted on the polarized face phenotypes along the European-Han dimensions, and Uyghur was used as the study cohort to dissect the genotype-phenotype association. Uyghur is a minority group living in Xinjiang province, China, and was found to have arisen from ancient admixtures between East-Asian and European ancestries at a roughly equal ratio, followed by a long period of isolation (Xu and Jin, 2008; Xu et al., 2008, 2009). Furthermore, Uyghur facial traits demonstrated a wide range of shape gradients between the characteristic Europeans and Han-Chinese faces (Guo et al., 2014). These properties made Uyghur an ideal group to study the genetic variants of divergent facial features across Eurasia. We performed GWAS in 694 Uyghurs using both landmark based and 3dDFM based phenotypes. Significant loci were replicated in an independent Uyghur sample and a Han Chinese cohort. Next, we carried out a prospective investigation on whether 3D faces could be predicted to certain degree by using the top associated single nucleotide polymorphisms (SNPs). A quantitative model was established to summarize the phenotypic effects of multiple loci and to simulate realistic 3D face models. The prediction model was further tested in hypothetic forensic scenarios to evaluate the potential enhancement of face identification based on DNA.

2. Results

The studied cohorts included two independent Uyghur panels (694 and 171 individuals) from Xinjiang, China, used as GWAS discovery panel (Uyghur discovery, UIG-D) and replication panel (Uyghur replication, UIG-R), respectively. UIG-D and UIG-R were genotyped on Illumina Omni ZhongHua-8 and Affymetrix Genome-Wide Human SNP Array 6.0, respectively. In order to obtain a common set of SNPs, the array data of UIG-D and UIG-R were imputed to obtain the whole genome sequencing (WGS) SNPs. The SNPs genotyped on arrays were referred to as genotyped SNPs hereafter, in order to distinguish from the imputed SNPs. In addition, 1504 Han Chinese were sampled from Chenzhou, Hunan province, China (HAN-CZ) as a replication panel for candidate SNPs (Table 1). Furthermore, a Han Chinese cohort from Taizhou, Jiangsu province, China (HAN-TZ, 929) and 86 Shanghai residents of European ancestry (EUR) were used as the phenotype reference groups (Table 1). The average age of all participants was 20.02 ± 2.16 (s.d.) years old. The three-dimensional facial images were collected from the participants and mapped to a common 32,251 points' spatial dense mesh automatically (Guo et al., 2013). This was jointly analyzed among EUR, UIG-D and HAN-TZ, and the registered face images were then decomposed to various phenotype measurements. A candidate phenotype was chosen and termed as an ancestry-divergent phenotype if there existed a strong phenotypic divergence between EUR and HAN-TZ, and showed a wide distribution in UIG-D (Fig. 1). Three types of ancestry-divergent phenotypes were defined. First, 10 inter-landmark distances were selected (Fig. S1A and Tables S1–S3). The other two types of phenotypes were based on decomposing the high dimensional 3dDFM data. Briefly, we first extracted six facial features, namely, the brow ridge, eyes, side faces, cheeks, nose and mouth, grossly based on the reported strong inter-population differentiation (Guo et al., 2014) (Fig. S1B). The extracted 3dDFM data were decomposed by either partial least square (PLS) or principle component (PC) analysis (Guo et al., 2014), and the PLS model and PC model that defined the strongest segregation between EUR and HAN-TZ were selected as ancestry-divergent phenotypes, hereafter termed as sPLS and sPC respectively in each feature (Figs. 2, S2–S4 and Table S4). In total, six sPLS and sPC phenotypes were defined, corresponding to the six facial features.

2.1. Genome-wide association studies of Eurasian facial shape

We carried out sex-stratified and sex-mixed GWAS in UIG-D on 8,100,752 genome-wide SNPs for the 22 ancestry-divergent phenotypes (Table S5). In general, the lower *P* value inflation assessed by Quantile-Quantile (Q-Q) plot (D. Turner, 2018) analyses revealed little subpopulation stratification (Fig. S5). Six regions revealed signals of genome-wide significance (Pruim et al., 2010) at the level of $P < 5 \times 10^{-8}$ (Fig. 2 and Table 2). We focus on the discussion on the genotyped SNPs of the highest signal (index SNP) within each region, as their genotypes were more reliable. These include rs1868752 (at 11q24.1) associated with distance between external canthus and internal canthus (ExtCan-IntCan) in mixed genders, rs118078182 (on *COL23A1* at 5q35.3) associated with distance of Nasion point-Pronasale-Subnasale (Nsn-Prn-Sbn) in mixed genders, rs60159418 (on *PCDH7* at 4p15.1) associated with mouth sPC in males, rs17868256 (at 2p16.3) associated with cheek sPLS in females, rs3920540 (near *BMP2* at 20p12.3) associated with nasal sPLS in females, and rs61672954 (at 3p12.2) associated with the sPLS of side-faces in mixed genders. In order to control potential confounding effects from varying ancestry makeup (Fig. S6), we inferred the ancestry proportions of either East-Asian or European origin for each UIG-D individual. The six signals remained significant after accounting for the inferred ancestry in the association model (Table S6). As a total of 66 genome-wide analyses (22 traits in females, males and gender mixed) were carried out, the genome-wide *P* values of the index SNPs should be taken as nominal. Noticing that the ancestry-divergent phenotypes were highly correlated among each other (Table S7), we applied Benjamini-Hochberg procedure for FDR (Benjamini and Hochberg, 1995) to address the multiple testing problem. Four index SNPs remained genome-widely significant, including rs1868752, rs118078182, rs60159418 and rs17868256 (Table 2 and S8).

Ideally, the effects of facial shape related loci should be directly visualized on facial images. We modeled the effects of index SNPs

Table 1
Characteristics of study samples.

Location	Abb.	For	Ethnic	Total *N*	Male *N*	Mean age	Range (s.d.)
Urumchi, China	UIG-D	Discovery	Uyghur	694	270	20.09	17–25 (1.24)
Urumchi, China	UIG-R	Replication	Uyghur	171	63	20.55	17–25 (1.42)
Chenzhou, China	HAN-CZ	Replication	Han Chinese	1504	424	19.69	17–32 (1.67)
Taizhou, China	HAN-TZ	Discovery	Han Chinese	929	363	19.81	17–25 (0.95)
Europeans living in Shanghai	EUR	Discovery	European	86	57	27.59	16–42 (5.59)

N, sample size.

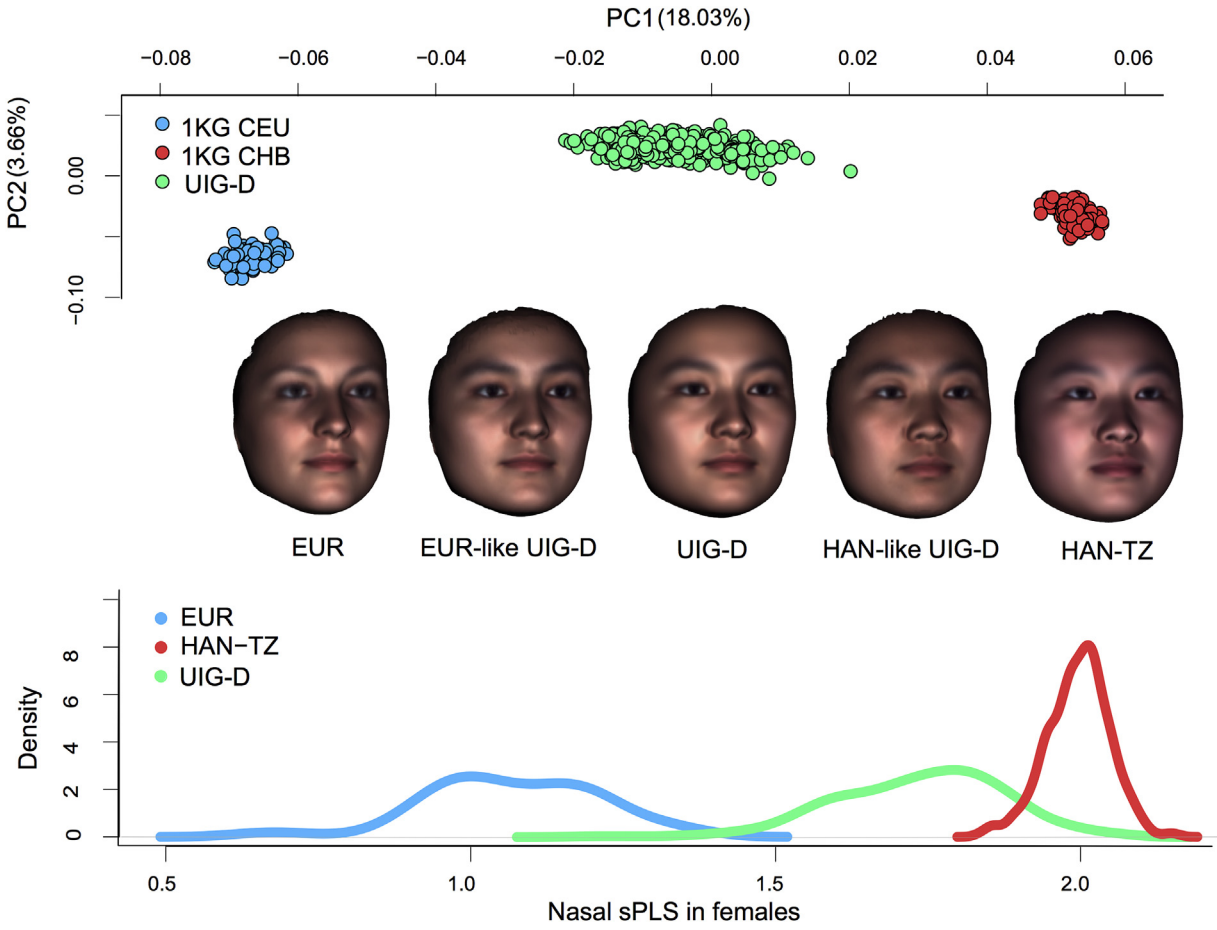

Fig. 1. Overall scheme of the study design. In the top panel, the genetic structure of three Eurasian populations was analyzed by PCA based on the 1 KG genome data of 85 CEU (blue), 97 CHB (red) and the whole-genome sequencing of 694 UIG-D (green). Clear clustering can be seen based on the ethnic backgrounds. In particular, UIG-D individuals clearly lie in the half way between CEU and CHB along PC1, and all consisted of a roughly equal ratio of CHB and CEU ancestries. Compared to the genetic composition, Uyghur individuals exhibit broad gradients of admixture in the facial phenotypes. The middle panel shows the average face models for EUR, European-like Uyghurs (EUR-like UIG-D), UIG-D, Han-like Uyghurs (HAN-like UIG-D) and HAN-TZ from left to right. EUR-like UIG-D and HAN-like UIG-D were obtained by averaging over 20 UIG-D individuals visually accessed to resemble Europeans or Han Chinese. The bottom panel shows the distribution of sPLS in nose, revealing a distinct segregation between EUR and HAN-TZ and a wide spread of UIG-D stretching between EUR and HAN-TZ along this phenotype dimension. In this study, the highly divergent phenotypes as shown above were selected and tested for association loci genome-widely in UIG-D.

using heat plots as well as extrapolated faces (Hu et al., 2017) (Fig. 3, Movies S1-S6). As can be seen in Fig. 3A (Movie S1), the extrapolated face towards the effect of rs1868752T had narrower eyes (smaller ExtCan-IntCan distance) compared to the G allele, resulting in a substantial displacement on the X axis; G also seemed to be associated with elevated nose ridge. SNP rs118078182 showed an obvious impact on the nasal shape along the Y and Z axes. Compared to rs118078182A, rs118078182G seemed to make the nose longer and more protrusive (taller) from the face, consistent with the association with Nsn-Prn-Sbn distance (Fig. 3B, Movie S2). For rs60159418 in males (Fig. 3C, Movie S3), the main shape changed the mouth along the Y axis, followed by Z axis. Allele G seemed to make the whole mouth area recessed from the face plane; in comparison, the mouth-chin curve bended convexly from the facial plane in the extrapolated face of the A allele. rs60159418 also seemed to influence other facial features: for G allele, nose and chin looked relatively protrusive outwards, and eye brow ridges seemed to elevate. The SNP rs17868256 mainly affected the shape of cheeks (Fig. 3D, Movie S4), with G allele associated with laterally expanded cheeks, making the face look wider on the X axis. On the Y axis, rs17868256G also seemed to lift the cheek protrusion upwards. For SNP rs3920540 in females (Fig. 3E, Movie S5), G allele was mainly associated with repressed nasal bridge and nasal tip along Z axis compared to T allele; G allele also seemed to link to more protrusive chin on the extrapolated face. The most notable effect of rs61672954 occurred around the jaw lines (Fig. 3F, Movie S6), with the A allele associated with stronger jawlines and therefore comparatively wider lower face than the extrapolated face of G allele.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.jgg.2018.07.009.

2.2. Replication studies and meta-analysis

We replicated the six GWAS significant loci in an independent Uyghur cohort (UIG-R) and a Han Chinese cohort (HAN-CZ) (Table S9). The former has the same ethnic background with UIG-D (Xu and Jin, 2008; Xu et al., 2008, 2009) and the latter represents a pool of Han Chinese ancestry from southern China (Kilpelainen et al., 2016). Face is highly complex, and the effects of genetic variants on face can be subtle and strongly depend on other factors such as ethnicity and gender (Hopman et al., 2014; Talbert et al., 2014). On the other hand, face related genetic loci may be pleiotropic, e.g., that a single variant may influence facial morphology on different parts and/or in different ways (Liu et al., 2012; Paternoster et al., 2012; Kamberov et al., 2013; Dick et al., 2014). In view of this,

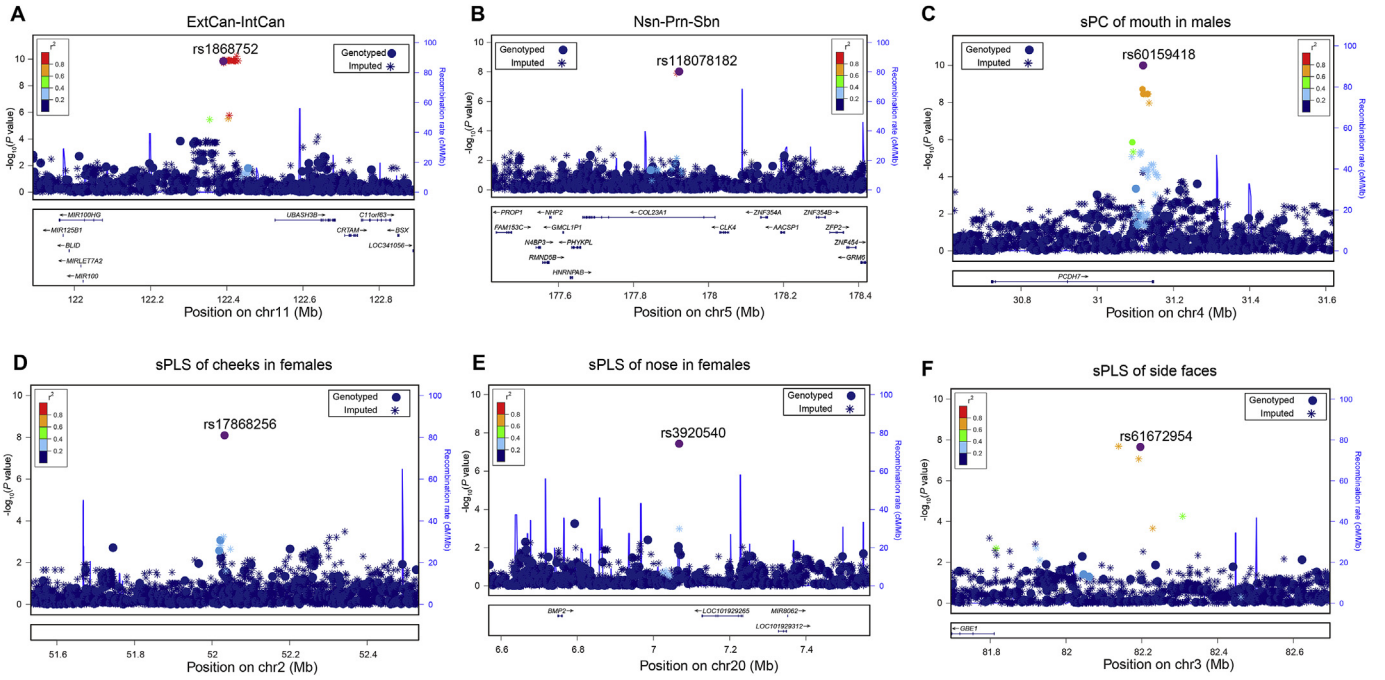

Fig. 2. Six genomic regions harboring SNPs of genome-widely significant associations with facial shape. **A**: 11q24. **B**: 5q35. **C**: 4q15. **D**: 2q16. **E**: 20q12. **F**: 3q12. The LocusZoom plots show the association (left y-axis) with corresponding ancestry-divergent traits labeled on the top. Genotyped SNPs are denoted by circles and imputed SNPs are denoted by stars. The plots are given for the 500 kb flanking region centered by the genotyped index loci indicated by purple circles. Color gradients of SNPs depict the linkage disequilibrium (r^2) between each SNP and index SNP.

we defined two types of association replications: the narrow-sense and broad-sense replications. Narrow-sense replication stands for the association signals replicated on exactly the same ancestry-divergent phenotypic measurement; whereas broad-sense replication requires the index loci to show evidences of association with any shape changes in the same facial feature. In this study, both the narrow- and broad-sense replications were conditioned in the same gender group as for the discovery panel. As a result (Table 2), the association of rs1868752 with ExtCan-IntCan exhibited narrow-sense replication ($P = 0.04$) in a dominant model in HAN-CZ. The SNP rs118078182 showed narrow-sense replications in an additive ($P = 0.004$) and a dominant model ($P = 0.003$) in UIG-R, in a dominant model ($P = 0.02$) in HAN-CZ, as well as in the additive model of UIG-R and HAN-CZ combined ($P = 0.02$). For rs60159418, the narrow-sense replication was tested by projecting the UIG-R and HAN-CZ 3dDFM data to the sPC of mouth where the GWAS signal was found in UIG-D males. It revealed marginal significance in HAN-CZ ($P = 0.09$) and combined group of UIG-R and HAN-CZ ($P = 0.05$). For rs17868256, the narrow-sense association in females was successfully replicated for cheek sPLS in UIG-R for both the additive model ($P = 0.04$) and a dominant model ($P = 0.02$). The other two index loci didn't show evidences of narrow-sense replication.

To systematically test the broad-sense replication, we carried out pair-wise shape distance (PSD) permutation as previous proposed (Peng et al., 2013) and PLS-based permutation for the 3dDFM data. Table 3 summarizes the results of broad-sense replications. In general, these tests confirmed the results of narrow-sense replications, showing that rs1868752, rs118078182, rs60159418 and rs17868256 affect the overall shapes of the corresponding features. Furthermore, rs3920540, the nose related locus that failed to replicate in the narrow-sense test, turned out to significantly affect the overall nasal shape in PSD permutation test ($P = 0.01$) and PLS-based permutation test ($P = 0.03$) in HAN-CZ.

Visualization in UIG-R and HAN-CZ revealed highly consistent effects of the ancestry-divergent variants as in UIG-D (Hu et al., 2017) (Fig. 4), despite the distinct ethnicity of HAN-CZ. Intriguingly in UIG-R, the effects of some index SNPs were strongly persistent not only within the facial features of GWAS signals, but also across the whole face. In HAN-CZ, similar influence on the whole face could also be observed. In particular, rs1868752T was involved in recessive eye-sockets and repressed nasal bridge in all three cohorts (Fig. 4A); rs118078182A seemed to make the whole face more flatter and rounder in addition to its effects on nasal shape (Fig. 4B); other than affecting the mouth shape (Fig. 4C), rs60159418G was also associated with stronger brow ridges and more protrusive chin among the three groups. These indicated that the identified association signals were authentic, and the candidate variants were in general highly pleiotropic.

2.3. Replication of reported variants of facial variation

We explored whether the candidate SNPs that were previously reported to affect normal facial variation (Liu et al., 2012; Paternoster et al., 2012; Adhikari et al., 2016) also showed signals of association in our combined Uyghur cohort (UIG-D + UIG-R). Replications were carried out either on the original or related measurements for 12 SNPs (Table S10). Notably, numerous candidate loci were re-validated to varying degrees. Briefly, rs4648379 in *PRDM16*, rs7559271 in *PAX3*, rs2045323 in *DCHS2*, rs17640804 in *GLI3*, rs805722 in *COL17A1* and rs927833 in *PAX1* reported to affect nasal phenotypes in different ethnic groups (Liu et al., 2012; Paternoster et al., 2012; Adhikari et al., 2016), were found to also modulate normal nasal shape in Uyghurs (Table S10). The SNPs rs3827760 in *EDAR* and rs6184 in *GHR* that were previously linked to mandibular shape variation turned out to be significant or marginally significant in our study (Dick et al., 2014; Adhikari et al., 2016) (Table S10). Interestingly, the SNP rs642961 in *IRF6*, previously found to be associated with the mouth shape in Han Chinese females (Rahimov et al., 2008; Peng et al., 2013), also showed

Table 2
SNPs with GWAS signals and their narrow-sense replications.

SNP	Chr. BP[a]	At or near gene	MAF	AA[b]	Ancestral allele frequency		Fst[e]	Ratio[f]	Gender group	Feature	Discovery in additive model[g]			Replication in additive model						Replication in dominant model[h]				Meta-analysis UIG-D+UIG-R in additive model	
					UIG[c]	CHB[d]	CEU[d]				UIG-D			UIG-R		HAN-CZ		UIG-R+HAN-CZ		aa+ab:bb		aa:ab+bb			
											Beta	*P* value	Adjusted *P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	*P* value	UIG-R	*P* value	HAN-CZ	*P* value	UIG-R	*P* value	HAN-CZ	Beta	*P* value
rs1868752	11	122391442	UBASH3B	0.015	T	0.99	0.961	0.985	0.0002	0.95	Mixed	ExtCan-IntCan	3.8	**1×10^{-10}**	**3×10^{-3}**	-	-	0.1	6×10^{-1}	0.1	6×10^{-1}	-	**4×10^{-2}**	6×10^{-1}	8×10^{-1}	3.19	**7.70×10^{-9}**
rs118078182	5	177922198	COL23A1	0.10	G	0.90	0.791	0.995	0.2	0.11	Mixed	Nsn-Prn-Sbn	-2.0	**8×10^{-9}**	**5×10^{-2}**	-2.3	**4×10^{-3}**	-0.3	**7×10^{-2}**	-0.4	**2×10^{-2}**	3×10^{-1}	**2×10^{-2}**	**3×10^{-3}**	2×10^{-1}	-2.07	**2.64×10^{-10}**
rs60159418	4	31120752	PCDH7	0.36	A	0.36	0.631	0.04	0.5	0.003	Male	Mouth sPC	-19.8	**9×10^{-11}**	**3×10^{-3}**	-5.7	3×10^{-1}	-4.4	**9×10^{-2}**	-4.6	**5×10^{-2}**	7×10^{-1}	1×10^{-1}	2×10^{-1}	2×10^{-1}	-17	**9.78×10^{-10}**
rs17868256	2	52032773	-	0.36	A	0.64	0.476	0.753	0.1	0.20	Female	Cheek sPLS	0.05	**7×10^{-9}**	**5×10^{-2}**	0.05	**4×10^{-2}**	0.003	5×10^{-1}	0.005	3×10^{-1}	**2×10^{-2}**	2×10^{-1}	2×10^{-1}	9×10^{-1}	0.0522	**1.07×10^{-9}**
rs3920540	20	7067738	BMP2	0.15	T	0.85	0.883	0.859	0.00002	0.99	Female	Nose sPLS	0.07	**3×10^{-8}**	1×10^{-1}	0.03	4×10^{-1}	-0.0007	9×10^{-1}	-0.0008	9×10^{-1}	-	4×10^{-1}	4×10^{-1}	6×10^{-1}	0.0652	**4.69×10^{-8}**
rs61672954	3	82196528	-	0.033	G	0.97	0.879	1.000	0.08	0.25	Mixed	side-faces sPLS	0.1	**2×10^{-8}**	9×10^{-2}	-	-	0.002	8×10^{-1}	0.001	9×10^{-1}	-	4×10^{-1}	7×10^{-1}	1	0.101	**2.61×10^{-7}**

[a]NCBI build 37; [b]ancestral allele; [c]in meta-UIGhurs (UIG-D + UIG-R); [d]1000 Genomes phase 3 database; [e]Fst calculated between CHB and CEU in 1KG phase1 release 2 database; [f]ratio means the rank of Fst in all genotyped SNPs; [g]in additive linear regression model, homozygote ancestral alleles as 0, heterozygote as 1, homozygote derived alleles as 2; [h]a allele stands for ancestral allele; "-" indicates null results in genes or null test in replications. The values in bold represent the SNP is significant tested by the model.

marginal significance ($P = 0.05$) in Uyghur females but not in males or mixed gender group, implying that the dependence of the genetic effect of rs642961 on gender was shared among different populations.

2.4. A prospective 3D face prediction model based on genome-wide SNPs

Hypothetically, if a quantitative trait is highly heritable, a proper model featuring major genetic factors should lead to true prediction (Yang et al., 2010, 2011; Wood et al., 2014). For the purpose of face prediction within Uyghurs, GWAS should be carried out on common facial variations in all dimensions. Therefore, in addition to the 22 ancestry-divergent phenotypes, we further included GWAS on a comprehensive list of phenotypes of various perspectives within Uyghurs, including 26 inter-landmark, 18 PC and 12 PLS phenotypes (Supplementary data). The total number of GWAS used for face prediction was exceptionally large (234) and a conventional cutoff for GWAS is impractical. We therefore assumed that SNPs reaching a nominal genome-wide threshold of P value $< 1 \times 10^{-6}$ were enriched for facial shape related loci. All SNPs that satisfy the threshold were combined into a panel of 277 top SNPs (Table S11). Based on these top SNPs, a simple quantitative model was constructed using UIG-D data. Briefly, for each SNP, a residual face was obtained for each genotype by subtracting the genotype average face by the global average. To compose a predicted face (PF), the 277 residual faces were scaled by a global "effect coefficient" α (Supplementary data; Fig. S7), and then added to the base face (the average face stratified by gender) according to the specific genotypes of the individual of interest. Details of the model construction can be referred to the Supplementary data.

The prediction model was applied to Uyghur individuals from the independent replication panel (UIG-R). The PF and actual faces can be visualized but quantitative evaluation of their similarity is difficult with human observers (Fig. 5A, Movie S7). To formally access the resemblance between predicted and actual faces, we constructed a robust shape similarity statistic: the shape space angle (SSA). SSA was defined as the angle between two shapes in the 3dDFM data space (Smeets et al., 2010; Claes et al., 2014) (Supplementary data). SSA would achieve 0 if the two shapes were the same. An SSA of 90-degree stands for statistical independence, whereas an SSA greater than 90 indicates that two faces are deviated from the mean face in reverse directions. Tests of shape similarity were carried out in UIG-R stratified by gender to see whether the differences from the actual face to their corresponding PF was indeed smaller than to random faces. We used two types of random faces. The first was the random-genotype predicted face (RGF; Supplementary data), adopted from a previous study (Claes et al., 2014). Briefly, an RGF was generated in the same way as the PF, except that the genotype set used for prediction was replaced by a random genotype set permuted from the known frequencies of genotypes in the combined UIG cohort. The other was the random-actual face (RAF) set, which was obtained by randomly sampling the UIG-R actual faces. We constructed two types of tests, one was to compare the SSA distributions (inter-distribution test) between the random predicted faces and the PF faces; the other was to test the average similarity score (single-score test) between the PF and actual faces, against the null distributions based on multiple iterations using RGF (Claes et al., 2014) or RAF as prediction (Supplementary data). Both tests generated similar results (Fig. 5B and C), that the females of UIG-R did not show marked differences between the PF and random predictions (the inter-distribution test $P = 0.87$ in RGF and $P = 0.73$ in RAF; the single-statistic test $P = 0.35$ in RGF and $P = 0.34$ in RAF). However, in males, the true prediction significantly outperformed the random predictions in all the tests

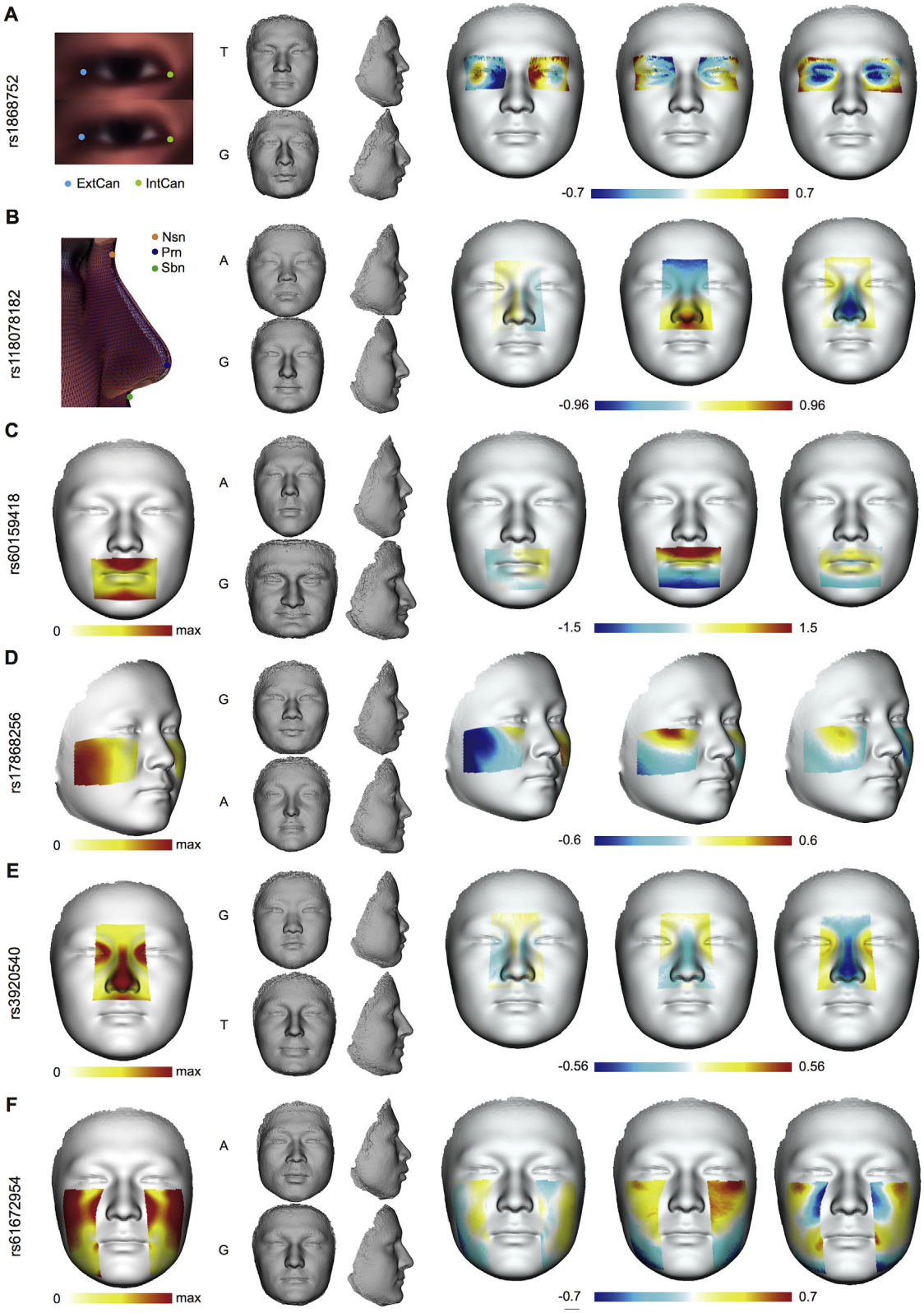

Fig. 3. Heat plots and extrapolated faces of six index SNPs affected on responding partial shape. **A**: The association of rs1868752 with the distance between ExtCan and IntCan. **B**: The association of rs118078182 with the distance of Nsn-Prn-Sbn. **C**: The association of rs60159418 with the sPC of mouth in males. **D**: The association of rs17868256 with the sPLS of cheeks in females. **E**: The association of rs3920540 with nasal sPLS in females. **F**: The association of rs61672954 with the sPLS of side faces. For each SNP, the first face shows the general effect on the corresponding feature as the displacement of landmarks or meshes. The mid-panel of four miniature faces gives the extrapolations towards the Han trend on the top, or the European trend on the bottom, with the associated allele labeled at the left side. The extrapolated faces were morphed by exemplifying the difference between the average faces of the opposite homozygotes if both are more than 10% frequent in UIG-D, or the major homozygote and the heterozygote if otherwise. The last three faces depict the signed displacement of the average faces of the fore-mentioned genotypes in X, Y and Z axes; obtained by subtracting the average face of European-trend from that of the Han-trend.

Table 3
Broad-sense replications of SNPs with GWAS signals.

SNP	PSD-based permutation in HAN-CZ[a]		PLS-based permutation	
	Between genotypes	HAN-CZ	UIG-R	HAN-CZ
rs1868752	TG:TT	$\mathbf{3 \times 10^{-2}}$	–	7×10^{-1}
rs118078182	AA + GA:GG	$\mathbf{3 \times 10^{-2}}$	$\mathbf{3 \times 10^{-3}}$	3×10^{-1}
rs60159418	AA:GG	2×10^{-1}	2×10^{-1}	$<\mathbf{1 \times 10^{-3}}$
rs17868256	AA:GG	$\mathbf{9 \times 10^{-3}}$	1	$<\mathbf{1 \times 10^{-3}}$
rs3920540	TG:TT	$\mathbf{1 \times 10^{-2}}$	7×10^{-1}	$\mathbf{3 \times 10^{-2}}$
rs61672954	AA:GG	4×10^{-1}	–	3×10^{-1}

[a] Replication only in HAN-CZ as sample sizes in different genotype groups are too small in UIG-R. The values in bold represent *P* value less than 0.05.

(the inter-distribution test $P = 0.01$ in RGF and $P = 0.04$ in RAF; the single-statistic test $P = 0.02$ in RGF and $P = 0.02$ in RAF).

Supplementary video related to this article can be found at https://doi.org/10.1016/j.jgg.2018.07.009.

Intriguingly, face prediction based on genotypes would promote the suspect identification in pragmatic forensic scenarios. We simulated toy forensic scenarios where a single true suspect should be picked up from a group of 8 candidates (Fig. S8). We supposed that the 3D facial data and genotypes could be obtained from each candidate to allow mathematical comparison between the actual and PF. Interestingly, there was a significant increase in accuracy (+6.3%, empirical $P = 0.025$) in males compared to the simulated neutral distribution (12.4% ± 0.03%) assuming no prediction power at all. In females, the prediction did not substantially enhance the accuracy (+3.3%, empirical $P = 0.076$) (Fig. 5D).

3. Discussion

This GWAS study was aimed at identifying genetic loci associated with normal facial variations. The facial features were based on complex 3dDFM data to reveal multiple genetic determinants. The genome-wide significant loci were located on independent regions and respectively associated with shape of eyes, nose, mouth, cheeks and side faces. We successfully replicated four loci, rs1868752, rs118078182, rs60159418 and rs17868256 in independent cohorts on the same phenotype measurements. In addition, the association signal of rs3920540 was replicated in broad-sense.

Among the significant SNP loci, some showed strong inter-population differentiations, and seemed to account for the morphological divergence among the Eurasian groups. The SNP rs118078182 is an intronic variant in collagen type 23 alpha 1 (*COL23A1*). A possible role of spatial/temporal regulation in facial morphogenesis was noted for *COL23A1* (Lu et al., 2016). Interestingly, the G allele of rs118078182 associated with the European trait of longer and taller nose is almost fixed (99.5%) in Utah residents with Northern and Western European ancestry (CEU), compared to the sequentially lower frequencies of ~90% in Uyghur and ~79% in Han Chinese in Beijing, China (CHB), suggesting that rs118078182 contributes to the nasal shape divergence across Eurasia. SNP rs60159418 is situated in gene protocadherin 7 (*PCDH7*). Previous studies showed that *PCDH7* played a key role in osteoclastogenesis (Ried et al., 2016), and its homologue gene is a pivotal regulator in the head formation of the mouse embryo (Loos, 2016). Interestingly, the derived G allele associated with the concaved European mouth shape demonstrated the strongest genetic differentiation among the three populations (G allele frequency is 0.96 in CEU, 0.64 in the Uyghurs (UIG) and 0.37 in CHB; $Fst_{CEU\text{-}CHB} = 0.5$; Table 2), suggesting an involvement of local adaptations in this region. For the SNP rs17868256, the derived G allele is associated with the Han Chinese trait of higher zygomatic arches and more cambered outwards and backwards zygomatics, and the corresponding allele frequency is also the highest in CHB (0.52), followed by UIG (0.36) and CEU (0.25), indicating that rs17868256 is involved in the "high cheek" phenotype commonly found in East Asians.

However, some other candidate SNPs may not have directly contributed to the inter-population facial divergence. SNP rs1868752, neighboring gene ubiquitin associated and SH3 domain containing B (*UBASH3B*), has low and similar frequencies in both CEU (~1.5%) and CHB (~4%), suggesting that this SNP affects eye lengths within rather than between populations. The other nasal shape related SNP rs3920540 is in affinity (300 kb) from the gene bone morphogenetic protein 2 (*BMP2*), which is involved in the development of bone and cartilage (Geiger et al., 2003; Bulik-Sullivan et al., 2015). The allele rs3920540T associated with taller nose does not seem to differ in frequency among the three Eurasian populations (0.86, 0.85 and 0.88 in CEU, UIG and CHB, respectively, $Fst = 0.00002$; Table 2). We assume that these loci only account for within population facial variations. On the other hand, face is highly complex and polygenic, and phenotypic differentiation may have resulted from accumulation and interaction of many rather small effects. Further discussions on potential genetic interactions and explained variance by SNPs are given in Supplementary data.

One evident limitation of this study is the relatively small sample sizes for the UIG cohorts. Despite the limited sample sizes, we were still able to detect 6 genome-wide association signals, among which five were replicated to various degrees. This may be attributed mainly to the specific study designs: first, as the Uyghur was examined on the ancestry-divergent phenotypic dimensions, the search was thus focused on genetic variants of large effect size. This rendered higher test power for a given sample size. Second, the 3dDFM data densely annotated each face by over 30,000 points, resulting in highly comprehensive face phenome data. Based on this, the association signals were searched both phenome-widely and genome-widely, providing enough test power to detect the phenotype-genotype associations.

Several trends are worth noting about the genetic architecture of facial morphology. First, no GWAS loci of "major effects" were identified in spite of the strong overall inter-population divergence, e.g., in nasal shape across Eurasia (Guo et al., 2014). This is in contrast to the case of skin pigmentation whose major genetic factors explain substantial phenotypic variance (Relethford, 2002, 2004; Myles et al., 2007). This suggests that the human face should be best described by a typical polygenic model of complex trait. Second, facial shape related variants may be highly pleiotropic. All the candidate loci in our study were not only associated with specific features, but also seemed to affect the shape of whole face, with relatively consistent patterns across different sample panels (Figs. 3 and 4).

In the end, we showed that an additive genetic model of whole face shape, based on a set of SNPs of top association signals, would lead to measurable predictive power. We showed that for independent individuals (UIG-R), the prediction model can also construct realistic 3D faces significantly closer to the actual face than random expectation (Fig. 5B and C); test in the hypothetic forensic scenarios revealed a moderate yet significant enhancement of the identification rate in males (Fig. 5D). There may be redundant markers within the 277 top SNPs, as the SNP sets filtered for strong LD gave very similar performance (Table S12 for prediction, Table S13 for forensic). Fundamentally, it can be argued that this prediction model is not really additive, as the best-fitting effect coefficient α is far lower than 1 (Fig. S7), even for the SNP set (209) well controlled for the physical LD (Fig. S7C). This may implicate epistatic interactions between the causal SNPs, or due to a proportion of false-positive signals in the top SNP set. Further studies specifically designed for examination of the genetic architecture is needed to address this question. Furthermore, we noticed that

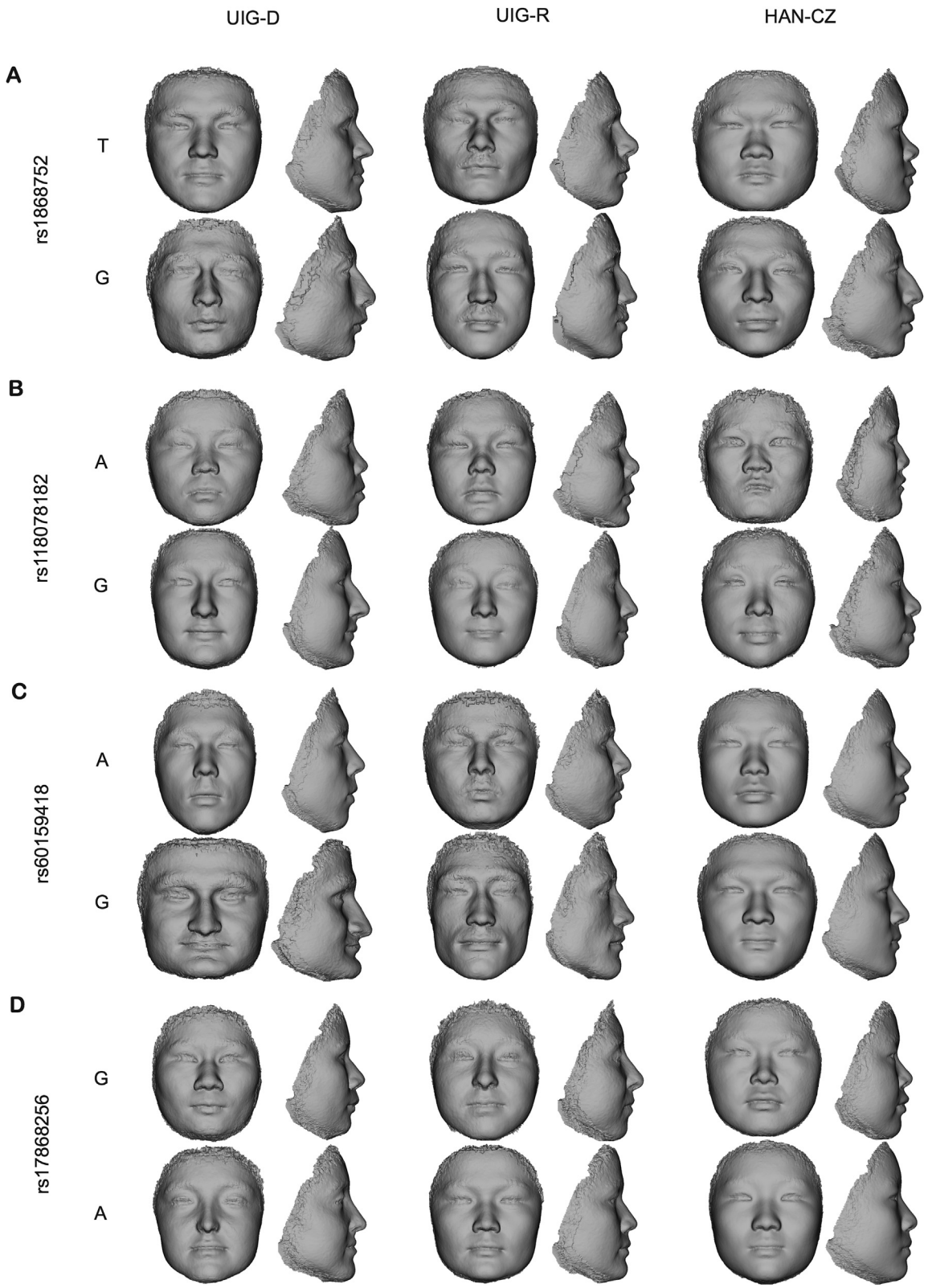

Fig. 4. Visualization in UIG-R and HAN-CZ revealed largely consistent effects of the ancestry-divergent variants as in UIG-D. **A**: rs1868752. **B**: rs118078182. **C**: rs60159418. **D**: rs17868256. We compared the extrapolated faces in UIG-D, UIG-R and HAN-CZ from left to right. For each locus, the top faces are in the trend of Han Chinese and the bottom ones are in European trend for the corresponding feature.

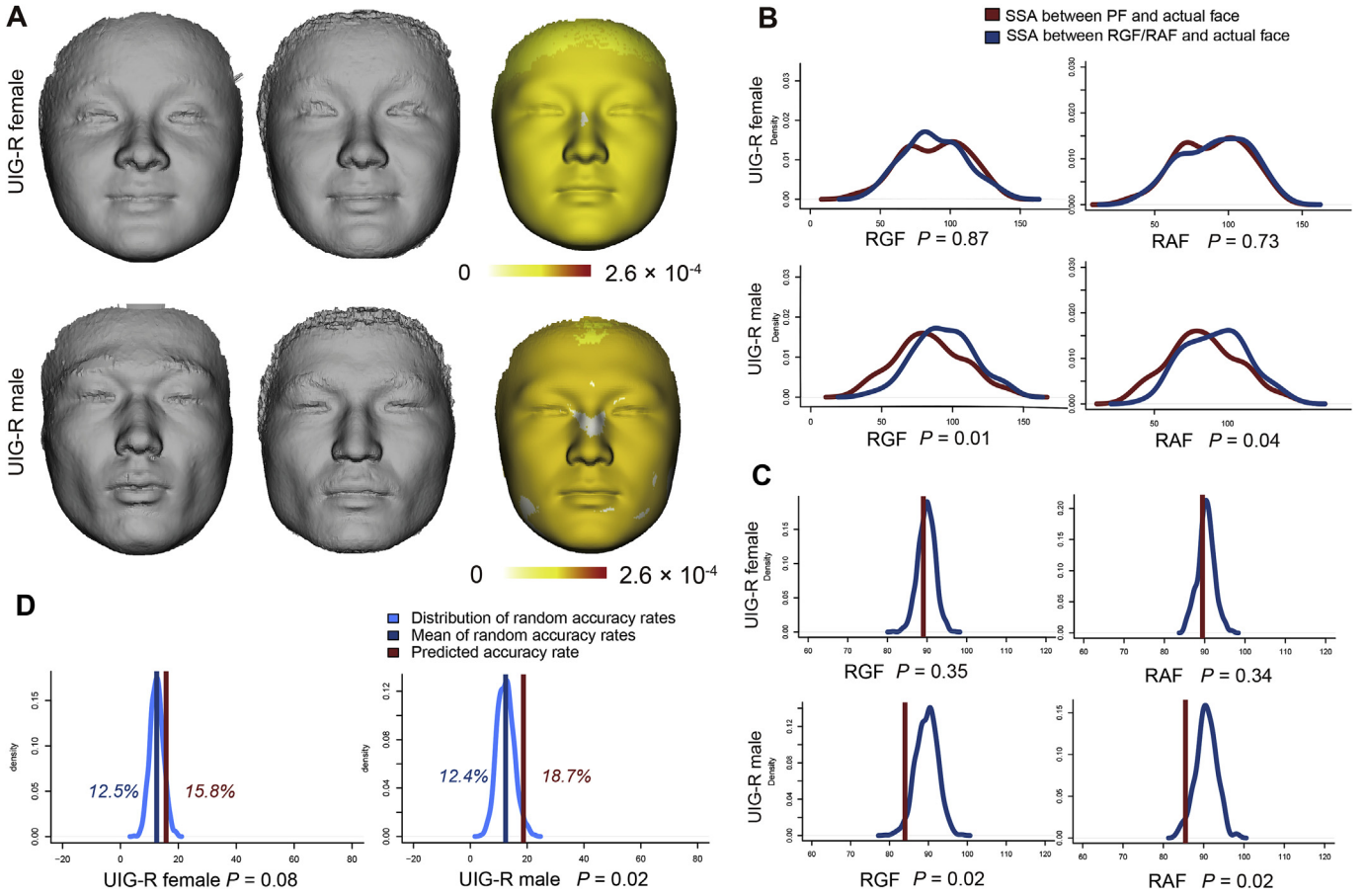

Fig. 5. Test of the prediction model in UIG-R. **A**: Cases of visualization of actual face (left column), the PF (middle column) and the displacement between the pair in heat plot (right column). **B**: Evaluation by inter-distance test. The comparison between true SSA distribution (in brown) and random SSA distribution in RGF (left column) or RAF (right column) was tested by Student's *t*-test. **C**: Evaluation by single-statistic test. The average SSA determined for the cohorts (in brown) was compared to the RGF (left column) and RAF (right column) distributions under null hypothesis (in dodgerblue). *P* values are the probability of predicted statistic distributed on the relative random normal distribution calculated like normal one-side *P* value. **D**: Evaluation of the face prediction in a hypothetic forensic scenario. SSA as accuracy statistics was evaluated in UIG-R females (left column) and UIG-R males (right column). The true accuracy rate based on SSA, in brown, were determined by examining how many cases of successful identification were achieved in all combined iterations. The random accuracy rate was calculated after reshuffling the pairwise corresponds between actual face and PF. This process was repeated 1000 times to obtain a distribution of random accuracy rates (in dodgerblue) under null hypothesis. The *P* values are the proportion of how many random accuracy rates were larger than true accuracy rate calculated as the empirical *P* value.

increasing the number of top SNPs in the prediction model (either for P value $< 1 \times 10^{-7}$ or 1×10^{-5} SNPs) could not improve the predictive power. Such saturation analysis suggests a finite number of loci affecting the normal facial shape. After all, this prediction model is highly simplified and explorative. Much work is needed to improve its performance to be formally tested in real forensic scenarios.

4. Materials and methods

4.1. Study cohorts

EUR was a resident cohort living in Shanghai with self-reported European ancestry between 16 and 42 years old. The HAN-TZ participants were self-reported Han Chinese samples collected from Taizhou, Jiangsu province. College students of self-reported Han ethnicity from Xiangnan University in Chenzhou, Hunan province were collected as HAN-CZ. The UIG-D and UIG-R were composed of college students of self-reported Uyghurs collected from Xinjiang Medical University in Urumchi, Xinjiang province. The self-reported ancestry information was requested for the last three generations, and individuals with mixed ancestry or missing information were excluded for further analyses. For EUR, a participant was used only if his/her ancestries of the last three generations were all from European countries. Individuals with obvious health problems or any history of facial surgery were ruled out. All sample collection in this study was carried out in accordance with the ethical standards of the ethics committee of the Shanghai Institutes for Biological Sciences (SIBS) and the Declaration of Helsinki, and has been specifically surveyed and approved by SIBS. All methods were carried out in accordance with the corresponding guidelines and regulations. A written statement of informed consent was obtained from every participant, with his/her authorizing signature. The participants, whose transformed facial images were used in this study as necessary illustrations of our methodology, have been shown in corresponding figures.

4.2. High-density 3D facial image alignment

The 3dMDface® system (www.3dmd.com/3dMDface) was used to collect high-resolution 3D facial images. We first established dense anatomical correspondence across dense surfaces of 3D facial

images automatically as described previously (Guo et al., 2013). Briefly, 15 salient facial landmarks were annotated automatically using principal component analysis (PCA) projection of texture and shape information. A reference face was selected for high image quality and smooth surface, and its mesh was resampled to achieve an even density of one vertex in each 1 mm × 1 mm grid. There were 32,251 vertices in total for the reference mesh. Afterwards, the reference face was warped to register to each objective face to ensure the complete overlapping of the 15 landmarks via a thin-plate spline (TPS) transformation. The vertices on the reference face then found their closest projections on the sample face to define the samples' new vertices, resulting in a point-to-point correspondence. At last, the Generalized Procrustes analysis (GPA) was used to align the sample grids into a common coordinate system. As a result, we obtained a set of 32,251 3D points to represent each participant's face. Samples with defective images were removed from the study.

4.3. Genotyping, quality control and imputation

Genomic DNA extracted from blood samples of UIG-D and UIG-R were genotyped on corresponding genotyping arrays. Quality control (QC) was performed using PLINK v1.07 (Purcell et al., 2007). Furthermore, 92 individuals from UIG-D were whole-genome sequenced at high coverage (30 ×). We didn't consider SNPs on mitochondria. SNPs with MAF <0.01, call rate <90%, or rejection in the Hardy-Weinberg Equilibrium test with $P < 1 \times 10^{-6}$ were omitted from the study. Genomic ancestry was detected using EIGENSTRAT 5.0.2 (Patterson et al., 2006; Price et al., 2006) with CHB and CEU from 1000 Genomes Project (Howie et al., 2012) (1 KG phase1 release v2) to remove samples who were not Uyghurs ancestry. Specifically, PCA using 17,552 autosomal SNPs pruned from UIG-D panel based on call rate (>90%), MAF (>0.01), and LD (pairwise $r^2 < 0.1$) was used to assess population structure. Samples with genotype missing rate >0.1 were removed. Samples were further examined by pairwise IBD estimation and inbreeding coefficients to remove individuals of close genetic relationships. Specifically, we used 17,552 independent SNPs to estimate the pairwise IBD to find pairs of individuals who genetically look too similar to each other and set inbreeding coefficients >0.2 or < −0.2 as in inbreeding. Individuals with aberrant gender information checked using X chromosome data were discarded. Finally, samples with defective 3D images were removed from this study. Detailed results of QC were given in Table S5. In the end, a total of 847,046 SNPs were used in 694 UIG-D and 758,453 SNPs were used in the 171 UIG-R. Genomic DNA of HAN-CZ was extracted from saliva according to a modified Phenol-chloroform protocol (Hoff-Olsen et al., 1999). Targeted genotyping for the index SNPs were carried out by SNaPshot multiplex system on an ABI3130xl genetic analyzer by Genesky Biotech, Shanghai, China. The results of SNaPshot genotyping QC are shown in Table S14.

In order to merge the data between UIG-D and UIG-R, which used different genotyping platform, we carried out imputation within UIG-D and UIG-R separately, using whole genome sequencing data as reference. Specifically, the chip data were pre-phased using SHAPEIT v2.r790 (Delaneau et al., 2014), and then imputed for probabilistic genotypes by Impute2 (Howie et al., 2009) taking 1092 individuals from 1000 Genomes project phase 1 and 92 high-coverage (30 ×) whole-genome sequencing data of Uyghurs (unpublished data) as reference. The genotyped SNPs with a large difference between the information and concordance values (info_type0 − concord_type0 > 0.1) or a low concordance value (<0.7) were excluded for further analysis. Additionally, SNPs were removed with low imputation quality score (<0.4), followed by converting the genotype probabilities to corresponding allele pair when it was over the threshold > 0.9 using GTOOL software (Delaneau et al., 2014). SNPs with MAF <0.01 or a call rate <90% were omitted.

4.4. Extraction of ancestry-divergent phenotypes

Three types of ancestry-divergent phenotypes were defined. The first type, the landmark-based ancestry-divergent phenotypes were defined based on the 15 landmarks (Fig. S1A). Briefly, the whole face was divided into five parts (Table S1). Thirty-six distances or angles were then defined as traditional anthropometric traits, many of which measure highly correlated facial features (Table S2). We chose 10 traits (Table S3) as the ancestry-divergent phenotypes, which surrogated the 36 traits and showed higher segregation (lower Student's *t*-test *P* value, Table S3) between EUR and HAN-TZ.

The second type is the PCA based ancestry-divergent phenotypes (sPC). Six facial features were extracted from the 3dDFM face data (Fig. 1B), which were then decomposed using PCA separately. The top PCs that together explained 98% total variance were examined for phenotypic segregation between EUR and HAN-TZ using Student's *t*-test (Fig. S2 and Table S4). The PC that defined the most significant segregation was taken as the ancestry-divergent phenotype, named as sPC.

PLS based ancestry-divergent phenotypes (sPLS) were defined as follows: In the PLS equation "response ~ predictor", we labeled the individuals from EUR as 1 and HAN-TZ as 2 in response. The predictor term featured $3 \times n$ matrix of the 3dDFM data, where n stands for the number of vertices in the corresponding facial feature. The PLS regression was first carried out only with EUR and HAN-TZ data. An optimal PLS model is a regression model composed of the first k top ranking PLS component(s), where k renders the best prediction from predictor to response. Leave-one-out (LOO) was used to cross-validate PLS models, with increasing number of top components, and the root mean squared error of prediction (RMSEP) was used to evaluate the overall prediction. There exist two cross-validation (CV) estimates based on RMSEP (Fig. S3 and Table S4): 'CV' is the ordinary CV estimate, and 'adjCV' is a bias-corrected CV estimate. A PLS regression model achieved the maximal discrimination between EUR and HAN-TZ (Table S4) when the RMSEP showed turning point on adjCV curve (Fig. S3), and the corresponding k was used. An optimal PLS model was found in each facial feature and termed as sPLS model.

4.5. Statistical analyses and replications

Genome-wide association studies were carried out using PLINK v1.07 (Purcell et al., 2007) and a linear regression model with additive SNP effects was used. For multiple testing correction, we performed the Benjamini-Hochberg procedure for FDR (Benjamini and Hochberg, 1995) to the combined results of the 66 autosomal GWASs. An FDR level of 0.05 was used to determine whether a test remained significant after multiple-testing correction.

For narrow-sense replication, the 3dDFM data from UIG-R and HAN-CZ were either projected to the sPC spaces or input to the sPLS models defined in UIG-D, to obtain the corresponding phenotype scores in narrow-sense replication. An additive linear regression was used to test the association between index SNPs and the corresponding narrow-sense phenotype. This test required a minimum of 10 individuals that carry at least one minor allele. UIG-D and UIG-R samples were combined to do the meta-analysis same as the narrow-sense replication method.

Index SNPs were also tested for broad-sense replication. For PLS-based permutation, genotypes (0, 1, 2 in additive model and 0, 1 in dominant model) of each SNP were taken as predictor. PCs that

together explained 99% of the total variance of each facial feature were adopted as response. A scheme of LOO was used in which *n-1* individuals were used as training and the left-out was used as test. This was repeated until every individual was used as a test sample. The training set was used to build a PLS regression model whose optimal number of components was determined by RMSEP under CV, as described before. Based on this model, a predicted genotype was given to the test sample. By such analogy, every sample obtained a predicted genotype. Afterwards, the correlation between actual genotype and predicted genotype can be calculated. Next, we performed a permutation procedure to control the supervised effect of PLS (Indahl, 2005; Indahl et al., 2009; Liland and Indahl, 2009; Krishnan et al., 2011). Briefly, the genotypes were reshuffled randomly among the samples for 1000 times, followed by the same procedure as above to calculate correlations between the actual and predicted genotypes. The null distribution of correlations can be thus established based on the permutation sets. The corrected PLS-based permutation *P* value was generated by ranking the empirical raw correlation against the null distribution.

For the PSD permutation among genotypes (Peng et al., 2013), we calculated the Euclidean distances between the mean shapes of any two genotype groups. The mean shape can be denoted as a vector,

$$s = [x_1, y_1, z_1, \ldots, x_n, y_n, z_n] \quad (1)$$

where x_i, y_i, z_i were the X, Y, Z coordinate values of the *i*th points, and n was the number of point.

For a targeted mean shape s and reference mean shape s', the PSD was defined as

$$PSD = \frac{\sum_{i=1}^{n} d_i^2}{3n} \quad (2)$$

where $d_i^2 = (x_i - x_i')^2 + (y_i - y_i')^2 + (z_i - z_i')^2$.

We also randomly permuted the genotypes among the samples and then calculated the PSD between the peudo-genotype groups. The PSD scores resulted from permutation formed the null distribution. The one-side *P* value was calculated by the proportion of permuted PSD smaller than or equal to the true PSD.

Acknowledgments

We are grateful to all the volunteers taking part in this study. We thank Chen Liu, Nianhao Cheng for the assistance in sample collection; thank Wei Qian for PLS-based methods discussion; thank Bertram Müller-Myhsok, Benno Pütz for non-linear methods discussion; thank David Allen Hughes Jr. for manuscript polishing and comments. We are grateful for the experimental support of the Uli Schwarz Quantitative Biology Core Facility in PICB, thank Dr. Yujie Chen for assistance with this study. This work was funded by the Max-Planck-Gesellschaft Partner Group Grant (KT), the National Natural Science Foundation of China (Nos. 31371267, 31322030, 91331108 (KT); 91731303, 31771388, and 31711530221 (SX); 91631307 (SW); 31501011 (YL) and 31260263 (YG)). This work was supported by Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) (XDB13040100, SX; XDB13041000, SW), the National Science Fund for Distinguished Young Scholars (31525014, SX), the Program of Shanghai Academic Research Leader (16XD1404700, to SX), the support of a National Thousand Young Talents Award and a Max Planck-CAS Paul Gerson Unna Independent Research Group Leadership Award (SW), the Science and Technology Commission of Shanghai Municipality (16JC1400504, SW; 14YF1406800, YL; 16YF1413900, HL).

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jgg.2018.07.009.

References

Adhikari, K., Fuentes-Guajardo, M., Quinto-Sanchez, M., Mendoza-Revilla, J., Camilo Chacon-Duque, J., Acuna-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R.B., Perez, G.M., Gomez-Valdes, J., Villamil-Ramirez, H., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C.C., Hurtado, M., Villegas, V., Granja, V., Gallo, C., Poletti, G., Schuler-Faccini, L., Salzano, F.M., Bortolini, M.C., Canizales-Quinteros, S., Cheeseman, M., Rosique, J., Bedoya, G., Rothhammer, F., Headon, D., Gonzalez-Jose, R., Balding, D., Ruiz-Linares, A., 2016. A genome-wide association scan implicates *DCHS2*, *RUNX2*, *GLI3*, *PAX1* and *EDAR* in human facial variation. Nat. Commun. 7, 11616.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate : a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 57, 289–300.

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control, C., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., Daly, M.J., Price, A.L., Neale, B.M., 2015. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241.

Claes, P., Hill, H., Shriver, M.D., 2014. Toward DNA-based facial composites: preliminary results and validation. Forensic. Sci. Int. Genet. 13, 208–216.

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., Mio, W., Prochazkova, M., Williams, T., Li, H., Jones, K.L., Klein, O.D., Santorico, S.A., Hallgrimsson, B., Spritz, R.A., 2016. Genomewide association study of African children identifies association of *SCHIP1* and *PDE8A* with facial size and shape. PLoS Genet. 12, 1–19.

Delaneau, O., Marchini, J., McVeanh, G.A., Donnelly, P., Lunter, G., Marchini, J.L., Myers, S., Gupta-Hinch, A., Iqbal, Z., Mathieson, I., Rimmer, A., Xifara, D.K., Kerasidou, A., Churchhouse, C., Altshuler, D.M., Gabriel, S.B., Lander, E.S., Gupta, N., Daly, M.J., DePristo, M.A., Banks, E., Bhatia, G., Carneiro, M.O., Del Angel, G., Genovese, G., Handsaker, R.E., Hartl, C., McCarroll, S.A., Nemesh, J.C., Poplin, R.E., Schaffner, S.F., Shakir, K., Sabeti, P.C., Grossman, S.R., Tabrizi, S., Tariyal, R., Li, H., Reich, D., Durbin, R.M., Hurles, M.E., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T.M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Ayub, Q., Chen, Y., Coffey, A.J., Colonna, V., Huang, N., Jostins, L., Scally, A., Walter, K., Xue, Y., Zhang, Y., Blackburne, B., Lindsay, S.J., Ning, Z., Frankish, A., Harrow, J., Chris, T.S., Abecasis, G.R., Kang, H.M., Anderson, P., Blackwell, T., Busonero, F., Fuchsberger, C., Jun, G., Maschio, A., Porcu, E., Sidore, C., Tan, A., Trost, M.K., Bentley, D.R., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Bauer, M., Cheetham, R.K., Cox, T., Eberle, M., Murray, L., Shaw, R., Chakravarti, A., Clark, A.G., Keinan, A., Rodriguez-Flores, J.L., De LaVega, F.M., Degenhardt, J., Eichler, E.E., Flicek, P., Clarke, L., Leinonen, R., Smith, R.E., Zheng-Bradley, X., Beal, K., Cunningham, F., Herrero, J., McLaren, W.M., Ritchie, G.R.S., Barker, J., Kelman, G., Kulesha, E., Radhakrishnan, R., Roa, A., Smirnov, D., Streeter, I., Toneva, I., Gibbs, R.A., Dinh, H., Kovar, C., Lee, S., Lewis, L., Muzny, D., Reid, J., Wang, M., Yu, F., Bainbridge, M., Challis, D., Evani, U.S., Lu, J., Nagaswamy, U., Sabo, A., Wang, Y., Yu, J., Fowler, G., Hale, W., Kalra, D., Green, E.D., Knoppers, B.M., Korbel, J.O., Rausch, T., Sttz, A.M., Lee, C., Griffin, L., Hsieh, C.H., Mills, R.E., Von Grotthuss, M., Zhang, C., Shi, X., Lehrach, H., Sudbrak, R., Amstislavskiy, V.S., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M.L., Herwig, S.R., Mardis, E.R., Wilson, R.K., Fulton, L., Fulton, R., Weinstock, G.M., Chinwalla, A., Ding, L., Dooling, D., Koboldt, D.C., McLellan, M.D., Wallis, J.W., Wendl, M.C., Zhang, Q., Marth, G.T., Garrison, E.P., Kural, D., Lee, W.P., Leong, W.F., Ward, A.N., Wu, J., Zhang, M., Nickerson, D.A., Alkan, C., Hormozdiari, F., Ko, A., Sudmant, P.H., Schmidt, J.P., Davies, C.J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Sherry, S.T., Xiao, C., Church, D., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., Wang, J., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Li, G., Li, J., Li, Y., Liu, X., Lu, Y., Ma, X., Tai, S., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Yin, Y., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Lachlan, H., Fang, L., Li, Q., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Wang, B., Xie, Y., Ye, C., Yu, C., Zheng, H., Zhu, H., Cai, H., Cao, H., Su, Y., Tian, Z., Yang, H., Yang, L., Zhu, J., Cai, Z., Wang, J., Albrecht, M.W., Borodina, T.A., Auton, A., Yoon, S.C., Lihm, J., Makarov, V., Jin, H., Kim, W., Kim, K.C., Gottipati, S., Jones, D., Cooper, D.N., Ball, E.V., Stenson, P.D., Barnes, B., Kahn, S., Ye, K., Batzer, M.A., Konkel, M.K., Walker, J.A., MacArthur, D.G., Lek, M., Shriver, M.D., Bustamante, C.D., Gravel, S., Kenny, E.E., Kidd, J.M., Lacroute, P., Maples, B.K., Moreno-Estrada, A., Zakharia, F., Henn, B., Sandoval, K., Byrnes, J.K., Halperin, E., Baran, Y., Craig, D.W., Christoforides, A., Izatt, T., Kurdoglu, A.A., Sinari, S.A., Homer, N., Squire, K., Sebat, J., Bafna, V., Ye, K., Burchard, E.G., Hernandez, R.D., Gignoux, C.R., Haussler, D., Katzman, S.J., Kent, W.J., Howie, B., Ruiz-Linares, A., Dermitzakis, E.T., Lappalainen, T., Devine, S.E., Liu, X., Maroo, A., Tallon, L.J., Rosenfeld, J.A., Michelson, L.P., Angius, A., Cucca, F., Sanna, S., Bigham, A., Jones, C., Reinier, F., Li, Y., Lyons, R., Schlessinger, D., Awadalla, P., Hodgkinson, A., Oleksyk, T.K., Martinez-Cruzado, J.C., Fu, Y., Liu, X., Xiong, M., Jorde, L., Witherspoon, D., Xing, J., Browning, B.L., Hajirasouliha, I., Chen, K.,

Albers, C.A., Gerstein, M.B., Abyzov, A., Chen, J., Fu, Y., Habegger, L., Harmanci, A.O., Mu, X.J., Sisu, C., Balasubramanian, S., Jin, M., Khurana, E., Clarke, D., Michaelson, J.J., OSullivan, C., Barnes, K.C., Gharani, N., Toji, L.H., Gerry, N., Kaye, J.S., Kent, A., Mathias, R., Ossorio, P.N., Parker, M., Rotimi, C.N., Royal, C.D., Tishkoff, S., Via, M., Bodmer, W., Bedoya, G., Yang, G., You, C.J., Garcia-Montero, A., Orfao, A., Dutil, J., Brooks, L.D., Felsenfeld, A.L., McEwen, J.E., Clemm, N.C., Guyer, M.S., Peterson, J.L., Duncanson, A., Dunn, M., Peltonen, L., 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat. Commun. 5, 3934.

Dick, K.J., Nelson, C.P., Tsaprouni, L., Sandling, J.K., Aïssi, D., Wahl, S., Meduri, E., Morange, P.-E., Gagnon, F., Grallert, H., Waldenberger, M., Peters, A., Erdmann, J., Hengstenberg, C., Cambien, F., Goodall, A.H., Ouwehand, W.H., Schunkert, H., Thompson, J.R., Spector, T.D., Gieger, C., Trégouët, D.-A., Deloukas, P., Samani, N.J., 2014. DNA methylation and body-mass index: a genome-wide analysis. Lancet 383, 1990–1998.

Geiger, M., Li, R.H., Friess, W., 2003. Collagen sponges for bone regeneration with rhBMP-2. Adv. Drug Deliv. Rev. 55, 1613–1629.

Guo, J., Mei, X., Tang, K., 2013. Automatic landmark annotation and dense correspondence registration for 3D human facial images. BMC Bioinf. 14, 232.

Guo, J., Tan, J., Yang, Y., Zhou, H., Hu, S., Hashan, A., Bahaxar, N., Xu, S., Weaver, T.D., Jin, L., Stoneking, M., Tang, K., 2014. Variation and signatures of selection on the human face. J. Hum. Evol. 75, 143–152.

Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., 2014. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). CA Sage, Thousand Oaks. ISBN 97814.

Hoff-Olsen, P., Mevåg, B., Staalstrøm, E., Hovde, B., Egeland, T., Olaisen, B., 1999. Extraction of DNA from decomposed human tissue: an evaluation of five extraction methods for short tandem repeat typing. Forensic Sci. Int. 105, 171–183.

Hopman, S.M., Merks, J.H., Suttie, M., Hennekam, R.C., Hammond, P., 2014. Face shape differs in phylogenetically related populations. Eur. J. Hum. Genet. 22, 1268–1271.

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, 1–15.

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

Hu, S., Xiong, J., Fu, P., Qiao, L., Tan, J., Jin, L., Tang, K., 2017. Signatures of personality on dense 3D facial images. Sci. Rep. 7, 73.

Indahl, U., 2005. A twist to partial least squares regression. J. Chemom. 19, 32–44.

Indahl, U.G., Liland, K.H., Næs, T., 2009. Canonical partial least squares-a unified PLS approach to classification and regression problems. J. Chemom. 23, 495–504.

Kamberov, Y.G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., Yang, Y., Li, S., Tang, K., Chen, H., Powell, A., Itan, Y., Fuller, D., Lohmueller, J., Mao, J., Schachar, A., Paymer, M., Hostetter, E., Byrne, E., Burnett, M., McMahon, A.P., Thomas, M.G., Lieberman, D.E., Jin, L., Tabin, C.J., Morgan, B.A., Sabeti, P.C., 2013. Modeling recent human evolution in mice by expression of a selected *EDAR* variant. Cell 152, 691–702.

Kilpelainen, T.O., Carli, J.F., Skowronski, A.A., Sun, Q., Kriebel, J., Feitosa, M.F., Hedman, A.K., Drong, A.W., Hayes, J.E., Zhao, J., Pers, T.H., Schick, U., Grarup, N., Kutalik, Z., Trompet, S., Mangino, M., Kristiansson, K., Beekman, M., Lyytikainen, L.P., Eriksson, J., Henneman, P., Lahti, J., Tanaka, T., Luan, J., Del Greco, M.F., Pasko, D., Renstrom, F., Willems, S.M., Mahajan, A., Rose, L.M., Guo, X., Liu, Y., Kleber, M.E., Perusse, L., Gaunt, T., Ahluwalia, T.S., Ju Sung, Y., Ramos, Y.F., Amin, N., Amuzu, A., Barroso, I., Bellis, C., Blangero, J., Buckley, B.M., Bohringer, S., YD, I.C., de Craen, A.J., Crosslin, D.R., Dale, C.E., Dastani, Z., Day, F.R., Deelen, J., Delgado, G.E., Demirkan, A., Finucane, F.M., Ford, I., Garcia, M.E., Gieger, C., Gustafsson, S., Hallmans, G., Hankinson, S.E., Havulinna, A.S., Herder, C., Hernandez, D., Hicks, A.A., Hunter, D.J., Illig, T., Ingelsson, E., Ioan-Facsinay, A., Jansson, J.O., Jenny, N.S., Jorgensen, M.E., Jorgensen, T., Karlsson, M., Koenig, W., Kraft, P., Kwekkeboom, J., Laatikainen, T., Ladwig, K.H., LeDuc, C.A., Lowe, G., Lu, Y., Marques-Vidal, P., Meisinger, C., Menni, C., Morris, A.P., Myers, R.H., Mannisto, S., Nalls, M.A., Paternoster, L., Peters, A., Pradhan, A.D., Rankinen, T., Rasmussen-Torvik, L.J., Rathmann, W., Rice, T.K., Brent Richards, J., Ridker, P.M., Sattar, N., Savage, D.B., Soderberg, S., Timpson, N.J., Vandenput, L., van Heemst, D., Uh, H.W., Vohl, M.C., Walker, M., Wichmann, H.E., Widen, E., Wood, A.R., Yao, J., Zeller, T., Zhang, Y., Meulenbelt, I., Kloppenburg, M., Astrup, A., Sorensen, T.I., Sarzynski, M.A., Rao, D.C., Jousilahti, P., Vartiainen, E., Hofman, A., Rivadeneira, F., Uitterlinden, A.G., Kajantie, E., Osmond, C., Palotie, A., Eriksson, J.G., Heliovaara, M., Knekt, P.B., Koskinen, S., Jula, A., Perola, M., Huupponen, R.K., Viikari, J.S., Kahonen, M., Lehtimaki, T., Raitakari, O.T., Mellstrom, D., Lorentzon, M., Casas, J.P., Bandinelli, S., Marz, W., Isaacs, A., van Dijk, K.W., van Duijn, C.M., Harris, T.B., Bouchard, C., Allison, M.A., Chasman, D.I., Ohlsson, C., Lind, L., Scott, R.A., Langenberg, C., Wareham, N.J., Ferrucci, L., Frayling, T.M., Pramstaller, P.P., Borecki, I.B., Waterworth, D.M., Bergmann, S., Waeber, G., Vollenweider, P., Vestergaard, H., Hansen, T., Pedersen, O., Hu, F.B., Eline Slagboom, P., Grallert, H., Spector, T.D., Jukema, J.W., Klein, R.J., Schadt, E.E., Franks, P.W., Lindgren, C.M., Leibel, R.L., Loos, R.J., 2016. Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. Nat. Commun. 7, 10494.

Krishnan, A., Williams, L.J., McIntosh, A.R., Abdi, H., 2011. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage 56, 455–475.

Liland, K.H., Indahl, U.G., 2009. Powered partial least squares discriminant analysis. J. Chemom. 23, 7–18.

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., van der Lugt, A., Rivadeneira, F., Uitterlinden, A.G., Hofman, A., Niessen, W.J., Homuth, G., de Zubicaray, G., McMahon, K.L., Thompson, P.M., Daboul, A., Puls, R., Hegenscheid, K., Bevan, L., Pausova, Z., Medland, S.E., Montgomery, G.W., Wright, M.J., Wicking, C., Boehringer, S., Spector, T.D., Paus, T., Martin, N.G., Biffar, R., Kayser, M., 2012. A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genet. 8, 1–13.

Liu, F., Hamer, M.A., Deelen, J., Lall, J.S., Jacobs, L., van Heemst, D., Murray, P.G., Wollstein, A., de Craen, A.J., Uh, H.W., Zeng, C., Hofman, A., Uitterlinden, A.G., Houwing-Duistermaat, J.J., Pardo, L.M., Beekman, M., Slagboom, P.E., Nijsten, T., Kayser, M., Gunn, D.A., 2016. The *MC1R* gene and Youthful looks. Curr. Biol. 26, 1213–1220.

Loos, R.J., 2016. *CREBRF* variant increases obesity risk and protects against diabetes in Samoans. Nat. Genet. 48, 976–978.

Loos, R.J., Janssens, A.C., 2017. Predicting polygenic obesity using genetic information. Cell Metabol. 25, 535–543.

Lu, Y., Day, F.R., Gustafsson, S., Buchkovich, M.L., Na, J., Bataille, V., Cousminer, D.L., Dastani, Z., Drong, A.W., Esko, T., Evans, D.M., Falchi, M., Feitosa, M.F., Ferreira, T., Hedman, A.K., Haring, R., Hysi, P.G., Iles, M.M., Justice, A.E., Kanoni, S., Lagou, V., Li, R., Li, X., Locke, A., Lu, C., Magi, R., Perry, J.R., Pers, T.H., Qi, Q., Sanna, M., Schmidt, E.M., Scott, W.R., Shungin, D., Teumer, A., Vinkhuyzen, A.A., Walker, R.W., Westra, H.J., Zhang, M., Zhang, W., Zhao, J.H., Zhu, Z., Afzal, U., Ahluwalia, T.S., Bakker, S.J., Bellis, C., Bonnefond, A., Borodulin, K., Buchman, A.S., Cederholm, T., Choh, A.C., Choi, H.J., Curran, J.E., de Groot, L.C., De Jager, P.L., Dhonukshe-Rutten, R.A., Enneman, A.W., Eury, E., Evans, D.S., Forsen, T., Friedrich, N., Fumeron, F., Garcia, M.E., Gartner, S., Han, B.G., Havulinna, A.S., Hayward, C., Hernandez, D., Hillege, H., Ittermann, T., Kent, J.W., Kolcic, I., Laatikainen, T., Lahti, J., Mateo Leach, I., Lee, C.G., Lee, J.Y., Liu, T., Liu, Y., Lobbens, S., Loh, M., Lyytikainen, L.P., Medina-Gomez, C., Michaelsson, K., Nalls, M.A., Nielson, C.M., Oozageer, L., Pascoe, L., Paternoster, L., Polasek, O., Ripatti, S., Sarzynski, M.A., Shin, C.S., Narancic, N.S., Spira, D., Srikanth, P., Steinhagen-Thiessen, E., Sung, Y.J., Swart, K.M., Taittonen, L., Tanaka, T., Tikkanen, E., van der Velde, N., van Schoor, N.M., Verweij, N., Wright, A.F., Yu, L., Zmuda, J.M., Eklund, N., Forrester, T., Grarup, N., Jackson, A.U., Kristiansson, K., Kuulasmaa, T., Kuusisto, J., Lichtner, P., Luan, J., Mahajan, A., Mannisto, S., Palmer, C.D., Ried, J.S., Scott, R.A., Stancakova, A., Wagner, P.J., Demirkan, A., Doring, A., Gudnason, V., Kiel, D.P., Kuhnel, B., Mangino, M., McKnight, B., Menni, C., O'Connell, J.R., Oostra, B.A., Shuldiner, A.R., Song, K., Vandenput, L., van Duijn, C.M., Vollenweider, P., White, C.C., Boehnke, M., Boettcher, Y., Cooper, R.S., Forouhi, N.G., Gieger, C., Grallert, H., Hingorani, A., Jorgensen, T., Jousilahti, P., Kivimaki, M., Kumari, M., Laakso, M., Langenberg, C., Linneberg, A., Luke, A., McKenzie, C.A., Palotie, A., Pedersen, O., Peters, A., Strauch, K., Tayo, B.O., Wareham, N.J., Bennett, D.A., Bertram, L., Blangero, J., Bluher, M., Bouchard, C., Campbell, H., Cho, N.H., Cummings, S.R., Czerwinski, S.A., Demuth, I., Eckardt, R., Eriksson, J.G., Ferrucci, L., Franco, O.H., Froguel, P., Gansevoort, R.T., Hansen, T., Harris, T.B., Hastie, N., Heliovaara, M., Hofman, A., Jordan, J.M., Jula, A., Kahonen, M., Kajantie, E., Knekt, P.B., Koskinen, S., Kovacs, P., Lehtimaki, T., Lind, L., Liu, Y., Orwoll, E.S., Osmond, C., Perola, M., Perusse, L., Raitakari, O.T., Rankinen, T., Rao, D.C., Rice, T.K., Rivadeneira, F., Rudan, I., Salomaa, V., Sorensen, T.I., Stumvoll, M., Tonjes, A., Towne, B., Tranah, G.J., Tremblay, A., Uitterlinden, A.G., van der Harst, P., Vartiainen, E., Viikari, J.S., Vitart, V., Vohl, M.C., Volzke, H., Walker, M., Wallaschofski, H., Wild, S., Wilson, J.F., Yengo, L., Bishop, D.T., Borecki, I.B., Chambers, J.C., Cupples, L.A., Dehghan, A., Deloukas, P., Fatemifar, G., Fox, C., Furey, T.S., Franke, L., Han, J., Hunter, D.J., Karjalainen, J., Karpe, F., Kaplan, R.C., Kooner, J.S., McCarthy, M.I., Murabito, J.M., Morris, A.P., Bishop, J.A., North, K.E., Ohlsson, C., Ong, K.K., Prokopenko, I., Richards, J.B., Schadt, E.E., Spector, T.D., Widen, E., Willer, C.J., Yang, J., Ingelsson, E., Mohlke, K.L., Hirschhorn, J.N., Pospisilik, J.A., Zillikens, M.C., Lindgren, C., Kilpelainen, T.O., Loos, R.J., 2016. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. Nat. Commun. 7, 10495.

Myles, S., Somel, M., Tang, K., Kelso, J., Stoneking, M., 2007. Identifying genes underlying skin pigmentation differences among human populations. Hum. Genet. 120, 613–621.

Ng, M.C.Y., Graff, M., Lu, Y., Justice, A.E., Mudgal, P., Liu, C.T., Young, K., Yanek, L.R., Feitosa, M.F., Wojczynski, M.K., Rand, K., Brody, J.A., Cade, B.E., Dimitrov, L., Duan, Q., Guo, X., Lange, L.A., Nalls, M.A., Okut, H., Tajuddin, S.M., Tayo, B.O., Vedantam, S., Bradfield, J.P., Chen, G., Chen, W.M., Chesi, A., Irvin, M.R., Padhukasahasram, B., Smith, J.A., Zheng, W., Allison, M.A., Ambrosone, C.B., Bandera, E.V., Bartz, T.M., Berndt, S.I., Bernstein, L., Blot, W.J., Bottinger, E.P., Carpten, J., Chanock, S.J., Chen, Y.I., Conti, D.V., Cooper, R.S., Fornage, M., Freedman, B.I., Garcia, M., Goodman, P.J., Hsu, Y.H., Hu, J., Huff, C.D., Ingles, S.A., John, E.M., Kittles, R., Klein, E., Li, J., McKnight, B., Nayak, U., Nemesure, B., Ogunniyi, A., Olshan, A., Press, M.F., Rohde, R., Rybicki, B.A., Salako, B., Sanderson, M., Shao, Y., Siscovick, D.S., Stanford, J.L., Stevens, V.L., Stram, A., Strom, S.S., Vaidya, D., Witte, J.S., Yao, J., Zhu, X., Ziegler, R.G., Zonderman, A.B., Adeyemo, A., Ambs, S., Cushman, M., Faul, J.D., Hakonarson, H., Levin, A.M., Nathanson, K.L., Ware, E.B., Weir, D.R., Zhao, W., Zhi, D., Bone Mineral Density in Childhood Study, G., Arnett, D.K., Grant, S.F.A., Kardia, S.L.R., Oloapde, O.I., Rao, D.C., Rotimi, C.N., Sale, M.M., Williams, L.K., Zemel, B.S., Becker, D.M., Borecki, I.B., Evans, M.K., Harris, T.B., Hirschhorn, J.N., Li, Y., Patel, S.R., Psaty, B.M., Rotter, J.I., Wilson, J.G., Bowden, D.W., Cupples, L.A., Haiman, C.A., Loos, R.J.F., North, K.E., 2017. Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of

African ancestry: African Ancestry Anthropometry Genetics Consortium. PLoS Genet. 13, 1–25.
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., Richmond, S., Evans, D.M., 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.
Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet. 2, 2074–2093.
Peng, S., Tan, J., Hu, S., Zhou, H., Guo, J., Jin, L., Tang, K., 2013. Detecting genetic association of common human facial morphological variation using high density 3D image registration. PLoS Comput. Biol. 9, 1–11.
Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909.
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., 2010. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., 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.
Rahimov, F., Marazita, M.L., Visel, A., Cooper, M.E., Hitchler, M.J., Rubini, M., Domann, F.E., Govil, M., Christensen, K., Bille, C., Melbye, M., Jugessur, A., Lie, R.T., Wilcox, A.J., Fitzpatrick, D.R., Green, E.D., Mossey, P.A., Little, J., Steegers-Theunissen, R.P., Pennacchio, L.A., Schutte, B.C., Murray, J.C., 2008. Disruption of an AP-2alpha binding site in an *IRF6* enhancer is associated with cleft lip. Nat. Genet. 40, 1341–1347.
Relethford, J.H., 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. Am. J. Phys. Anthropol. 118, 393–398.
Relethford, J., 2004. Global patterns of isolation by distance based on genetic and morphological data. Hum. Biol. 76, 499–513.
Ried, J.S., Jeff, M.J., Chu, A.Y., Bragg-Gresham, J.L., van Dongen, J., Huffman, J.E., Ahluwalia, T.S., Cadby, G., Eklund, N., Eriksson, J., Esko, T., Feitosa, M.F., Goel, A., Gorski, M., Hayward, C., Heard-Costa, N.L., Jackson, A.U., Jokinen, E., Kanoni, S., Kristiansson, K., Kutalik, Z., Lahti, J., Luan, J., Magi, R., Mahajan, A., Mangino, M., Medina-Gomez, C., Monda, K.L., Nolte, I.M., Perusse, L., Prokopenko, I., Qi, L., Rose, L.M., Salvi, E., Smith, M.T., Snieder, H., Stancakova, A., Ju Sung, Y., Tachmazidou, I., Teumer, A., Thorleifsson, G., van der Harst, P., Walker, R.W., Wang, S.R., Wild, S.H., Willems, S.M., Wong, A., Zhang, W., Albrecht, E., Couto Alves, A., Bakker, S.J., Barlassina, C., Bartz, T.M., Beilby, J., Bellis, C., Bergman, R.N., Bergmann, S., Blangero, J., Bluher, M., Boerwinkle, E., Bonnycastle, L.L., Bornstein, S.R., Bruinenberg, M., Campbell, H., Chen, Y.I., Chiang, C.W., Chines, P.S., Collins, F.S., Cucca, F., Cupples, L.A., D'Avila, F., de Geus, E.J., Dedoussis, G., Dimitriou, M., Doring, A., Eriksson, J.G., Farmaki, A.E., Farrall, M., Ferreira, T., Fischer, K., Forouhi, N.G., Friedrich, N., Gjesing, A.P., Glorioso, N., Graff, M., Grallert, H., Grarup, N., Grassler, J., Grewal, J., Hamsten, A., Harder, M.N., Hartman, C.A., Hassinen, M., Hastie, N., Hattersley, A.T., Havulinna, A.S., Heliovaara, M., Hillege, H., Hofman, A., Holmen, O., Homuth, G., Hottenga, J.J., Hui, J., Husemoen, L.L., Hysi, P.G., Isaacs, A., Ittermann, T., Jalilzadeh, S., James, A.L., Jorgensen, T., Jousilahti, P., Jula, A., Marie Justesen, J., Justice, A.E., Kahonen, M., Karaleftheri, M., Tee Khaw, K., Keinanen-Kiukaanniemi, S.M., Kinnunen, L., Knekt, P.B., Koistinen, H.A., Kolcic, I., Kooner, I.K., Koskinen, S., Kovacs, P., Kyriakou, T., Laitinen, T., Langenberg, C., Lewin, A.M., Lichtner, P., Lindgren, C.M., Lindstrom, J., Linneberg, A., Lorbeer, R., Lorentzon, M., Luben, R., Lyssenko, V., Mannisto, S., Manunta, P., Leach, I.M., McArdle, W.L., McKnight, B., Mohlke, K.L., Mihailov, E., Milani, L., Mills, R., Montasser, M.E., Morris, A.P., Muller, G., Musk, A.W., Narisu, N., Ong, K.K., Oostra, B.A., Osmond, C., Palotie, A., Pankow, J.S., Paternoster, L., Penninx, B.W., Pichler, I., Pilia, M.G., Polasek, O., Pramstaller, P.P., Raitakari, O.T., Rankinen, T., Rao, D.C., Rayner, N.W., Ribel-Madsen, R., Rice, T.K., Richards, M., Ridker, P.M., Rivadeneira, F., Ryan, K.A., Sanna, S., Sarzynski, M.A., Scholtens, S., Scott, R.A., Sebert, S., Southam, L., Sparso, T.H., Steinthorsdottir, V., Stirrups, K., Stolk, R.P., Strauch, K., Stringham, H.M., Swertz, M.A., Swift, A.J., Tonjes, A., Tsafantakis, E., van der Most, P.J., Van Vliet-Ostaptchouk, J.V., Vandenput, L., Vartiainen, E., Venturini, C., Verweij, N., Viikari, J.S., Vitart, V., Vohl, M.C., Vonk, J.M., Waeber, G., Widen, E., Willemsen, G., Wilsgaard, T., Winkler, T.W., Wright, A.F., Yerges-Armstrong, L.M., Hua Zhao, J., Zillikens, M.C., Boomsma, D.I., Bouchard, C., Chambers, J.C., Chasman, D.I., Cusi, D., Gansevoort, R.T., Gieger, C., Hansen, T., Hicks, A.A., Hu, F., Hveem, K., Jarvelin, M.R., Kajantie, E., Kooner, J.S., Kuh, D., Kuusisto, J., Laakso, M., Lakka, T.A., Lehtimaki, T., Metspalu, A., Njolstad, I., Ohlsson, C., Oldehinkel, A.J., Palmer, L.J., Pedersen, O., Perola, M., Peters, A., Psaty, B.M., Puolijoki, H., Rauramaa, R., Rudan, I., Salomaa, V., Schwarz, P.E., Shudiner, A.R., Smit, J.H., Sorensen, T.I., Spector, T.D., Stefansson, K., Stumvoll, M., Tremblay, A., Tuomilehto, J., Uitterlinden, A.G., Uusitupa, M., Volker, U., Vollenweider, P., Wareham, N.J., Watkins, H., Wilson, J.F., Zeggini, E., Abecasis, G.R., Boehnke, M., Borecki, I.B., Deloukas, P., van Duijn, C.M., Fox, C., Groop, L.C., Heid, I.M., Hunter, D.J., Kaplan, R.C., McCarthy, M.I., North, K.E., O'Connell, J.R., Schlessinger, D., Thorsteinsdottir, U., Strachan, D.P., Frayling, T., Hirschhorn, J.N., Muller-Nurasyid, M., Loos, R.J., 2016. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. Nat. Commun. 7, 13357.
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., Moreno, L.M., Wehby, G.L., Murray, J.C., Laurie, C.A., Laurie, C.C., Cole, J., Ferrara, T., Santorico, S., Klein, O., Mio, W., Feingold, E., Hallgrimsson, B., Spritz, R.A., Marazita, M.L., Weinberg, S.M., 2016. Genome-wide association study reveals multiple loci influencing normal human facial morphology. PLoS Genet. 12, 1–21.
Smeets, D., Claes, P., Vandermeulen, D., Clement, J.G., 2010. Objective 3D face recognition: evolution, approaches and challenges. Forensic Sci. Int. 201, 125–132.
Talbert, L., Kau, C.H., Christou, T., Vlachos, C., Souccar, N., 2014. A 3D analysis of Caucasian and African American facial morphologies in a US population. J. Orthod. 41, 19–29.
Turner, D., S, 2018. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J. Open Source Softw. 3, 731.
Weinberg, S.M., Parsons, T.E., Marazita, M.L., Maher, B.S., 2013. Heritability of face shape in twins: a preliminary study using 3D stereophotogrammetry and geometric morphometrics. Inside Dent. 3000, 1, a004.
Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., Amin, N., Buchkovich, M.L., Croteau-Chonka, D.C., Day, F.R., Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A.U., Karjalainen, J., Lo, K.S., Locke, A.E., Magi, R., Mihailov, E., Porcu, E., Randall, J.C., Scherag, A., Vinkhuyzen, A.A., Westra, H.J., Winkler, T.W., Workalemahu, T., Zhao, J.H., Absher, D., Albrecht, E., Anderson, D., Baron, J., Beekman, M., Demirkan, A., Ehret, G.B., Feenstra, B., Feitosa, M.F., Fischer, K., Fraser, R.M., Goel, A., Gong, J., Justice, A.E., Kanoni, S., Kleber, M.E., Kristiansson, K., Lim, U., Lotay, V., Lui, J.C., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Nalls, M.A., Nyholt, D.R., Palmer, C.D., Pasko, D., Pechlivanis, S., Prokopenko, I., Ried, J.S., Ripke, S., Shungin, D., Stancakova, A., Strawbridge, R.J., Sung, Y.J., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S.W., van Setten, J., Van Vliet-Ostaptchouk, J.V., Wang, Z., Yengo, L., Zhang, W., Afzal, U., Arnlov, J., Arscott, G.M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A.J., Berne, C., Bluher, M., Bolton, J.L., Bottcher, Y., Boyd, H.A., Bruinenberg, M., Buckley, B.M., Buyske, S., Caspersen, I.H., Chines, P.S., Clarke, R., Claudi-Boehm, S., Cooper, M., Daw, E.W., De Jong, P.A., Deelen, J., Delgado, G., Denny, J.C., Dhonukshe-Rutten, R., Dimitriou, M., Doney, A.S., Dorr, M., Eklund, N., Eury, E., Folkersen, L., Garcia, M.E., Geller, F., Giedraitis, V., Go, A.S., Grallert, H., Grammer, T.B., Grassler, J., Gronberg, H., de Groot, L.C., Groves, C.J., Haessler, J., Hall, P., Haller, T., Hallmans, G., Hannemann, A., Hartman, C.A., Hassinen, M., Hayward, C., Heard-Costa, N.L., Helmer, Q., Hemani, G., Henders, A.K., Hillege, H.L., Hlatky, M.A., Hoffmann, W., Hoffmann, P., Holmen, O., Houwing-Duistermaat, J.J., Illig, T., Isaacs, A., James, A.L., Jeff, J., Johansen, B., Johansson, A., Jolley, J., Juliusdottir, T., Junttila, J., Kho, A.N., Kinnunen, L., Klopp, N., Kocher, T., Kratzer, W., Lichtner, P., Lind, L., Lindstrom, J., Lobbens, S., Lorentzon, M., Lu, Y., Lyssenko, V., Magnusson, P.K., Mahajan, A., Maillard, M., McArdle, W.L., McKenzie, C.A., McLachlan, S., McLaren, P.J., Menni, C., Merger, S., Milani, L., Moayyeri, A., Monda, K.L., Morken, M.A., Muller, G., Muller-Nurasyid, M., Musk, A.W., Narisu, N., Nauck, M., Nolte, I.M., Nothen, M.M., Oozageer, L., Pilz, S., Rayner, N.W., Renstrom, F., Robertson, N.R., Rose, L.M., Roussel, R., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F.R., Schunkert, H., Scott, R.A., Sehmi, J., Seufferlein, T., Shi, J., Silventoinen, K., Smit, J.H., Smith, A.V., Smolonska, J., Stanton, A.V., Stirrups, K., Stott, D.J., Stringham, H.M., Sundstrom, J., Swertz, M.A., Syvanen, A.C., Tayo, B.O., Thorleifsson, G., Tyrer, J.P., van Dijk, S., van Schoor, N.M., van der Velde, N., van Heemst, D., van Oort, F.V., Vermeulen, S.H., Verweij, N., Vonk, J.M., Waite, L.L., Waldenberger, M., Wennauer, R., Wilkens, L.R., Willenborg, C., Wilsgaard, T., Wojczynski, M.K., Wong, A., Wright, A.F., Zhang, Q., Arveiler, D., Bakker, S.J., Beilby, J., Bergman, R.N., Bergmann, S., Biffar, R., Blangero, J., Boomsma, D.I., Bornstein, S.R., Bovet, P., Brambilla, P., Brown, M.J., Campbell, H., Caulfield, M.J., Chakravarti, A., Collins, R., Collins, F.S., Crawford, D.C., Cupples, L.A., Danesh, J., de Faire, U., den Ruijter, H.M., Erbel, R., Erdmann, J., Eriksson, J.G., Farrall, M., Ferrannini, E., Ferrieres, J., Ford, I., Forouhi, N.G., Forrester, T., Gansevoort, R.T., Gejman, P.V., Gieger, C., Golay, A., Gottesman, O., Gudnason, V., Gyllensten, U., Haas, D.W., Hall, A.S., Harris, T.B., Hattersley, A.T., Heath, A.C., Hengstenberg, C., Hicks, A.A., Hindorff, L.A., Hingorani, A.D., Hofman, A., Hovingh, G.K., Humphries, S.E., Hunt, S.C., Hypponen, E., Jacobs, K.B., Jarvelin, M.R., Jousilahti, P., Jula, A.M., Kaprio, J., Kastelein, J.J., Kayser, M., Kee, F., Keinanen-Kiukaanniemi, S.M., Kiemeney, L.A., Kooner, J.S., Kooperberg, C., Koskinen, S., Kovacs, P., Kraja, A.T., Kumari, M., Kuusisto, J., Lakka, T.A., Langenberg, C., Le Marchand, L., Lehtimaki, T., Lupoli, S., Madden, P.A., Mannisto, S., Manunta, P., Marette, A., Matise, T.C., McKnight, B., Meitinger, T., Moll, F.L., Montgomery, G.W., Morris, A.D., Morris, A.P., Murray, J.C., Nelis, M., Ohlsson, C., Oldehinkel, A.J., Ong, K.K., Ouwehand, W.H., Pasterkamp, G., Peters, A., Pramstaller, P.P., Price, J.F., Qi, L., Raitakari, O.T., Rankinen, T., Rao, D.C., Rice, T.K., Ritchie, M., Rudan, I., Salomaa, V., Samani, N.J., Saramies, J., Sarzynski, M.A., Schwarz, P.E., Sebert, S., Sever, P., Shuldiner, A.R., Sinisalo, J., Steinthorsdottir, V., Stolk, R.P., Tardif, J.C., Tonjes, A., Tremblay, A., Tremoli, E., Virtamo, J., Vohl, M.C., Electronic Medical, R., Genomics, C., Consortium, M.Ig, Consortium, P., LifeLines Cohort, S., Amouyel, P., Asselbergs, F.W., Assimes, T.L., Bochud, M., Boehm, B.O., Boerwinkle, E., Bottinger, E.P., Bouchard, C., Cauchi, S., Chambers, J.C., Chanock, S.J., Cooper, R.S., de Bakker, P.I., Dedoussis, G., Ferrucci, L., Franks, P.W., Froguel, P., Groop, L.C., Haiman, C.A., Hamsten, A., Hayes, M.G., Hui, J., Hunter, D.J., Hveem, K., Jukema, J.W., Kaplan, R.C., Kivimaki, M., Kuh, D., Laakso, M., Liu, Y., Martin, N.G., Marz, W., Melbye, M., Moebus, S., Munroe, P.B., Njolstad, I., Oostra, B.A., Palmer, C.N., Pedersen, N.L., Perola, M., Perusse, L., Peters, U., Powell, J.E., Power, C., Quertermous, T., Rauramaa, R., Reinmaa, E., Ridker, P.M., Rivadeneira, F., Rotter, J.I., Saaristo, T.E., Saleheen, D., Schlessinger, D., Slagboom, P.E., Snieder, H., Spector, T.D., Strauch, K., Stumvoll, M., Tuomilehto, J., Uusitupa, M., van der Harst, P., Volzke, H., Walker, M., Wareham, N.J., Watkins, H., Wichmann, H.E., Wilson, J.F., Zanen, P.,

Deloukas, P., Heid, I.M., Lindgren, C.M., Mohlke, K.L., Speliotes, E.K., Thorsteinsdottir, U., Barroso, I., Fox, C.S., North, K.E., Strachan, D.P., Beckmann, J.S., Berndt, S.I., Boehnke, M., Borecki, I.B., McCarthy, M.I., Metspalu, A., Stefansson, K., Uitterlinden, A.G., van Duijn, C.M., Franke, L., Willer, C.J., Price, A.L., Lettre, G., Loos, R.J., Weedon, M.N., Ingelsson, E., O'Connell, J.R., Abecasis, G.R., Chasman, D.I., Goddard, M.E., Visscher, P.M., Hirschhorn, J.N., Frayling, T.M., 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46, 1173–1186.

Xu, S., Jin, L., 2008. A genome-wide analysis of admixture in Uyghurs and a high-density admixture map for disease-gene discovery. Am. J. Hum. Genet. 83, 322–336.

Xu, S., Huang, W., Qian, J., Jin, L., 2008. Analysis of genomic admixture in Uyghur and its implication in mapping strategy. Am. J. Hum. Genet. 82, 883–894.

Xu, S., Jin, W., Jin, L., 2009. Haplotype-sharing analysis showing Uyghurs are unlikely genetic donors. Mol. Biol. Evol. 26, 2197–2206.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565–569.

Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., Hill, W.G., Landi, M.T., Alonso, A., Lettre, G., Lin, P., Ling, H., Lowe, W., Mathias, R.A., Melbye, M., Pugh, E., Cornelis, M.C., Weir, B.S., Goddard, M.E., Visscher, P.M., 2011. Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. 43, 519–525.