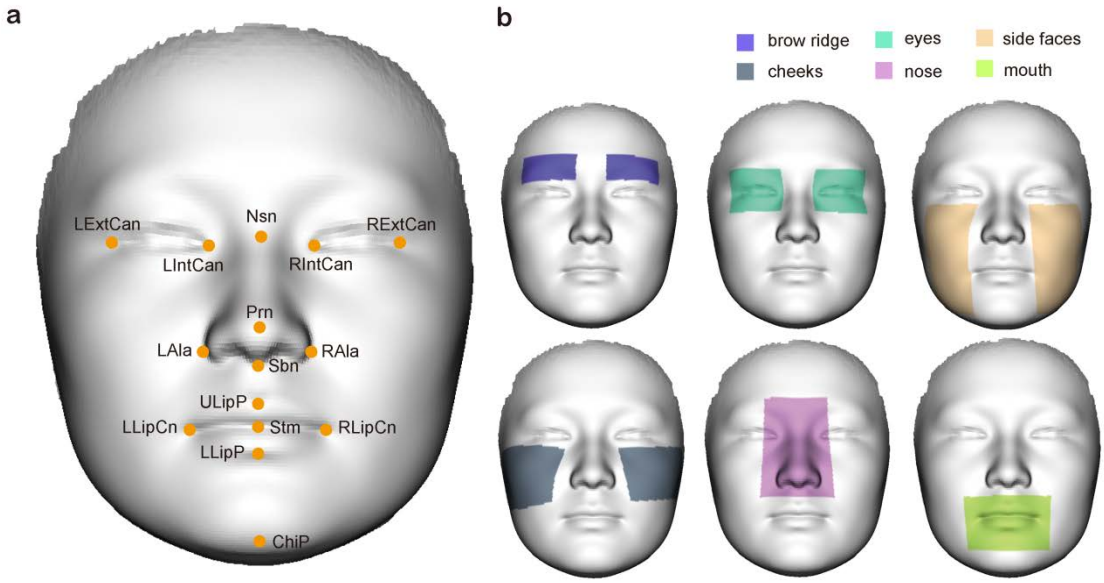
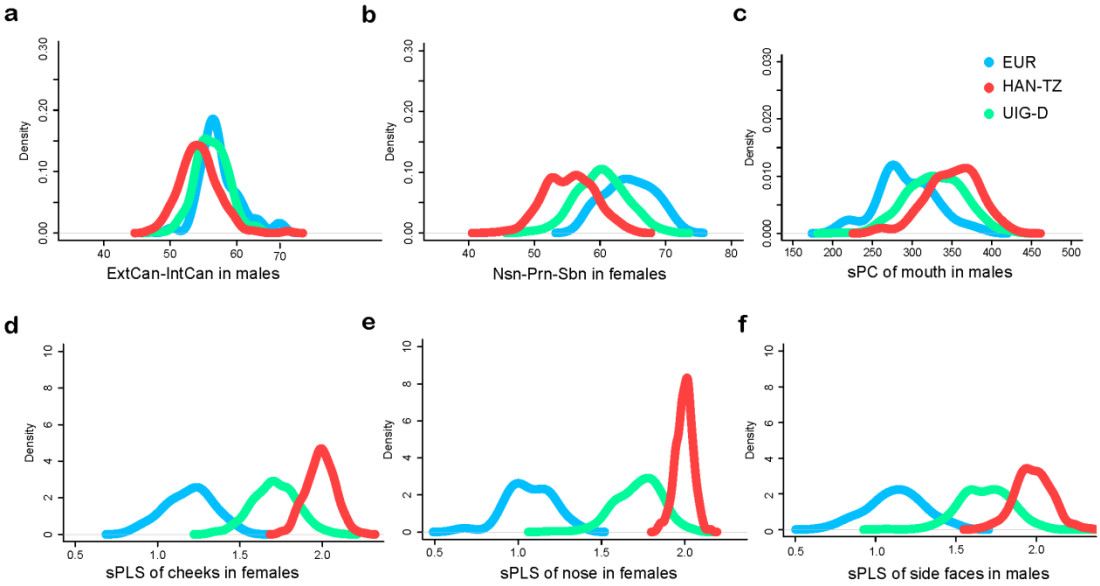
Genome-wide Variants of Eurasian Facial Shape Differentiation and DNA based Face Prediction

Lu Qiao[1,2], Yajun Yang[3,4], Pengcheng Fu[5], Sile Hu[1,2], Hang Zhou[1,2], Jingze Tan[3,4], Yan Lu[1], Haiyi Lou[1], Dongsheng Lu[1,2], Sijie Wu[1,2], Jing Guo[1], Shouneng Peng[1,2], Li Jin[1,2,3,4,7], Yaqun Guan[6], Sijia Wang[1,2,7,*], Shuhua Xu[1,2,7,8,*] & Kun Tang[1,2,*]

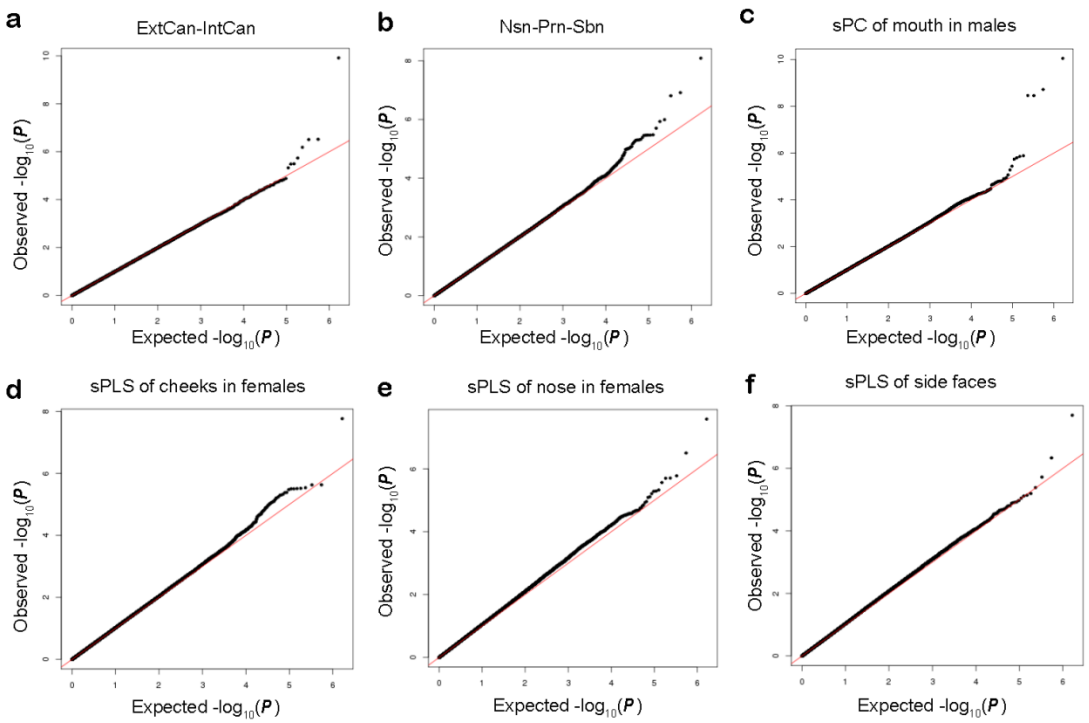
[1] Chinese Academy of Sciences (CAS) Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Science, Shanghai 200031, China;
[2]University of Chinese Academy of Sciences, Beijing 100049, China;
[3]State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China;
[3]Fudan-Taizhou Institute of Health Sciences, 1 Yaocheng Road, Taizhou, Jiangsu 225300, China;
[5]Department of neurology, the First People's Hospital of Chenzhou, Hunan, 423000, China;
[6]Department of Biochemistry and Molecular Biology, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, P. R China;
[7]Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China;
[8]School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China.
*Correspondence could be addressed to K.T. (tangkun@picb.ac.cn), S.X. (xushua@picb.ac.cn), S.W. (wangsijia@picb.ac.cn),



Supplementary Figure 1 | 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.



Supplementary Figure 2 | Examples of candidate 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.



Supplementary Figure 3 | Q-Q plot of GWAS showing six genome-wide significant SNPs. (a) rs1868752, (b) rs118078182, (c) rs60159418, (d) rs17868256, (e) rs3920540, (f) rs61672954.

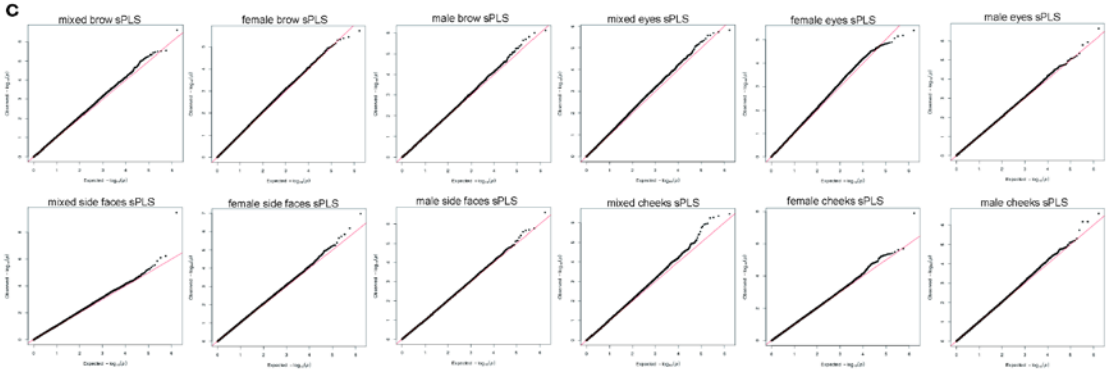

a
mixed LAla-Prn-RAla
female LAla-Prn-RAla
male LAla-Prn-RAla
mixed angle of LAla-Prn-RAla
female angle of LAla-Prn-RAla
male angle of LAla-Prn-RAla
mixed Nsn-Prn-Sbn
female Nsn-Prn-Sbn
male Nsn-Prn-Sbn
mixed angle of Nsn-Prn-Sbn
female angle of Nsn-Prn-Sbn
male angle of Nsn-Prn-Sbn
mixed Nsn-Prn
female Nsn-Prn
male Nsn-Prn
mixed Prn-Sbn
female Prn-Sbn
male Prn-Sbn
mixed LAla-RAla
female LAla-RAla
male LAla-RAla
mixed LIntCan-Nsn-RIntCan
female LIntCan-Nsn-RIntCan
male LIntCan-Nsn-RIntCan
mixed angle of LIntCan-Nsn-RIntCan
female angle of LIntCan-Nsn-RIntCan
male angle of LIntCan-Nsn-RIntCan
mixed height of Nsn to eyes
female height of Nsn to eyes
male height of Nsn to eyes
mixed height of Prn to Nsn-Sbn
female height of Prn to Nsn-Sbn
male height of Prn to Nsn-Sbn
mixed height of Prn to LAla-RAla
female height of Prn to LAla-RAla
male height of Prn to LAla-RAla
mixed angle of Prn-Sbn-ULipP
female angle of Prn-Sbn-ULipP
male angle of Prn-Sbn-ULipP
mixed Sbn-ULipP
female Sbn-ULipP
male Sbn-ULipP
mixed Z axis of Prn
female Z axis of Prn
male Z axis of Prn
mixed Z axis of Ala
female Z axis of Ala
male Z axis of Ala


mixed Z axes of Prn to Ala
female Z axes of Prn to Ala
male Z axes of Prn to Ala
mixed Nsn-Sbn
female Nsn-Sbn
male Nsn-Sbn
mixed ExtCan-IntCan
female ExtCan-IntCan
male ExtCan-IntCan
mixed LExtCan-RExtCan
female LExtCan-RExtCan
male LExtCan-RExtCan
mixed LIntCan-RIntCan
female LIntCan-RIntCan
male LIntCan-RIntCan
mixed eyes slope
female eyes slope
male eyes slope
mixed LLipCn-RLipCn
female LLipCn-RLipCn
male LLipCn-RLipCn
mixed ULipP-LLipP
female ULipP-LLipP
male ULipP-LLipP
mixed ULipP-LLipCn-LLipP-RLipCn
female ULipP-LLipCn-LLipP-RLipCn
male ULipP-LLipCn-LLipP-RLipCn
mixed angle of ULipP-Stm-LLipP
female angle of ULipP-Stm-LLipP
male angle of ULipP-Stm-LLipP
mixed ULipP-Stm-LLipP
female ULipP-Stm-LLipP
male ULipP-Stm-LLipP
mixed angle of LLipCn-Stm-RLipCn
female angle of LLipCn-Stm-RLipCn
male angle of LLipCn-Stm-RLipCn
mixed angle of Sbn-ULipP-Stm
female angle of Sbn-ULipP-Stm
male angle of Sbn-ULipP-Stm
mixed angle of Stm-LLipP-ChiP
female angle of Stm-LLipP-ChiP
male angle of Stm-LLipP-ChiP
mixed angle of LLipCn-ULipP-RLipCn
female angle of LLipCn-ULipP-RLipCn
male angle of LLipCn-ULipP-RLipCn
mixed LLipP-ChiP
female LLipP-ChiP
male LLipP-ChiP

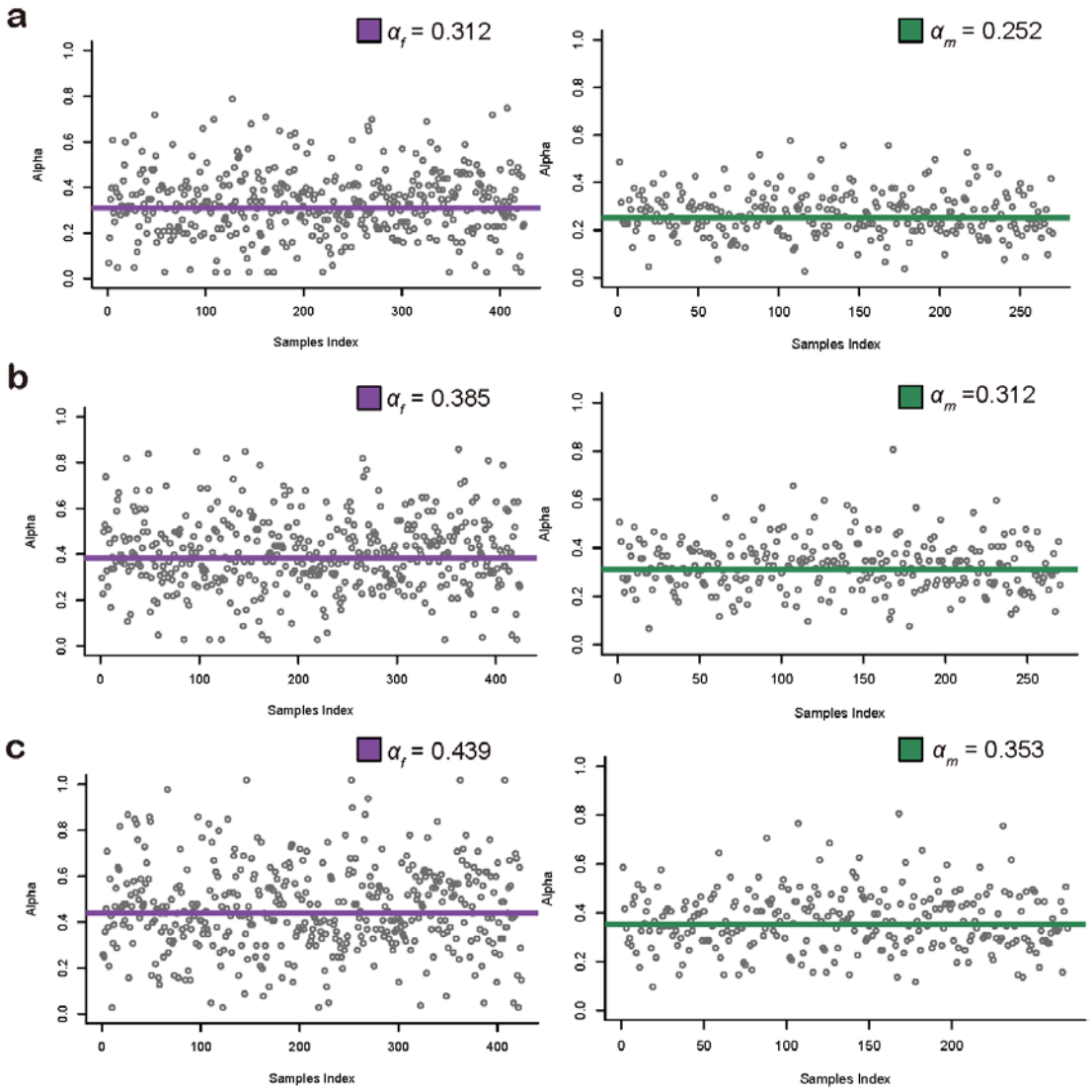

mixed Sbn-ULipP-ChiP
female Sbn-ULipP-ChiP
male Sbn-ULipP-ChiP
mixed angle of Sbn-ULipP-ChiP
female angle of Sbn-ULipP-ChiP
male angle of Sbn-ULipP-ChiP
mixed Sbn-ULipP-LLipP
female Sbn-ULipP-LLipP
male Sbn-ULipP-LLipP
mixed angle of Sbn-ULipP-LLipP
female angle of Sbn-ULipP-LLipP
male angle of Sbn-ULipP-LLipP
b
mixed brow sPC
mixed brow PC1
mixed brow PC2
mixed brow PC3
female brow sPC
female brow PC1
female brow PC2
female brow PC3
male brow sPC
male brow PC1
male brow PC2
male brow PC3
mixed eyes sPC
mixed eyes PC1
mixed eyes PC2
mixed eyes PC3
female eyes sPC
female eyes PC1
female eyes PC2
female eyes PC3
male eyes sPC
male eyes PC1
male eyes PC2
male eyes PC3
mixed side faces sPC
mixed side faces PC1
mixed side faces PC2
mixed side faces PC3
female side faces sPC
female side faces PC1
female side faces PC2
female side faces PC3
male side faces sPC
male side faces PC1
male side faces PC2
male side faces PC3


mixed cheeks sPC
mixed cheeks PC1
mixed cheeks PC2
mixed cheeks PC3
female cheeks sPC
female cheeks PC1
female cheeks PC2
female cheeks PC3
male cheeks sPC
male cheeks PC1
male cheeks PC2
male cheeks PC3
mixed nose sPC
mixed nose PC1
mixed nose PC2
mixed nose PC3
female nose sPC
female nose PC1
female nose PC2
female nose PC3
male nose sPC
male nose PC1
male nose PC2
male nose PC3
mixed mouth sPC
mixed mouth PC1
mixed mouth PC2
mixed mouth PC3
female mouth sPC
female mouth PC1
female mouth PC2
female mouth PC3
male mouth sPC
male mouth PC1
male mouth PC2
male mouth PC3

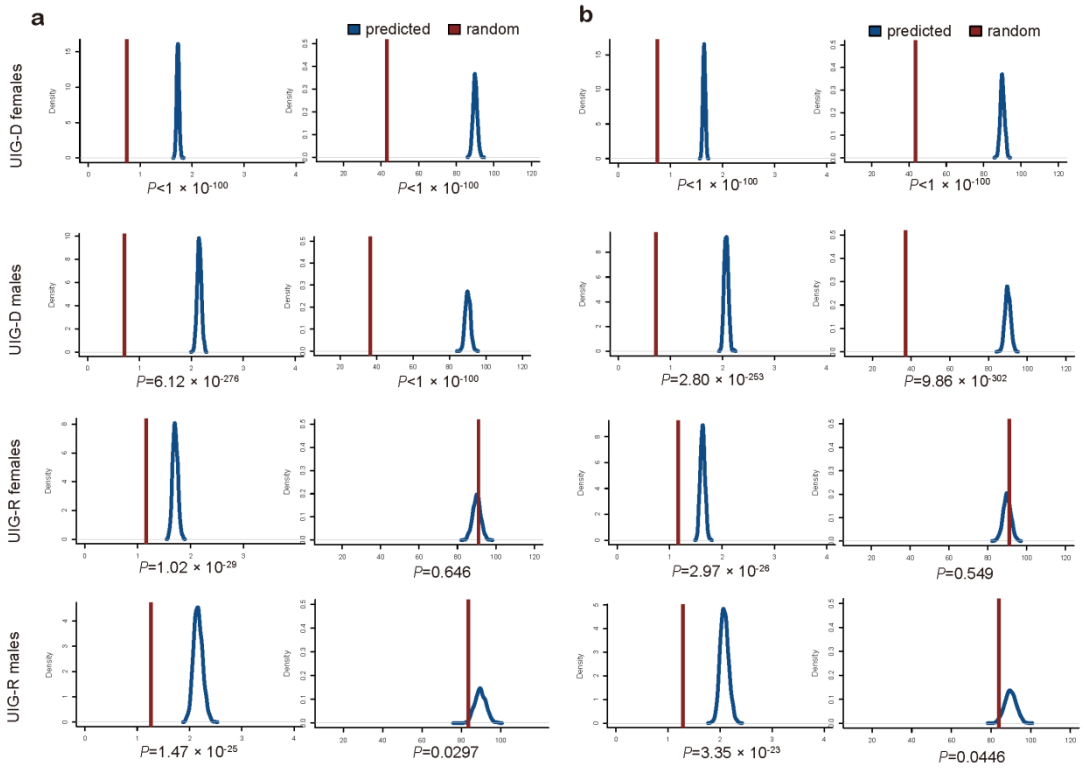
c


mixed brow sPLS
female brow sPLS
male brow sPLS
mixed eyes sPLS
female eyes sPLS
male eyes sPLS
mixed side faces sPLS
female side faces sPLS
male side faces sPLS
mixed cheeks sPLS
female cheeks sPLS
male cheeks sPLS

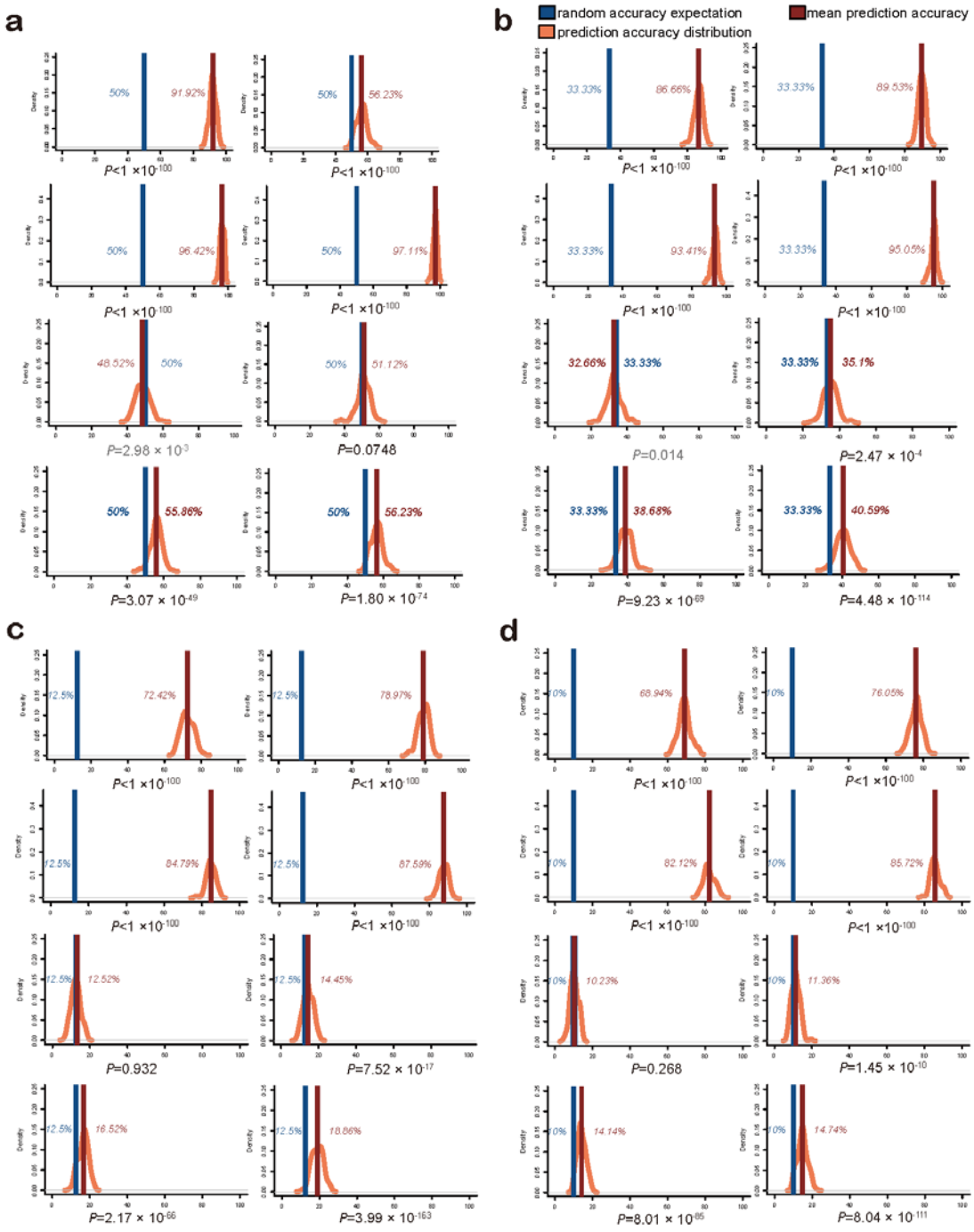
Supplementary Figure 4 | Q-Q plot of all GWAS. (a) inter-landmark phenotyping GWAS, (b) PCA-based phenotyping GWAS, (c) PLS-based phenotyping GWAS.



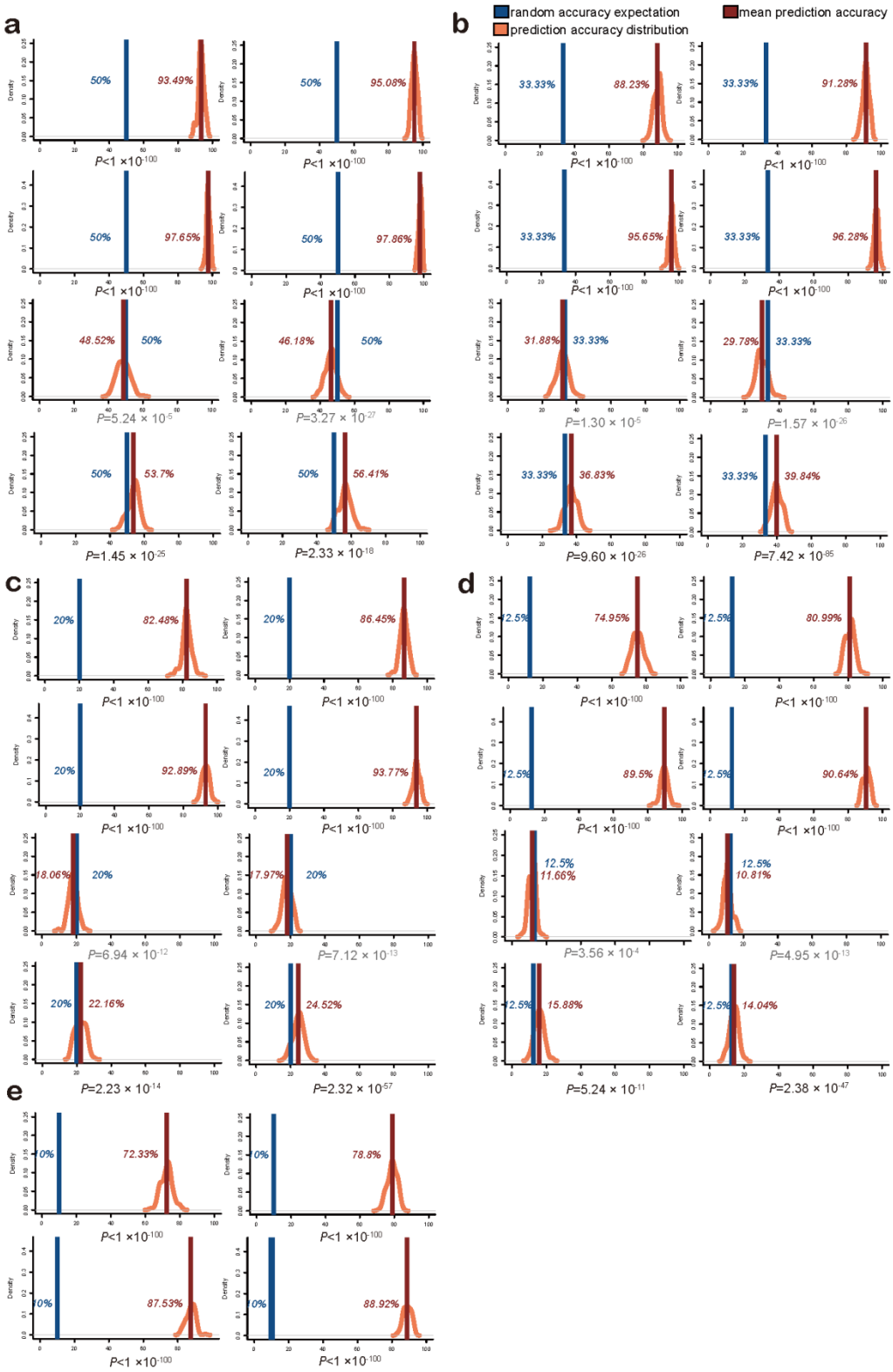
Supplementary Figure 5 | Global effect coefficient α_f and α_m from UIG-D females and UIG-D males. The coefficient α was plotted respectively for (a) the full prediction model of 277 top SNPs; (b) the prediction model of 240 top-SNPs after trimming the SNPs of pairwise LD > 0.8. (c) the prediction model of 209 top-SNPs after trimming SNPs within a physical distance of < 400kb. The global coefficient was obtained by averaging over all individuals.

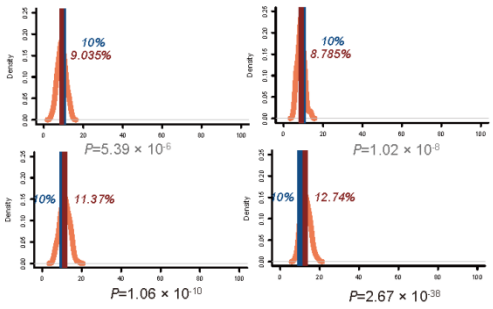


Supplementary Figure 6 | Test of prediction model using 240-SNP and 209-SNP prediction models. For the (a) 240-SNP prediction model, and (b) 209-SNP prediction model, the average PSD (left column) and SSA (right column) determined for the sample cohorts (in red) were compared to the random distributions under null hypothesis (in blue).

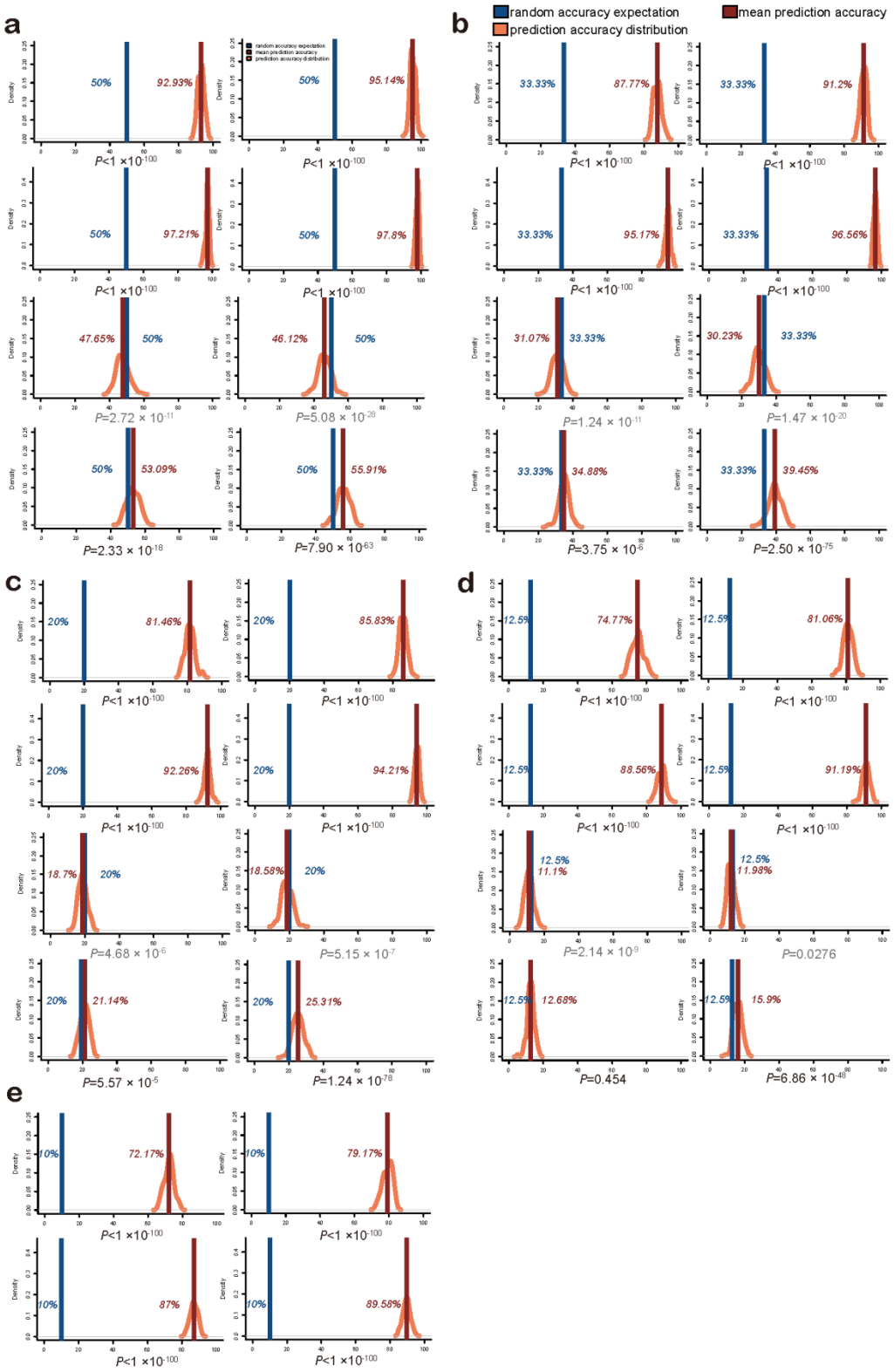


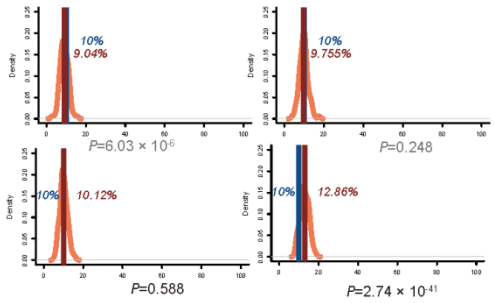
Supplementary Figure 7 | Evaluation of the face prediction in hypothetic forensic scenarios (2/3/8/10 candidates) using 277-SNP face prediction model. (a) N=2, (b) N=3, (c) N=8, (d) N=10. In each scenario, the samples from top row to bottom are UIG-D females, UIG-D males, UIG-R females, UIG-R males. For each cohort, the accuracy rates are decided based on PSD (left column) and SSA (right column). *P* values in black indicate that the prediction perform significantly better than random draws.


a
b
random accuracy expectation
mean prediction accuracy
prediction accuracy distribution
c
d
e
Density
50%
93.49%
P<1 ×10-100
50%
95.08%
P<1 ×10-100
50%
97.65%
P<1 ×10-100
50%
97.86%
P<1 ×10-100
48.52%
50%
P=5.24 × 10-5
46.18%
50%
P=3.27 × 10-27
50%
53.7%
P=1.45 × 10-25
50%
56.41%
P=2.33 × 10-18
33.33%
88.23%
P<1 ×10-100
33.33%
91.28%
P<1 ×10-100
33.33%
95.65%
P<1 ×10-100
33.33%
96.28%
P<1 ×10-100
31.88%
33.33%
P=1.30 × 10-5
29.78%
33.33%
P=1.57 × 10-26
33.33%
36.83%
P=9.60 × 10-26
33.33%
39.84%
P=7.42 × 10-85
20%
82.48%
P<1 ×10-100
20%
86.45%
P<1 ×10-100
20%
92.89%
P<1 ×10-100
20%
93.77%
P<1 ×10-100
18.06%
20%
P=6.94 × 10-12
17.97%
20%
P=7.12 × 10-13
20%
22.16%
P=2.23 × 10-14
20%
24.52%
P=2.32 × 10-57
12.5%
74.95%
P<1 ×10-100
12.5%
80.99%
P<1 ×10-100
12.5%
89.5%
P<1 ×10-100
12.5%
90.64%
P<1 ×10-100
12.5%
11.66%
P=3.56 × 10-4
12.5%
10.81%
P=4.95 × 10-13
12.5%
15.88%
P=5.24 × 10-11
12.5%
14.04%
P=2.38 × 10-47
10%
72.33%
P<1 ×10-100
10%
78.8%
P<1 ×10-100
10%
87.53%
P<1 ×10-100
10%
88.92%
P<1 ×10-100

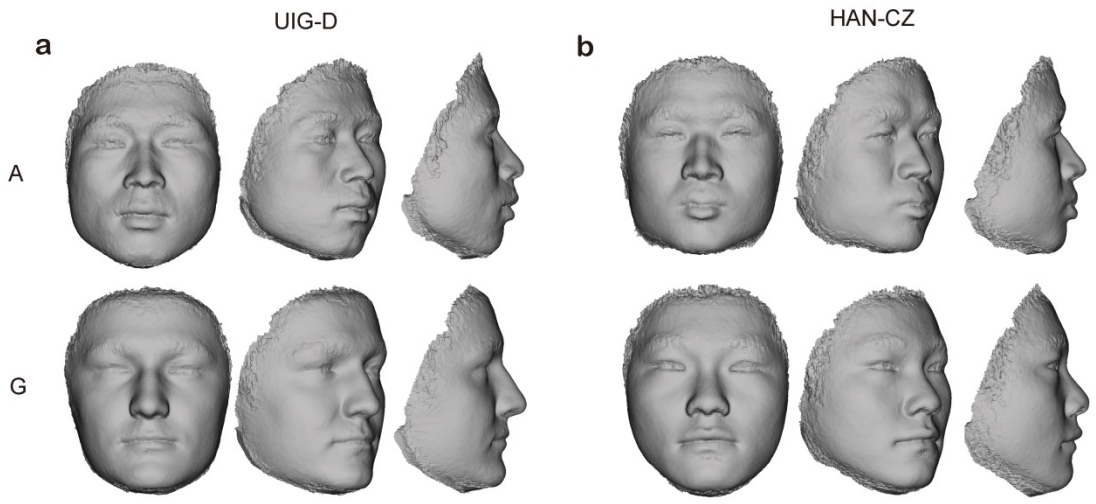


Supplementary Figure 8 | Evaluation of the face prediction in hypothetic forensic scenarios (2/3/5/8/10 candidates) using 240-SNP face prediction model. (a) N=2, (b) N=3, (c) N=5, (d) N=8, (e) N=10. In each scenario, the samples from top row to bottom are UIG-D females, UIG-D males, UIG-R females, UIG-R males. For each cohort, the accuracy rates are decided based on PSD (left column) and SSA (right column). *P* values in black indicate that the prediction perform significantly better than random draws.


a
b
random accuracy expectation
mean prediction accuracy
prediction accuracy distribution
c
d
e
50%
92.93%
95.14%
97.21%
97.8%
47.65%
46.12%
53.09%
55.91%
33.33%
87.77%
91.2%
95.17%
96.56%
31.07%
30.23%
34.88%
39.45%
20%
81.46%
85.83%
92.26%
94.21%
18.7%
18.58%
21.14%
25.31%
12.5%
74.77%
81.06%
88.56%
91.19%
12.5%
11.1%
12.5%
11.98%
12.68%
15.9%
10%
72.17%
79.17%
87%
89.58%
P<1 ×10-100
P=2.72 × 10-11
P=5.08 × 10-28
P=2.33 × 10-18
P=7.90 × 10-63
P=1.24 × 10-11
P=1.47 × 10-20
P=3.75 × 10-6
P=2.50 × 10-75
P=4.68 × 10-6
P=5.15 × 10-7
P=5.57 × 10-5
P=1.24 × 10-78
P=2.14 × 10-9
P=0.0276
P=0.454
P=6.86 × 10-48



Supplementary Figure 9 | Evaluation of the face prediction in hypothetic forensic scenarios (2/3/5/8/10 candidates) using 209-SNP face prediction model. (a) N=2, (b) N=3, (c) N=5, (d) N=8, (e) N=10. In each scenario, the samples from top row to bottom are UIG-D females, UIG-D males, UIG-R females, UIG-R males. For each cohort, the accuracy rates are decided based on PSD (left column) and SSA (right column). *P* values in black indicate that the prediction perform significantly better than random draws.



Supplementary Figure 10 | 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 Movie 1 | 3D visualization of extrapolated face affected by rs1868752. The animation show the extreme facial shape change from Han-trend to European-trend affected by rs1868752. (a) anterior view, (b) cant view, (c) lateral view.

Supplementary Movie 2 | 3D visualization of extrapolated face affected by rs118078182. The animation show the extreme facial shape change from Han-trend to European-trend affected by rs118078182. (a) anterior view, (b) cant view, (c) lateral view.

Supplementary Movie 3 | 3D visualization of extrapolated face affected by rs60159418. The animation show the extreme facial shape change from Han-trend to European-trend affected by rs60159418. (a) anterior view, (b) cant view, (c) lateral view.

Supplementary Movie 4 | 3D visualization of extrapolated face affected by rs17868256. The animation show the extreme facial shape change from Han-trend to European-trend affected by rs17868256. (a) anterior view, (b) cant view, (c) lateral view.

Supplementary Movie 5 | 3D visualization of extrapolated face affected by rs3920540. The animation show the extreme facial shape change from Han-trend to European-trend affected by rs3920540. (a) anterior view, (b) cant view, (c) lateral view.

Supplementary Movie 6 | 3D visualization of extrapolated face affected by rs61672954. The animation show the extreme facial shape change from Han-trend to European-trend affected by rs61672954. (a) anterior view, (b) cant view, (c) lateral view.

Supplementary Movie 7 | 3D visualization of actual face and predicted face. cases of visualization of actual face (left column) and predicted face (right column). (a) case from UIG-D female, (b) case from UIG-D male, (c) case from UIG-R female, (d) case from UIG-R male.

Supplementary Table 1 | Landmark-based phenotypes

Landmrk-based phenotypes	P	
	Female	**Male**
LAla-Prn-RAla	**2.70×10^{-11}**	**1.52×10^{-21}**
Nsn-Prn-Sbn	**2.43×10^{-15}**	**6.71×10^{-23}**
LIntCan-Nsn-RIntCan	**2.38×10^{-10}**	**5.27×10^{-27}**
height of Nsn to eyes	**1.22×10^{-21}**	**5.50×10^{-47}**
Sbn-ULipP	**3.28×10^{-8}**	**2.80×10^{-9}**
LLipCn-RLipCn	**0.00535**	**1.52×10^{-8}**
ULipP-Stm-LLipP	**0.0026**	**7.41×10^{-9}**
ExtCan-IntCan	**1.46×10^{-7}**	**2.58×10^{-10}**
LIntCan-RIntCan	**8.69×10^{-7}**	**4.31×10^{-9}**
Sbn-ULipP-ChiP	**0.00359**	**6.78×10^{-7}**

Supplementary Table 2 | Eurasian-specified phenotypes

		PCA based		PLS based	
Partial Features	**Gender[A]**	**ncomp.[B]**	**P[C]**	**ncomp.[B]**	**P[C]**
brow ridge	Female	2	$\mathbf{5.80 \times 10^{-19}}$	1:6	$\mathbf{6.41 \times 10^{-23}}$
	Male	2	$\mathbf{1.00 \times 10^{-29}}$	1:8	$\mathbf{2.78 \times 10^{-49}}$
eyes	Female	5	$\mathbf{1.69 \times 10^{-16}}$	1:8	$\mathbf{1.45 \times 10^{-25}}$
	Male	2	$\mathbf{8.62 \times 10^{-17}}$	1:11	$\mathbf{1.00 \times 10^{-60}}$
side faces	Female	1	$\mathbf{5.79 \times 10^{-13}}$	1:25	$\mathbf{6.48 \times 10^{-27}}$
	Male	2	$\mathbf{1.71 \times 10^{-17}}$	1:9	$\mathbf{1.94 \times 10^{-47}}$
cheeks	Female	1	$\mathbf{4.71 \times 10^{-15}}$	1:22	$\mathbf{1.10 \times 10^{-27}}$
	Male	2	$\mathbf{5.79 \times 10^{-21}}$	1:8	$\mathbf{4.03 \times 10^{-43}}$
nose	Female	2	$\mathbf{1.43 \times 10^{-22}}$	1:17	$\mathbf{2.94 \times 10^{-29}}$
	Male	2	$\mathbf{2.24 \times 10^{-46}}$	1:12	$\mathbf{6.09 \times 10^{-57}}$
mouth	Female	4	$\mathbf{3.41 \times 10^{-11}}$	1:20	$\mathbf{1.85 \times 10^{-24}}$
	Male	2	$\mathbf{8.73 \times 10^{-18}}$	1:15	$\mathbf{1.72 \times 10^{-48}}$

[A]sex specified test

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

[C]Student's test

Supplementary Table 3 | Inflation factor in GWAS

SNP	P	Inflation factor	genome PC corrected P values	genome PC corrected Inflation factor
rs1868752	$\mathbf{1.22 \times 10^{-10}}$	1	$\mathbf{1.02 \times 10^{-10}}$	1
rs118078182	$\mathbf{8.19 \times 10^{-9}}$	1.00961	$\mathbf{1.09 \times 10^{-8}}$	1.01032
rs60159418	$\mathbf{8.96 \times 10^{-11}}$	1.01437	$\mathbf{1.38 \times 10^{-9}}$	1
rs17868256	$\mathbf{7.22 \times 10^{-9}}$	1.03422	$\mathbf{1.20 \times 10^{-7}}$	1.0013
rs3920540	$\mathbf{3.31 \times 10^{-8}}$	1.06986	$\mathbf{2.28 \times 10^{-7}}$	1.00707
rs61672954	$\mathbf{2.00 \times 10^{-8}}$	1.03794	$\mathbf{1.04 \times 10^{-7}}$	1

Supplementary Table 4 | Previous candidate SNPs associated with Uyghur facial shape

SNP	Chr.	BP[A]	Gene	Discovered Trait[B]	Studied Trait	P		
						Female	Male	Mixed
rs4648379	1	3261516	PRDM16	LAla-Prn-RAla	LAla-Prn-RAla	0.7133	0.172	0.2357
					angle of LAla-Prn-Rala	0.2207	**0.02539**	**0.02028**
rs642961	1	209989270	IRF6	ULipP-Stm-LLipP	ULipP-Stm-LLipP	**0.05197**	0.778	0.2358
rs3827760	2	109513601	EDAR	Chin protrusion	angle of Sbn-ULipP-ChiP	**0.06131**	**0.09733**	**0.01238**
rs7559271	2	223068286	PAX3	Nasion	height of Nasion to eyes	**0.005811**	0.117	**0.001986**
rs17447439	3	189549423	TP63	EyeR-EyeL	LExtCan-RExtCan	0.7631	0.4777	0.9567
rs2045323	4	154831899	DCHS2	Columella inclination/Nose protrusion/Nose tip angle	height of Prn to Ala	**0.05813**	**0.04288**	0.815
rs6184	5	42719344	GHR	mandibular height (ear-Gn)	Ear-ChiP	**0.0915**	0.2674	0.5894
rs1852985	6	45329656	SUPT3H/RUNX2	Nose bridge breadth	LAla-RAla	0.6941	0.1894	0.244
rs7773292	6	132099761	ENPP1	upper and lower face height	Sbn-UlipP-ChiP	0.9239	0.9781	0.9585
rs17640804	7	42131390	GLI3	Nose wing breadth	LAla-RAla	0.1402	0.1242	**0.03421**
rs805722	10	105810400	COL17A1	EyeLR-Nsn	LIntCan-Nsn-RExtCan	0.4636	0.3311	0.9475
					angle of LIntCan-Nsn-RExtCan	**0.004021**	0.8374	**0.01856**
rs927833	20	22041577	PAX1	Nose wing breadth	LAla-RAla	0.4841	**0.007729**	**0.02703**

[A]NCBI build 37

[B]original measurements in previous studies