Genome-wide Variants of Eurasian Facial Shape Differentiation and a prospective model of DNA based Face Prediction

Lu Qiao[1], Yajun Yang[2,3], Pengcheng Fu[4], Sile Hu[1], Hang Zhou[1], Shouneng Peng[1], Jingze Tan[2,3], Yan Lu[1], Haiyi Lou[1], Dongsheng Lu[1], Sijie Wu[1], Jing Guo[1], Li Jin[1,2,3,6], Yaqun Guan[5], Sijia Wang[1,6,*], Shuhua Xu[1,6,7,*], Kun Tang[1,*]

SUPPLEMENTARY NOTE

This supplementary note mainly describes the methods and detailed results related to the prospective face prediction model.

GWA analyses on within-Uyghur phenotypes

To capture the majority of the common facial shape variance for prediction, additional phenotypes based on inter-landmark measurements, PCA and PLS were defined mainly for the within-Uyghur variations, in complement to the ancestry-divergent traits. For the landmark based phenotypes, other than the 10 ancestry-divergent phenotypes, the remaining 26 traits were defined as within-Uyghur traits. For PCA phenotypes, we carried out PCA on the UIG-D 3dDFM data for six facial features separately. The top three PCs (PC1, PC2, PC3) were used as within-Uyghur phenotypes that together explained ~60% variance. For PLS based traits, the models consist of the first component (PLS1) and the top two components (PLS2) were defined as the within-Uyghur phenotypes for each facial feature.

Hereto, the ancestry-divergent and within-Uyghur phenotypes summarized the facial features of Uyghurs. To reduce the computational burden and avoid imprecision in imputation, GWA analyses for facial prediction were carried out only on the genotyped SNPs in UIG-D.

Predicted face (PF) composition

For each of the top SNPs found in the ancestry-divergent GWAS and

within-Uyghur GWAS, we divided the 3dDFM images into three genotype groups, and the within-group mean shapes were obtained as μ'_g where g is the genotype. A residual face can be obtained for each genotype by subtracting the genotype mean by the global cohort mean shape as,

$$r_g = \mu'_g - \mu \qquad (1)$$

where μ was the global mean. We superimposed all the re-scaled residual faces of each SNP according to their true genotype information, onto the global mean to construct a face of prediction:

$$f = \mu + \alpha \sum r_{g(i)} \qquad (2)$$

where α was the effect coefficient and i is the SNP index. The coefficient α was proposed to maximize the similarity of the predicted and actual faces. Briefly, a PF was calculated for every individual in the discovery panel as described by equation (2), where α varies. PSD was used to calculate the distance between the PF and actual face for each individual. The value of α was determined when the minimum average PSD was achieved within a gender group. This same α was also used for the independent cohort as well as the randomly simulated faces. For the predicted model built by 277 top SNPs, the coefficient was determined to be 0.312 in females (α_f) and 0.252 in males (α_m) (Fig. S7).

Generation of random-genotype predicted faces (RGF) set

RGF was generated in the similar way as the PF, specifically by synthesis of a 3D face model via equation (2), based on a given set of genotypes. The only difference

was that a PF used an actual set of genotypes, whereas the genotype set for a RGF was randomly permuted from the known frequencies of genotypes in the combined UIG cohort .

Evaluation of the face prediction

For each sample, we described the degrees of similarity between predicted and actual face in the respects of shape space angle (SSA). The SSA was the angle between the predicted and actual faces in the 3 × 32,251 face-space. Either RGF or RAF was compared to PF to test whether the face prediction was significantly better than random, and two types of tests were carried out: (1) inter-distribution test: For a group of *N* individuals, we can generate a corresponding RGF/RAF set and derive the pairwise SSA distribution between the actual and random prediction faces, which is called the random SSA distribution. We repeat the RGF/RAF simulations ten times to obtain a much bigger and therefore more robust random SSA distribution. This was compared against the true SSA distribution between the PF and actual faces using Student's T test. (2) single-statistic test: For a group of *N* individuals, the average SSA score (SSA_{avg}) was calculated and used as the test statistic. Here, we named such similarity statistics between PF and actual face as true SSA_{avg}. For the case of RGF, we generated the same number (*N*) of RGF as described above and calculated the corresponding SSA_{avg}. This was repeated 1,000 times and the distributions of RGF SSA_{avg} under the null hypothesis were compared to the true SSA_{avg}. For RAF, as there was a limited number of actual face in the independent panel, we applied a

leave-one-out (LOO) procedure. For each iteration of random SSA_{avg} calculation, every individual in UIG-R was tested in turn. Its PF was generated first and its corresponding actual face was removed from the UIG-R actual faces panel. A RAF was randomly sampled from the remaining panel of actual faces without replacement and its SSA to the tested face was calculated and averaged over the *N* individuals. Doing so for *N*-1 times until the RAF panel was complete, the distributions of RAF SSA_{avg} under the null hypothesis was compared to the true SSA_{avg}. Assuming normality of the null distributions, the probability of the true statistics was either given as the percentage of random cases scored higher than the true ones if any, or the standard difference of the true statistics from the null expectation $t = (s - m)/\sigma$ was queried against the normal distribution function (pnorm function in the R statistic package) to give the estimated *P* value, where *s* was the observed value, and *m* and σ were the mean and standard deviation of the null distribution.

In the independent panel of UIG-R, for inter-distribution test (Fig. 5B), there are extensive overlaps between PF and RGF/RAF distributions in female (*P*=0.87 in RGF; *P*=0.73 in RAF), but PF showed moderately lower than RGF (P=0.01) and RAF (*P*=0.02) in males. For single-score test (Fig. 5C), the average SSA of prediction (SSA_{avg}=89.07) was not different (*P*=0.35) from the null RGF expectation in females, but was slightly and significantly lower (SSA_{avg}=83.97, *P*=0.02) than the null RGF expectation in males. When compared to RAF (Fig. 5C), the SSA distributions indicated a similar trend, that females did not show significant difference between PF and RAF (*P*=0.34), whereas in males PF was significantly more similar to the actual

face than the RAF (P=0.02). To account for the potential linkage disequilibrium (LD) within the top SNPs, we further trimmed the top SNPs set either by pairwise LD (r^2<0.8, 240 top SNPs set) or inter-marker physical distance (>400kb, 209 top SNPs set), the prediction became slightly better and the general trends remained unchanged (Fig. S7; Table S12).

Forensic scenarios simulation

In a simulative forensic scenario, we randomly set 8 individuals as the hypothetical candidates, one of them being the true suspect. The SSA scores were obtained between the actual face and PF for each individual as depicted in Figure 6A. If the true suspect happened to have the lowest SSA score among the 8 candidates, the identification was called successful, otherwise failure. For a fixed sample panel of UYG-R, the true accuracy rate can be obtained by averaging over all the possible combinations of 8 candidates. To test whether the true accuracy rate was significantly higher than random level, we reshuffled the pairwise corresponding similarity SSA scores randomly between the actual faces and PFs. For such a reshuffled dataset, we applied the pick-one-from-eight procedure 10,000 times as above to obtain the corresponding accuracy rate for this random dataset. We reshuffled the pairwise score matrix 1,000 times and obtained 1,000 random accuracy rates for the given UYG-R configuration. On account of missing the true corresponds between actual face and PF, the random accuracy rate would be around 1/8. We utilized the proportion (P) of how many random accuracy rates were larger than true accuracy rate as the empirical

p-value. The *P* level of 0.05 was used to determine whether the performance of PF showed significantly better accuracy rate than random situations in forensic scenario.

Compared with the accuracy rate from prediction model on 277 top SNPs (Fig. 5D), the LD controlled SNP panels showed weaker performance both in UIG-R males (increase 1.9% for 240 SNPs and 1.0% for 209 SNPs) and UIG-R females (increase +0.4% for 240 SNPs and +0.4% for 209 SNPs) (Table S13).

SUPPLEMENTARY FIGURES

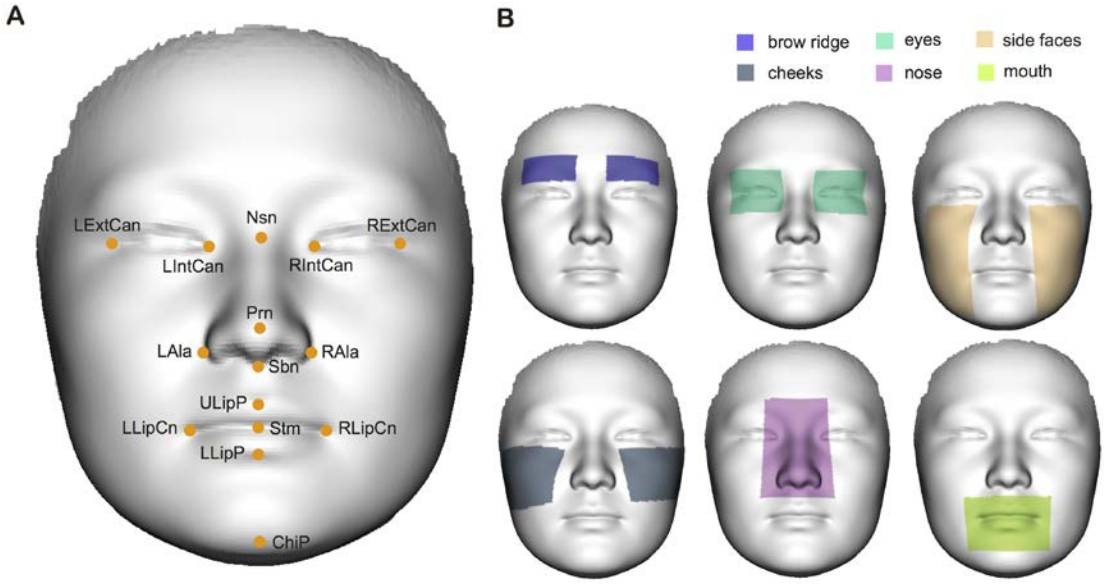

Figure S1. The 15 landmarks and facial features extracted from 3D images. (A) the annotation of the fifteen landmarks. (B) the extraction of six facial features on the 3dDFM data.

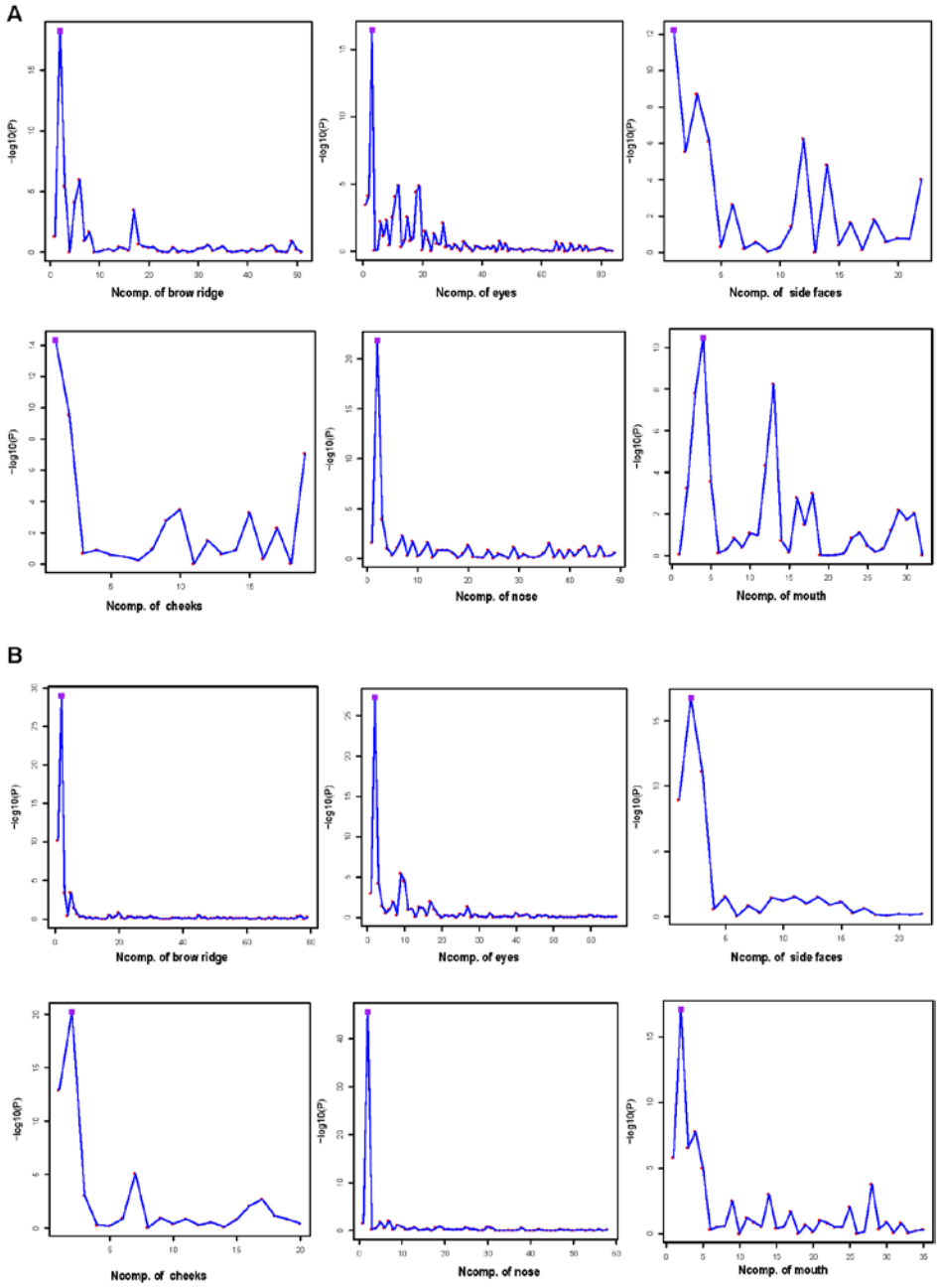

Figure S2. Selection of segregating Principal Components for ancestry-divergent phenotypes in the GWAS analyses. The 3dDFM data from EUR and HAN-TZ are decomposed into perpendicular dimensions using PCA for six features separately. The top PCs who explain 98% total variance are extracted and compared segregations between EUR and HAN-TZ one by one using Students' Test. The purple points in the scree plot show sPC numbers. (A) in females, (B) in males.

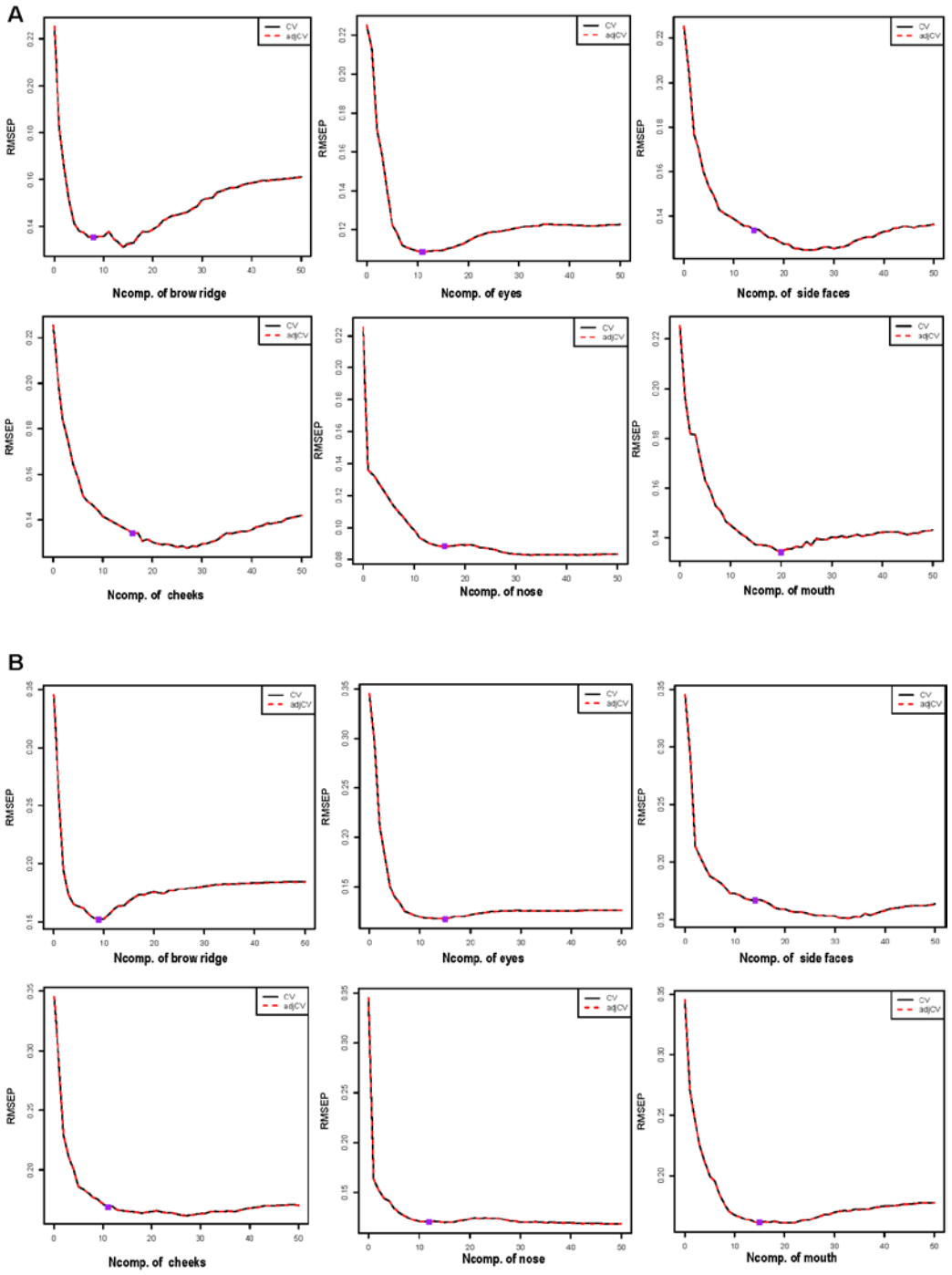

Figure S3. Selection of segregating partial least square space for ancestry-divergent phenotypes in the GWAS analyses. Leave-one-out (LOO) cross-validation are performed to optimize PLS model and validation results are root mean squared error of prediction (RMSEP). The purple points in cross-validated RMSEP curves show sPLS components. (A) in females, (B) in males.

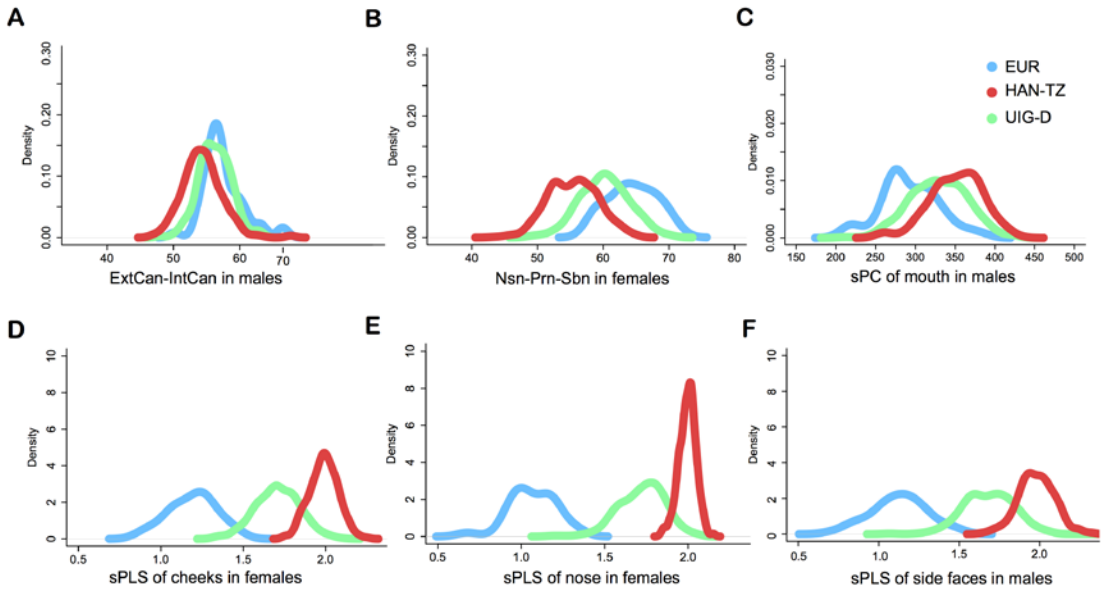

Figure S4. Examples of ancestry-divergent phenotypes showing high divergence between EUR and HAN-TZ. (A) distance of ExtCan-IntCan in males, (B) distance of Nsn-Prn-Sbn in females, (C) sPC of mouth in males, (D) sPLS of cheek in females, (E) nasal sPLS in females, (F) sPLS of side faces in males.

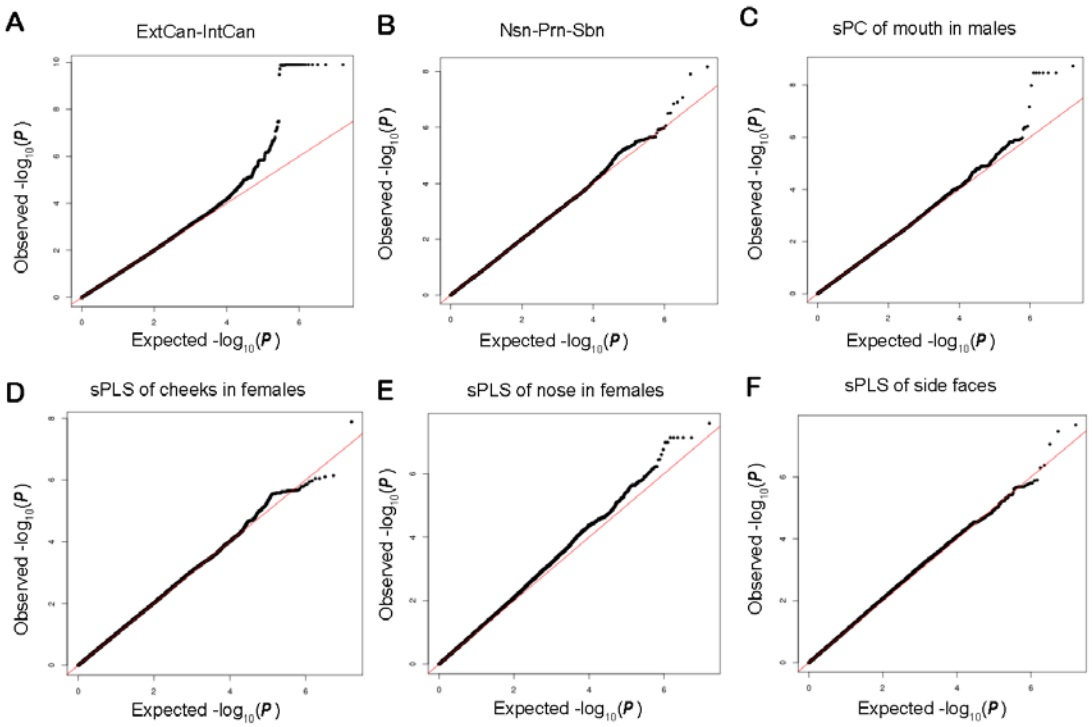

Figure S5. Quantile-Quantile (Q-Q) plots for the GWAS. Q-Q plots of (A) distance of ExtCan-IntCan in males, (B) distance of Nsn-Prn-Sbn in females, (C) sPC of mouth in males, (D) sPLS of cheek in females, (E) nasal sPLS in females, (F) sPLS of side faces in males.

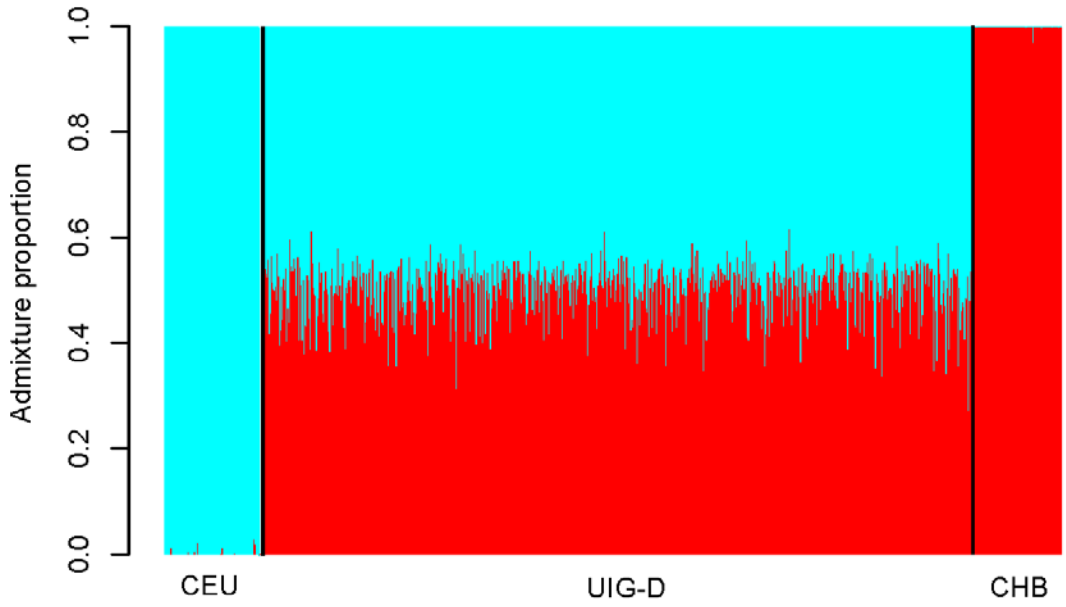

Figure S6. Summary plot of individual admixture proportions. Each individual is represented by a single vertical line broken into two colored segments, with lengths proportional to each of the two inferred clusters; blue indicates European ancestry and red indicates East-Asian ancestry. The admixture happened quite thoroughly in Uyghur at the individual level, making mean ancestry estimates for overall UIG-D are 49.96:50.04 of their European and East-Asian ancestries.

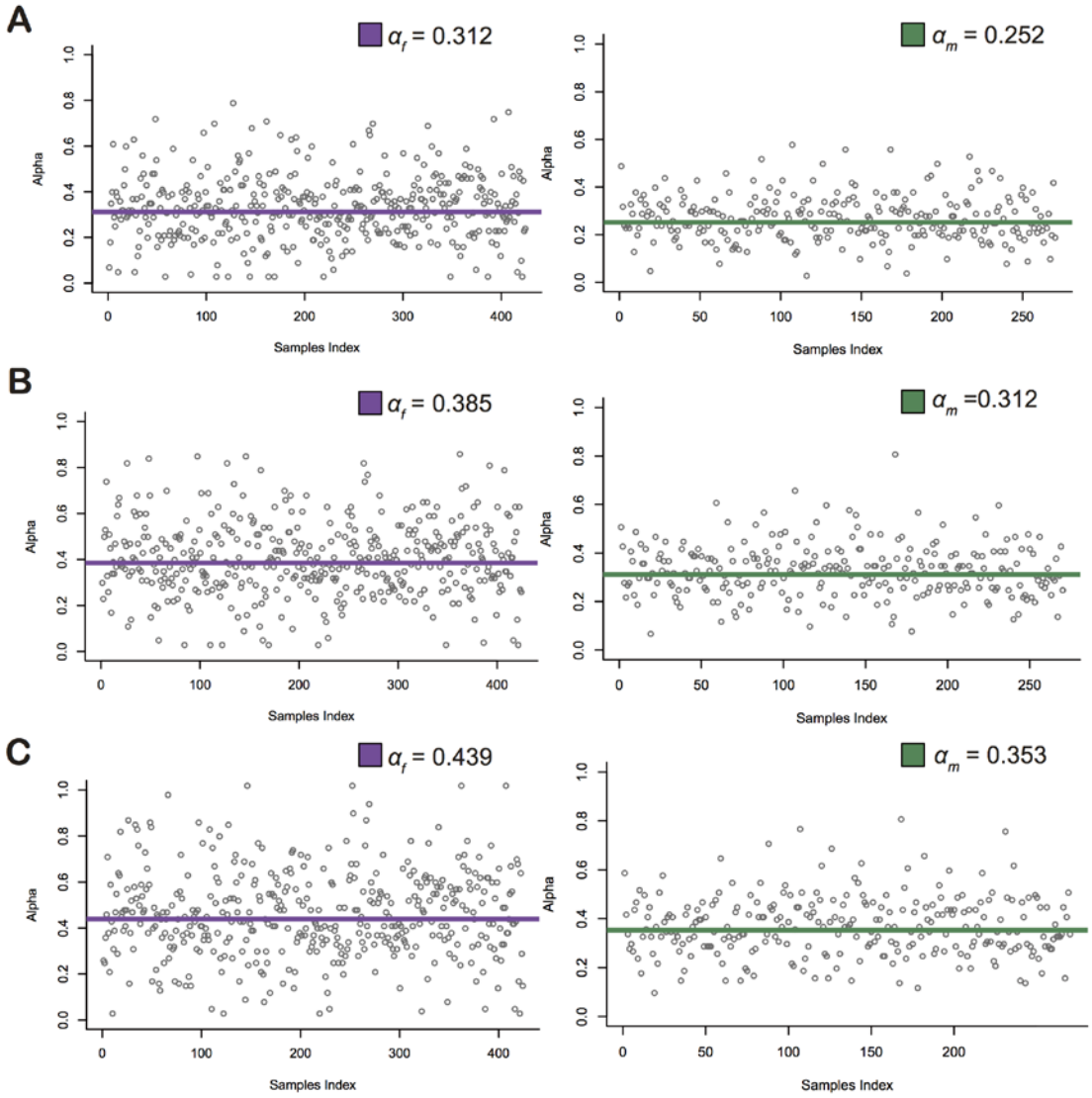

Figure S7. Global effect coefficient α_f and α_m from UIG-D females and UIG-D males. The coefficient α was plotted respectively for (A) the full prediction model of 277 top SNPs, (B) the prediction model of 240 top-SNPs after trimming the SNPs of pairwise LD > 0.8, (C) the prediction model of 209 top-SNPs after trimming SNPs within a physical distance of < 400kb. The global coefficient was obtained by averaging over all individuals.

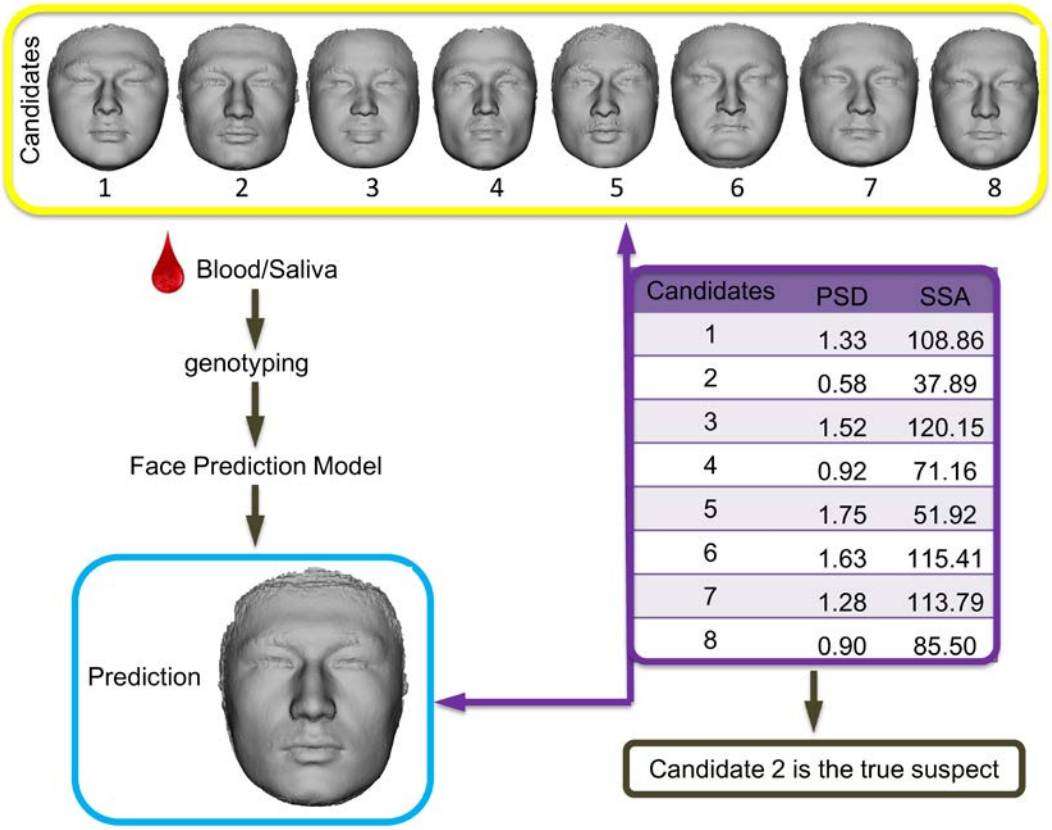

Figure S8. The overall scheme of the forensic test. As can be seen, a number of candidates were given and one true suspect is to be selected. A PF is determined based on the DNA information and compared to the actual 3dDFM data of the candidates and SSA is calculated. The face with minimum SSA is called as the true suspect.

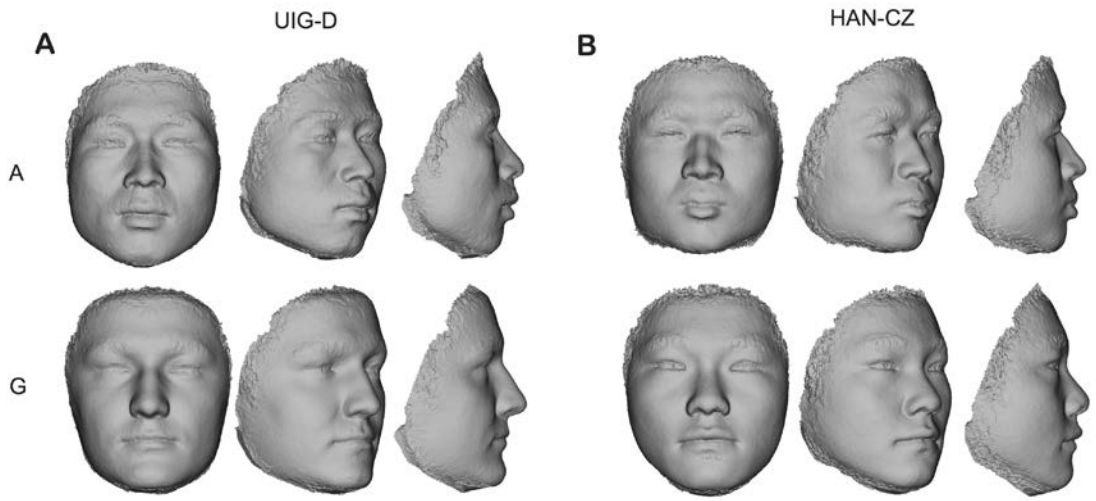

Figure S9. Extrapolated faces of rs61672954 in discovery and independent cohorts. (A) in UIG-D males, (B) in HAN-CZ males. We showed the extreme faces based on residual faces with their allele labeled in front. The top extrapolated faces and alleles are corresponding to Han Chinese liked faces. The bottom extrapolations are European liked faces.

SUPPLEMENTARY TABLES

Table S1. Facial traits defined using 3D facial landmarks

Face section	Features	Quantitative definition	Abbreviation	Measurements
Nose	Nose wing breadth	LAla-Prn-RAla	LaPRa	Distance between LAla-Prn-RAla
		angle LAla-Prn-RAla	anLaPRa	Angle between LAla-Prn-RAla
		LAla-RAla	LaRa	Distance between LAla-RAla
	Nose length	Nsn-Prn-Sbn	NPS	Distance between Nsn-Prn-Sbn
		angle Nsn-Prn-Sbn	anNPS	Angle between Nsn-Prn-Sbn
		Prn-Nsn	PN	Distance between landmarks Prn and Nsn
		Prn-Sbn	PS	Distance between landmarks Prn and Sbn
		angle Prn-Sn-Ls	anPSL	Angle between Prn-Sn-Ls
		Nsn-Sbn	NS	Distance between landmarks Nsn and Sbn
	Nosetip height	height Prn-LAla-RAla	hPLaRa	Distance from Prn to the mid-point of a line joining landmarks LAla and RAla
		height Ala	hA	z-coordinate mean value of LAla and RAla
		height Prn-Ala	hPA	Distance from Prn to mid-point of a line joining LAla and RAla on z-coordinate
		height Prn	hP	z-coordinate mean value of Prn
Nasion	Nasal root breadth	LIntCan-Nsn-RIntCan	LiNRi	Distance between LIntCan-Nsn-RIntCan
		angle LIntCan-Nsn-RIntCan	aLiNRi	Angle between LIntCan-Nsn-RIntCan
	Nasal root height	height Nsn to eyes	hNeyes	Perpendicular distance from Nsn to a line joining LIntCan and RIntCan
		height Nsn-Prn-Sbn	hNPS	Perpendicular distance from Prn to a line joining Nsn and Sbn
Eyes	Eyes length	ExtCan-IntCan	EI	Mean distance between landmarks ExtCan and IntCan of left(L) and right(R)
	Outercanthal width	LExtCan-RExtCan	LeRe	Distance between LExtCan and RExtCan
	Interorbital distance	LIntCan-RIntCan	LiRi	Distance between LIntCan and RIntCan

	Antimongoloid slant	Eyes slope	slope	Angle between a line joining ExtCan-IntCan and y-coordinate
Mouth	Labial fissure length	LLipCn-RLipCn	LlRl	Distance between LLipCn and RLipCn
		angle LLipCn-Stm-RLipCn	anLlSRl	Angle between LLipCn-Stm-RLipCn
		angle LLipCn-ULipP-RLipCn	anLlUlRl	Angle between LLipCn-ULipCn-RLipCn
	Mouth size	Mouth perimeter	perimeter	Distance between LLipCn-LLipP-RLipCn-ULipP
	Philtrum length	Sbn-ULipP	SU	Distance between Sbn and ULipP
	Lip thickness	ULipP-LLipP	UL	Distance between ULipP and LLipP
		angle ULipP-Stm-LLipP	anUSL	Angle between ULipP-Stm-LLipP
		ULipP-Stm-LLipP	USL	Distance between ULipP-Stm-LLipP
		angle Sbn-ULipP-Stm	anSUS	Angle between Sub-ULipP-Stm
		angle Stm-LLipP-ChiP	anSLlC	Angle between Stm-LLipP-ChiP
		Sbn-ULipP-LLipP	SUL	Distance between Sbn-ULipP-LLipP
		angle Sbn-ULipP-LLipP	anSUL	Angle between Sbn-ULipP-LLipP
Chip	Chin length	LLipP-ChiP	LC	Distance between LLipP and ChiP
	Lower face length	Sbn-ULipP-ChiP	SUC	Distance between Sbn-ULipP-ChiP
		angle Sbn-ULipP-ChiP	anSUC	Angle between Sbn-ULipP-ChiP

Table S2. Correlations for inter-landmark phenotypes defined using 3D facial landmarks. Shown are Pearson correlation coefficients (*r*) with color-coded based on magnitude of correlation. The lower left triangle is *r* tested in UIG-D females and upper right triangle in UIG-D males.

		Nose													Nasion				Eyes				Mouth												Chip		
	Quantitative definition	LaPRa	anLaPRa	LaRa	NPS	anNPS	PN	PS	anPSL	NS	hPLaRa	hA	hPA	hP	LiNRi	aLiNRi	hNeyes	hNPS	EI	LeRe	LiRi	slope	LlRl	perimeter	anLlSRl	anLlUlRl	SU	UL	anUSL	USL	anSUS	anSLlC	SUL	anSUL	LC	SUC	anSUC
Nose	LAla-Prn-RAla		-0.68	0.26	0.68	-0.59	0.47	0.74	-0.1	0.45	0.95	-0.25	0.94	0.81	0.56	-0.52	0.58	0.78	0.24	0.25	0.067	-0.094	0.28	0.28	-0.26	0.034	0.038	-0.16	-0.089	-0.023	0.078	0.11	-0.12	0.15	0.12	-0.014	0.28
Nose	angle LAla-Prn-RAla	-0.6		0.52	-0.63	0.48	-0.41	-0.72	-0.33	-0.44	-0.88	0.2	-0.87	-0.77	-0.51	0.71	-0.7	-0.72	-0.16	0.014	0.21	0.068	0.014	-0.013	0.23	-0.014	0.0026	0.27	0.15	0.2	-0.12	-0.089	0.24	-0.24	0.015	0.19	-0.39
Nose	LAla-RAla	0.33	0.55		-0.025	-0.052	0.022	-0.09	-0.55	-0.035	-0.061	-0.028	-0.067	-0.081	-0.04	0.35	-0.27	-0.035	0.066	0.32	0.36	-0.012	0.33	0.31	0.0068	0.011	0.042	0.18	0.1	0.24	-0.08	0.0026	0.19	-0.16	0.15	0.24	-0.22
Nose	Nsn-Prn-Sbn	0.65	-0.6	-0.028		-0.55	0.91	0.77	0.37	0.9	0.71	0.15	0.73	0.8	0.45	-0.46	0.49	0.83	0.15	0.15	0.022	-0.064	0.17	0.17	-0.21	0.06	-0.029	-0.17	-0.081	-0.034	0.082	0.13	-0.17	0.14	0.11	-0.061	0.29
Nose	angle Nsn-Prn-Sbn	-0.57	0.53	0.024	-0.6		-0.42	-0.54	-0.02	-0.15	-0.6	-0.035	-0.63	-0.65	-0.24	0.29	-0.3	-0.8	-0.11	-0.058	0.072	0.089	-0.18	-0.17	0.26	-0.094	-0.021	0.23	0.031	0.12	-0.064	-0.16	0.19	-0.079	-0.056	0.11	-0.21
Nose	Prn-Nsn	0.46	-0.4	0.0091	0.92	-0.43		0.43	0.38	0.91	0.49	0.16	0.51	0.59	0.24	-0.23	0.26	0.57	0.062	0.093	0.042	-0.026	0.17	0.2	-0.18	-0.036	6×10^{-4}	-0.049	-0.047	0.047	0.024	0.085	-0.044	0.05	0.097	0.03	0.15
Nose	Prn-Sbn	0.73	-0.71	-0.076	0.79	-0.66	0.48		0.21	0.54	0.8	0.076	0.78	0.82	0.59	-0.64	0.68	0.93	0.24	0.19	-0.018	-0.1	0.12	0.075	-0.17	0.18	-0.065	-0.28	-0.1	-0.14	0.14	0.15	-0.3	0.24	0.081	-0.18	0.39
Nose	angle Prn-Sn-Ls	-0.053	-0.45	-0.59	0.44	-0.15	0.45	0.28		0.43	0.076	0.37	0.079	0.26	0.11	-0.22	0.2	0.19	-0.003	-0.087	-0.12	-0.031	-0.12	-0.15	-0.035	0.15	0.51	-0.32	-0.27	-0.21	-0.1	0.059	0.068	0.1	-0.004	0.061	0.26
Nose	Nsn-Sbn	0.46	-0.42	-0.008	0.9	-0.21	0.94	0.54	0.47		0.49	0.16	0.49	0.57	0.34	-0.32	0.35	0.52	0.096	0.13	0.065	-0.022	0.12	0.13	-0.11	-0.0069	-0.034	-0.047	-0.068	0.049	0.044	0.062	-0.066	0.097	0.096	0.011	0.19
Nose	height Prn-LAla-RAla	0.92	-0.87	-0.067	0.7	-0.62	0.48	0.8	0.19	0.49		-0.25	0.99	0.87	0.58	-0.65	0.69	0.82	0.22	0.15	-0.05	-0.091	0.17	0.19	-0.27	0.028	0.024	-0.22	-0.12	-0.1	0.1	0.11	-0.18	0.2	0.068	-0.095	0.35
Nose	height Ala	-0.089	0.061	-0.027	0.3	-0.21	0.26	0.24	0.34	0.25	-0.087		-0.24	0.25	0.21	-0.098	0.14	0.083	0.059	0.14	0.16	-0.02	0.12	0.12	-0.093	-0.1	0.07	-0.049	-0.014	-0.068	0.0094	0.13	0.0046	-0.03	-0.075	-0.032	0.11
Nose	height Prn-Ala	0.9	-0.86	-0.079	0.73	-0.66	0.53	0.79	0.2	0.51	0.99	-0.07		0.88	0.58	-0.66	0.69	0.83	0.22	0.14	-0.063	-0.087	0.19	0.21	-0.3	0.047	-0.008	-0.25	-0.12	-0.12	0.12	0.14	-0.23	0.21	0.067	-0.13	0.38
Nose	height Prn	0.77	-0.75	-0.084	0.81	-0.7	0.61	0.84	0.36	0.58	0.85	0.44	0.87		0.69	-0.71	0.76	0.87	0.25	0.21	0.014	-0.097	0.25	0.27	-0.34	-0.0038	0.027	-0.27	-0.13	-0.15	0.12	0.2	-0.22	0.2	0.03	-0.14	0.43
Nasion	LIntCan-Nsn-RIntCan	0.42	-0.36	0.0042	0.37	-0.22	0.21	0.48	0.12	0.28	0.44	0.35	0.44	0.57		-0.66	0.83	0.51	0.12	0.45	0.47	-0.088	0.23	0.21	-0.23	0.054	0.051	-0.22	-0.097	-0.12	0.084	0.17	-0.16	0.16	0.037	-0.093	0.34
Nasion	angle LIntCan-Nsn-RIntCan	-0.44	0.66	0.33	-0.48	0.41	-0.29	-0.61	-0.31	-0.32	-0.6	-0.27	-0.62	-0.7	-0.54		-0.96	-0.56	-0.18	0.082	0.34	0.11	-0.051	-0.034	0.26	-0.09	0.036	0.36	0.14	0.29	-0.13	-0.22	0.35	-0.23	0.062	0.3	-0.47
Nasion	height Nsn to eyes	0.48	-0.64	-0.27	0.5	-0.39	0.29	0.64	0.28	0.34	0.62	0.32	0.64	0.74	0.74	-0.96		0.6	0.17	0.088	-0.094	-0.11	0.12	0.097	-0.27	0.085	-0.008	-0.35	-0.14	-0.26	0.12	0.22	-0.32	0.22	-0.034	-0.26	0.47
Nasion	height Nsn-Prn-Sbn	0.74	-0.7	-0.057	0.84	-0.86	0.58	0.94	0.28	0.53	0.8	0.27	0.82	0.87	0.41	-0.58	0.59		0.21	0.16	-0.034	-0.1	0.17	0.14	-0.24	0.15	-0.032	-0.29	-0.088	-0.14	0.12	0.17	-0.28	0.19	0.087	-0.16	0.37
Eyes	ExtCan-IntCan	0.25	-0.18	0.042	0.17	-0.12	0.11	0.21	0.019	0.13	0.24	0.061	0.23	0.24	0.017	-0.1	0.088	0.19		0.68	-0.084	-0.064	0.15	0.12	-0.051	0.055	0.044	-0.075	-0.043	-0.011	-0.003	0.053	-0.037	-6E-04	0.053	0.007	0.078
Eyes	LExtCan-RExtCan	0.26	-0.017	0.25	0.14	-0.035	0.087	0.17	-0.078	0.13	0.17	0.15	0.15	0.21	0.49	0.08	0.087	0.13	0.71		0.65	-0.053	0.27	0.25	-0.029	0.0074	0.11	0.059	0.01	0.14	-0.038	0.03	0.13	-0.054	0.12	0.18	-0.033
Eyes	LIntCan-RIntCan	0.06	0.22	0.32	-0.026	0.14	-0.015	-0.033	-0.15	0.036	-0.071	0.15	-0.088	-0.003	0.63	0.3	-0.045	-0.079	-0.086	0.63		0.011	0.2	0.2	0.018	-0.061	0.1	0.17	0.062	0.2	-0.058	-0.029	0.23	-0.087	0.097	0.24	-0.15
Eyes	Eyes slope	-0.079	0.093	0.03	-0.046	0.1	0.014	-0.12	-0.006	0.013	-0.095	-0.049	-0.097	-0.11	-0.11	0.18	-0.18	-0.12	-0.066	-0.036	0.039		-0.039	-0.027	0.051	-0.04	-0.044	0.067	0.0063	0.044	-0.021	-0.022	0.03	-0.037	-0.013	0.014	-0.04
Mouth	LLipCn-RLipCn	0.23	0.088	0.34	0.13	-0.081	0.17	0.012	-0.088	0.13	0.095	0.17	0.12	0.19	0.23	-0.03	0.093	0.062	0.11	0.25	0.23	-0.02		0.94	-0.36	0.2	0.01	0.077	0.093	0.18	0.13	-0.006	0.076	0.12	0.2	0.18	0.048
Mouth	Mouth perimeter	0.24	0.045	0.3	0.14	-0.082	0.2	-0.007	-0.1	0.14	0.13	0.17	0.15	0.22	0.23	-0.043	0.1	0.054	0.093	0.23	0.22	-0.008	0.94		-0.46	-0.081	-0.027	0.28	0.12	0.35	0.13	-0.11	0.23	0.09	0.23	0.31	-0.061
Mouth	angle LLipCn-Stm-RLipCn	-0.19	0.16	-0.002	-0.19	0.21	-0.21	-0.09	-0.084	-0.13	-0.2	-0.14	-0.24	-0.28	-0.17	0.23	-0.23	-0.17	-0.038	-0.022	0.02	0.026	-0.32	-0.43		0.11	0.013	0.21	-0.48	0.1	-0.37	-0.36	0.2	-0.14	-0.18	0.043	-0.15
Mouth	angle LLipCn-ULipP-RLipCn	-0.008	0.071	0.075	0.047	-0.05	0.0086	0.091	0.12	0.023	-0.041	-0.051	-0.006	-0.03	-0.008	0.0057	-0.005	0.077	0.029	0.024	-0.012	-0.023	0.28	-0.0049	0.12		0.12	-0.58	-0.094	-0.46	0.2	0.12	-0.44	0.41	0.0014	-0.34	0.46
Mouth	Sbn-ULipP	0.031	-0.0092	0.021	-0.079	0.04	-0.042	-0.11	0.45	-0.063	0.024	0.03	-0.025	-0.008	0.049	0.057	-0.032	-0.088	0.034	0.11	0.11	0.059	0.085	0.053	-0.016	0.1		-0.23	-0.32	-0.058	-0.39	-0.015	0.48	-0.18	0.11	0.46	-0.069
Mouth	ULipP-LLipP	-0.081	0.21	0.17	-0.16	0.24	-0.091	-0.22	-0.29	-0.064	-0.16	-0.085	-0.21	-0.23	-0.083	0.29	-0.26	-0.25	-0.025	0.097	0.19	0.058	0.088	0.26	0.25	-0.51	-0.13		0.3	0.69	0.079	-0.36	0.74	-0.1	-0.049	0.52	-0.49
Mouth	angle ULipP-Stm-LLipP	-0.039	0.13	0.12	-0.049	0.022	-0.019	-0.078	-0.15	-0.04	-0.089	-9E-04	-0.089	-0.081	-0.027	0.095	-0.085	-0.06	-0.015	0.029	0.065	0.011	0.097	0.1	-0.55	0.018	-0.15	0.16		0.086	0.65	0.44	0.042	0.095	-0.032	0.0028	-0.082
Mouth	ULipP-Stm-LLipP	-0.004	0.17	0.2	-0.055	0.17	0.0069	-0.13	-0.2	0.03	-0.087	-0.11	-0.13	-0.17	-0.043	0.28	-0.24	-0.15	0.0034	0.15	0.22	0.053	0.15	0.28	0.15	-0.39	-0.001	0.66	-0.002		0.028	-0.31	0.58	-0.018	0.66	0.85	-0.3
Mouth	angle Sbn-ULipP-Stm	0.0051	-0.006	-0.001	0.039	0.034	0.035	0.031	-0.012	0.055	0.0057	0.043	0.025	0.045	0.052	-0.051	0.056	0.011	-0.008	0.017	0.022	0.0055	0.14	0.11	-0.41	0.29	-0.24	-0.007	0.69	-0.044		0.17	-0.2	0.77	0.1	-0.11	0.54
Mouth	angle Stm-LLipP-ChiP	0.12	-0.098	0.011	0.17	-0.23	0.12	0.18	0.14	0.083	0.12	0.22	0.15	0.25	0.17	-0.23	0.24	0.22	0.043	0.045	-0.01	-0.056	0.039	-0.061	-0.4	0.16	0.033	-0.36	0.53	-0.33	0.29		-0.34	-0.14	-0.11	-0.28	0.37
Mouth	Sbn-ULipP-LLipP	-0.047	0.17	0.16	-0.19	0.23	-0.1	-0.26	0.053	-0.095	-0.11	-0.052	-0.19	-0.2	-0.037	0.28	-0.24	-0.27	0.0017	0.15	0.23	0.088	0.13	0.25	0.19	-0.36	0.56	0.75	0.039	0.55	-0.17	-0.28		-0.21	0.032	0.78	-0.49
Mouth	angle Sbn-ULipP-LLipP	0.042	-0.12	-0.099	0.097	0.033	0.069	0.11	0.14	0.12	0.083	0.023	0.11	0.11	0.093	-0.13	0.14	0.063	0.01	0.0092	-0.016	-0.006	0.15	0.11	-0.11	0.46	-0.15	-0.13	0.06	-0.053	0.73	-0.11			0.17	-0.077	0.73
Chip	LLipP-ChiP	0.045	0.057	0.11	0.05	0.03	0.072	-0.002	-0.024	0.079	-0.002	-0.12	-0.007	-0.066	-0.015	0.15	-0.13	-0.004	0.022	0.11	0.12	0.026	0.13	0.15	-0.12	-0.0019	0.098	-0.035	-0.02	0.71	0.067	-0.15	0.036	0.11		0.64	0.061
Chip	Sbn-ULipP-ChiP	0.0014	0.16	0.19	-0.1	0.18	-0.025	-0.18	0.04	-0.016	-0.079	-0.097	-0.14	-0.17	-0.025	0.29	-0.25	-0.19	0.016	0.18	0.25	0.08	0.18	0.28	0.073	-0.27	0.5	0.52	0.011	0.86	-0.092	-0.25	0.77	-0.1	0.67		-0.32
Chip	angle Sbn-ULipP-ChiP	0.15	-0.25	-0.15	0.24	-0.16	0.16	0.28	0.31	0.18	0.22	0.22	0.27	0.35	0.24	-0.36	0.36	0.26	0.048	0.022	-0.063	-0.048	0.086	-0.027	-0.092	0.46	-0.044	-0.47	-0.059	-0.31	0.5	0.43	-0.42	0.7	-0.048	-0.31	

Table S3. Divergence of landmark based ancestry-divergent phenotypes. The *P* values are measured by Students' test between EUR and HAN-TZ.

Ancestry-divergent phenotypes based on landmark	***P***	
	Female	**Male**
LAla-Prn-RAla	3×10^{-11}	2×10^{-21}
Nsn-Prn-Sbn	2×10^{-15}	7×10^{-23}
height Prn-Ala	1×10^{-18}	2×10^{-31}
LIntCan-Nsn-RIntCan	2×10^{-10}	5×10^{-27}
height Nsn to eyes	1×10^{-21}	6×10^{-47}
ExtCan-IntCan	1×10^{-7}	3×10^{-10}
LIntCan-RIntCan	9×10^{-7}	4×10^{-9}
LLipCn-RLipCn	5×10^{-3}	2×10^{-8}
ULipP-Stm-LLipP	3×10^{-3}	7×10^{-9}
Sbn-ULipP-ChiP	4×10^{-3}	7×10^{-7}

Table S4. Divergence of PCA based and PLS based ancestry-divergent phenotypes.

		PCA based		PLS based	
Partial Features	**Gender**[a]	**ncomp.**[b]	***P***[c]	**ncomp.**[b]	***P***[c]
brow ridge	Female	2	$\mathbf{6 \times 10^{-19}}$	1:6	$\mathbf{6 \times 10^{-23}}$
	Male	2	$\mathbf{1 \times 10^{-29}}$	1:8	$\mathbf{3 \times 10^{-49}}$
eyes	Female	5	$\mathbf{2 \times 10^{-16}}$	1:8	$\mathbf{1 \times 10^{-25}}$
	Male	2	$\mathbf{9 \times 10^{-17}}$	1:11	$\mathbf{1 \times 10^{-60}}$
side faces	Female	1	$\mathbf{6 \times 10^{-13}}$	1:25	$\mathbf{6 \times 10^{-27}}$
	Male	2	$\mathbf{2 \times 10^{-17}}$	1:9	$\mathbf{2 \times 10^{-47}}$
cheeks	Female	1	$\mathbf{5 \times 10^{-15}}$	1:22	$\mathbf{1 \times 10^{-27}}$
	Male	2	$\mathbf{6 \times 10^{-21}}$	1:8	$\mathbf{4 \times 10^{-43}}$
nose	Female	2	$\mathbf{1 \times 10^{-22}}$	1:17	$\mathbf{3 \times 10^{-29}}$
	Male	2	$\mathbf{2 \times 10^{-46}}$	1:12	$\mathbf{6 \times 10^{-57}}$
mouth	Female	4	$\mathbf{3 \times 10^{-11}}$	1:20	$\mathbf{2 \times 10^{-24}}$
	Male	2	$\mathbf{9 \times 10^{-18}}$	1:15	$\mathbf{2 \times 10^{-48}}$

[a]sex specified test

[b]Number of components that used to distinguish Han Chinese and European features

[c]Student's test

Table S5. Detailed information about quality control. Here summarized the number of SNPs and samples omitted and retained for each quality filter.

	UIG-D		**UIG-R**	
Plantform	Illumina Omni Zhonghua-8		Affymetrix Genome-Wide Human SNP Array 6.0	
	Number of SNPs	**Number of samples**	**Number of SNPs**	**Number of samples**
Before QC	894,956	726	934,968	192
Mitochondria SNPs	147	726	411	192
Meaningless SNPs	1,880	726	1,846	192
MAF <0.01	41,988	726	127,717	192
Missing rate per SNP >90%	1,544	726	66,295	192
HWE p-value <10-6	487	726	3,652	192
Positional duplicates	2,541	726	2,947	192
Uyghur Ancestry outliers	847,046	3	758,453	0
Missing rate per person >10%	847,046	1	758,453	0
Pairwise IBD estimation	847,046	0	758,453	12
Inbreeding coefficients >0.2 or <-0.2	847,046	0	758,453	0
Sex check	847,046	0	758,453	0
defecive 3D images	847,046	28	758,453	9
After QC	847,046	694	758,453	171

Table S6. Inflation factors of six significant ancestry-divergent phenotypic GWAS. Genome-wide association tests using multivariate linear regression were applied, as implemented in PLINK, taking an additive linear regression model adjusting for the first four genetic principal components (gPCs) of UIG-D which were decomposed using EIGENSTRAT.

SNP	P	Inflation factor	gPCs corrected P values	gPCs corrected Inflation factors
rs1868752	$\mathbf{1 \times 10^{-10}}$	1	$\mathbf{1 \times 10^{-10}}$	1
rs118078182	$\mathbf{8 \times 10^{-9}}$	1.01	$\mathbf{1 \times 10^{-8}}$	1.01
rs60159418	$\mathbf{9 \times 10^{-11}}$	1.01	$\mathbf{1 \times 10^{-9}}$	1
rs17868256	$\mathbf{7 \times 10^{-9}}$	1.04	$\mathbf{1 \times 10^{-7}}$	1.01
rs3920540	$\mathbf{3 \times 10^{-8}}$	1.06	$\mathbf{2 \times 10^{-7}}$	1.02
rs61672954	$\mathbf{2 \times 10^{-8}}$	1.04	$\mathbf{1 \times 10^{-7}}$	1

Table S7. Correlations for all ancestry-divergent phenotypes. Shown are Pearson correlation coefficients (*r*) with color-coded based on magnitude of correlation. The lower left triangle is *r* tested in UIG-D females and upper right triangle in UIG-D males.

		Landmark based ancestry-divergent phenotypes										PCA based ancestry-divergent phenotypes						PLS based ancestry-divergent phenotypes					
		LaPRa	**NPS**	**hPA**	**LiNRi**	**hNeyes**	**EI**	**LiRi**	**LlRl**	**USL**	**SUC**	**B.sPC**	**E.sPC**	**S.sPC**	**C.sPC**	**N.sPC**	**M.sPC**	**B.sPLS**	**E.sPLS**	**S.sPLS**	**C.sPLS**	**N.sPLS**	**M.sPLS**
Landmark based	**LAla-Prn-RAla**		0.6	0.94	0.39	0.33	0.0014	0.23	0.29	0.23	0.19	-0.039	0.0046	0.16	-0.11	-0.4	0.082	-0.22	-0.13	0.011	0.093	-0.31	-0.053
	Nsn-Prn-Sbn	0.56		0.63	0.26	0.24	-0.043	0.14	0.19	0.12	0.051	-0.043	0.14	0.37	0.002	-0.003	0.062	-0.01	-0.014	-0.038	-0.015	-0.29	-0.17
	height Prn-Ala	0.91	0.62		0.39	0.41	-0.035	0.13	0.2	0.18	0.1	-0.12	-0.046	0.11	-0.18	-0.51	0.059	-0.28	-0.22	-0.055	0.023	-0.43	-0.2
	LIntCan-Nsn-RIntCan	0.34	0.21	0.34		0.79	-0.016	0.66	0.14	0.08	0.077	-0.013	0.084	0.17	0.11	-0.5	0.028	-0.41	-0.28	-0.002	0.041	-0.35	-0.18
	height Nsn to eyes	0.32	0.23	0.42	0.73		-0.026	0.069	0.01	-0.024	-0.083	-0.35	-0.18	0.079	0.026	-0.69	-0.062	-0.59	-0.51	-0.13	-0.16	-0.54	-0.21
	ExtCan-IntCan	0.2	0.12	0.14	0.0093	-0.046		-0.002	0.04	0.01	0.029	0.2	-0.59	0.1	0.21	0.061	-0.007	-0.1	-0.18	-0.18	-0.027	-0.11	0.073
	LIntCan-RIntCan	0.2	0.1	0.11	0.78	0.14	0.059		0.22	0.16	0.22	0.41	0.36	0.18	0.15	0.024	0.12	0.057	0.17	0.16	0.25	0.093	-0.036
	LLipCn-RLipCn	0.25	0.17	0.17	0.27	0.13	0.15	0.28		0.22	0.22	0.096	0.13	0.26	0.18	0.016	-0.13	0.068	0.086	-0.14	-0.033	-0.007	-0.049
	ULipP-Stm-LLipP	0.054	0.047	-0.008	0.059	-0.1	-0.008	0.18	0.19		0.83	0.093	0.078	0.2	-0.092	0.14	0.76	0.037	0.15	0.17	0.12	0.095	0.24
	Sbn-ULipP-ChiP	0.056	-0.054	-0.033	0.057	-0.13	-0.02	0.2	0.22	0.82		0.16	0.08	0.33	-0.033	0.24	0.79	0.11	0.19	0.18	0.15	0.13	0.22
PCA based	**BrowRidge.sPC**	-0.039	0.061	-0.097	-0.13	-0.41	-0.05	0.19	0.026	0.27	0.23		0.19	-0.17	-0.087	0.4	0.086	0.57	0.49	0.14	0.29	0.35	0.14
	eyes.sPC	0.097	0.025	0.18	0.024	0.21	0.42	-0.15	-0.058	-0.024	-0.074	-0.37		-0.086	-0.16	0.27	0.095	0.51	0.43	0.45	0.4	0.34	0.073
	SideFace.sPC	0.28	0.19	0.081	0.33	-0.03	0.44	0.51	0.56	0.3	0.32	0.33	-0.17		0.66	0.17	0.24	-0.15	-0.069	-0.25	-0.43	-0.12	-0.041
	cheeks.sPC	0.24	0.16	0.052	0.33	-0.02	0.41	0.5	0.45	0.28	0.28	0.36	-0.18	0.97		0.073	-0.095	-0.2	-0.11	-0.38	-0.48	-0.1	-0.1
	nose.sPC	-0.43	-0.27	-0.62	-0.47	-0.79	0.054	0.034	-0.013	0.21	0.26	0.46	-0.35	0.28	0.26		0.19	0.63	0.58	0.26	0.23	0.57	0.33
	mouth.sPC	0.052	-0.076	-0.067	0.067	0.0035	0.098	0.093	0.024	-0.2	-0.18	0.065	-0.044	0.26	0.3	0.082		0.078	0.14	0.28	0.2	0.12	0.25
PLS based	**BrowRidge.sPLS**	-0.18	-0.073	-0.25	-0.38	-0.63	-0.098	0.02	-0.059	0.17	0.15	0.66	-0.38	0.16	0.18	0.62	0.076		0.61	0.27	0.27	0.57	0.26
	eyes.sPLS	-0.18	-0.12	-0.29	-0.29	-0.52	-0.23	0.054	0.054	0.13	0.18	0.46	-0.47	0.16	0.15	0.57	0.044	0.63		0.37	0.31	0.66	0.22
	SideFace.sPLS	0.028	-0.16	-0.12	-0.12	-0.26	-0.15	0.058	-0.037	-0.014	0.04	0.27	-0.45	0.24	0.26	0.43	0.3	0.33	0.41		0.76	0.34	0.3
	cheeks.sPLS	0.052	-0.19	-0.096	-0.14	-0.32	-0.12	0.086	-0.054	0.026	0.073	0.3	-0.42	0.22	0.25	0.43	0.25	0.4	0.43	0.84		0.29	0.21
	nose.sPLS	-0.26	-0.19	-0.42	-0.42	-0.67	-0.063	0.0068	-0.1	0.11	0.15	0.4	-0.32	0.11	0.12	0.75	0.16	0.67	0.7	0.45	0.5		0.35
	mouth.sPLS	-0.049	-0.21	-0.23	-0.12	-0.3	0.082	0.098	-0.068	0.15	0.15	0.23	-0.2	0.24	0.25	0.4	0.3	0.25	0.23	0.37	0.39	0.4	

Table S8. Benjamini-Hochberg procedure as multiple testing correction. After multiple testing correction, there were 39 significant SNPs with adjusted *P* <0.05 corresponding to 4 index SNPs (in bold) reported associated regions in Table 2.

Chr.	SNP	Gender	Trait	P value	Adjusted *P*
2	**rs17868256**	**female**	**Cheek sPLS**	**7.00E-09**	**4.90E-02**
4	**rs60159418**	**male**	**Mouth sPC**	**9.00E-11**	**3.30E-03**
4	rs60159418	male	Mouth sPC	1.00E-09	1.80E-02
4	rs56271018	male	Mouth sPC	2.00E-09	2.60E-02
4	rs16884372	male	Mouth sPC	3.00E-09	3.20E-02
4	rs11937565	male	Mouth sPC	3.00E-09	3.20E-02
4	rs11945187	male	Mouth sPC	3.00E-09	3.20E-02
4	rs16867997	male	Mouth sPC	3.00E-09	3.20E-02
4	rs16884388	male	Mouth sPC	3.00E-09	3.20E-02
4	rs12507539	male	Mouth sPC	3.00E-09	3.20E-02
5	**rs118078182**	**mixed**	**Nsn-Prn-Sbn**	**8.00E-09**	**5.00E-02**
11	rs34639545	mixed	ExtCan-IntCan	6.00E-11	2.20E-03
11	**rs1868752**	**mixed**	**ExtCan-IntCan**	**1.00E-10**	**3.30E-03**
11	rs150546335	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs117799641	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs1975637	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs964662	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs56699489	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs58390872	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs78156770	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs34143639	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs11822267	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs74483984	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs56656319	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs79259187	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs11826493	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs61141574	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs11823815	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs115392245	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs147681613	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs57524079	mixed	ExtCan-IntCan	1.00E-10	4.80E-03
11	rs78206044	mixed	ExtCan-IntCan	1.00E-10	4.90E-03
11	rs56048952	mixed	ExtCan-IntCan	1.00E-10	5.00E-03

11	rs77420038	mixed	ExtCan-IntCan	1.00E-10	5.00E-03
11	rs184981564	mixed	ExtCan-IntCan	1.00E-10	5.00E-03
11	rs144478480	mixed	ExtCan-IntCan	1.00E-10	5.00E-03
11	rs144514908	mixed	ExtCan-IntCan	1.00E-10	5.00E-03
11	rs144189728	mixed	ExtCan-IntCan	1.00E-10	5.00E-03
11	rs148657983	mixed	ExtCan-IntCan	1.00E-10	5.00E-03
11	rs1868754	mixed	ExtCan-IntCan	2.00E-10	7.20E-03

Table S9. Sample size of each genotype per SNP in studied cohorts. Shown are six index SNPs in the corresponding discovered gender groups. We used one criterion to eliminate the interference from small sample size that we only considered sample size with individuals carrying minor allele (homozygote*2 + heterozygote*1) larger than 10 which number was enough to guarantee a stable average face. The black genotypes set presented below illustrate studied groups used to discover or replicate the index SNPs.

		UIG-D			**UIG-R**			**HAN-CZ**		
rs1868752		**GG**	**TG**	**TT**	**GG**	**TG**	**TT**	**GG**	**TG**	**TT**
	Female	1	14	423	0	4	112	3	110	3091
	Male	0	7	279	0	1	66	0	37	450
rs118078182		**AA**	**GA**	**GG**	**AA**	**GA**	**GG**	**AA**	**GA**	**GG**
	Female	5	72	361	1	8	83	42	364	797
	Male	3	43	240	1	7	51	19	146	321
rs60159418		**AA**	**GA**	**GG**	**AA**	**GA**	**GG**	**AA**	**GA**	**GG**
	Male	39	127	120	7	24	31	251	201	30
rs17868256		**AA**	**GA**	**GG**	**AA**	**GA**	**GG**	**AA**	**GA**	**GG**
	Female	177	201	60	38	38	5	257	613	333
rs3920540		**GG**	**TG**	**TT**	**GG**	**TG**	**TT**	**GG**	**TG**	**TT**
	Female	9	112	317	0	21	72	24	252	922
rs61672954		**AA**	**GA**	**GG**	**AA**	**GA**	**GG**	**AA**	**GA**	**GG**
	Female	0	32	406	0	4	100	13	214	976
	Male	1	18	267	0	3	57	10	79	398

Table S10. Previous candidate SNPs associated with Uyghur facial shape

						P		
SNP	**Chr.**	**BP[a]**	**Gene**	**Discovered Trait[b]**	**Studied Trait**	**Female**	**Male**	**Mixed**
rs4648379	1	3261516	PRDM16	LAla-Prn-RAla	LAla-Prn-RAla	7×10^{-1}	2×10^{-1}	2×10^{-1}
					angle of LAla-Prn-Rala	2×10^{-1}	$\mathbf{3 \times 10^{-2}}$	$\mathbf{2 \times 10^{-2}}$
rs642961	1	209989270	IRF6	ULipP-Stm-LLipP	ULipP-Stm-LLipP	$\mathbf{5 \times 10^{-2}}$	8×10^{-1}	2×10^{-1}
rs3827760	2	109513601	EDAR	Chin protrusion	angle of Sbn-ULipP-ChiP	$\mathbf{6 \times 10^{-2}}$	1×10^{-1}	$\mathbf{1 \times 10^{-2}}$
rs7559271	2	223068286	PAX3	Nasion	height of Nasion to eyes	$\mathbf{6 \times 10^{-3}}$	1×10^{-1}	$\mathbf{2 \times 10^{-3}}$
rs17447439	3	189549423	TP63	EyeR-EyeL	LExtCan-RExtCan	8×10^{-1}	5×10^{-1}	1
rs2045323	4	154831899	DCHS2	Columella inclination/Nose protrusion/Nose tip angle	height of Prn to Ala	$\mathbf{6 \times 10^{-2}}$	$\mathbf{4 \times 10^{-2}}$	8×10^{-1}
rs6184	5	42719344	GHR	mandibular height (ear-Gn)	Ear-ChiP	$\mathbf{9 \times 10^{-2}}$	3×10^{-1}	6×10^{-1}
rs1852985	6	45329656	SUPT3H/RUNX2	Nose bridge breadth	LAla-RAla	7×10^{-1}	2×10^{-1}	2×10^{-1}
rs7773292	6	132099761	ENPP1	upper and lower face height	Sbn-UlipP-ChiP	9×10^{-1}	1	1
rs17640804	7	42131390	GLI3	Nose wing breadth	LAla-RAla	1×10^{-1}	1×10^{-1}	$\mathbf{3 \times 10^{-2}}$
rs805722	10	105810400	COL17A1	EyeLR-Nsn	LIntCan-Nsn-RExtCan	5×10^{-1}	3×10^{-1}	9×10^{-1}
					angle of LIntCan-Nsn-RExtCan	$\mathbf{4 \times 10^{-3}}$	8×10^{-1}	$\mathbf{2 \times 10^{-2}}$
rs927833	20	22041577	PAX1	Nose wing breadth	LAla-RAla	5×10^{-1}	$\mathbf{8 \times 10^{-3}}$	$\mathbf{3 \times 10^{-2}}$

[a]NCBI build 37

[b]original measurements in previous studies

Table S11. Top SNPs information used for prediction. To capture the majority of the facial shape variance for prediction, both ancestry-divergent phenotypes and within-Uyghur phenotypes were defined and analyzed by GWA, and any autosomal SNPs that passed the threshold of $P<1 \times 10^{-6}$ in any test were combined into the panel of the top 277 SNPs.

available in separate excel file, see S11_Table.xlsx

Table S12. Single-statistic test of LD-trimming prediction models. The top SNPs set were trimmed either by pairwise LD (r^2<0.8, 240 top SNPs set) or inter-marker physical distance (>400kb, 209 top SNPs set). The values are the probability of predicted statistic distributed on the relative random normal distribution calculated like normal one-side *P* value.

	240-SNP-set model		**209-SNP-set model**	
	PF vs. RGF	**PF vs. RAF**	**PF vs. RGF**	**PF vs. RAF**
Female	0.6	0.6	0.7	0.6
Male	0.02	0.015	0.015	0.02

.

Table S13. Evaluation of the face prediction in hypothetic forensic scenarios (8 candidates) using 240-SNP and 209-SNP face prediction models. SSA as accuracy statistics are evaluated in UIG-R. The true accuracy rate based on SSA are determined by examining how many cases of successful identification were achieved in all combined iterations. The random accuracy rate is calculated after reshuffling the pairwise corresponds between actual face and PF. This process was repeated 1,000 times to obtain random accuracy rates under null hypothesis, the mean of which summarize the null distribution. The *P* values are the proportion of how many random accuracy rates were larger than true accuracy rate calculated as the empirical p-value.

	240-SNP-set model			**209-SNP-set model**		
	predicted accuracy rate	**mean of random accuracy rates**	***P***	**predicted accuracy rate**	**mean of random accuracy rate**	***P***
Female	12.8%	12.4%	0.4	12.9%	12.5%	0.4
Male	15.2%	12.3%	0.2	13.9%	12.9%	0.4

Table S14. Quality control of six index SNPs genotyping by SNaPshot

SNP	Chr.	BP	call rate (%)	HWE (P value)	MAF (SNaPshot)	
rs1868752	11	122391442	98.51	1	G	0.045
rs118078182	5	177922198	98.34	0.7329	A	0.19
rs60159418	4	31120752	97.77	0.0666	G	0.28
rs17868256	2	52032773	98.57	0.2814	A	0.47
rs3920540	20	7067738	97.71	0.1015	G	0.13
rs61672954	3	82196528	98.46	0.1355	A	0.10

SUPPLEMENTARY MOVIES

(available in separate files)

Movie S1. 3D visualization of extrapolated face affected by rs1868752.

Movie S2. 3D visualization of extrapolated face affected by rs118078182.

Movie S3. 3D visualization of extrapolated face affected by rs60159418.

Movie S4. 3D visualization of extrapolated face affected by rs17868256.

Movie S5. 3D visualization of extrapolated face affected by rs3920540.

Movie S6. 3D visualization of extrapolated face affected by rs61672954.

Movie S7. 3D visualization of actual face and predicted face.