Supplemental information

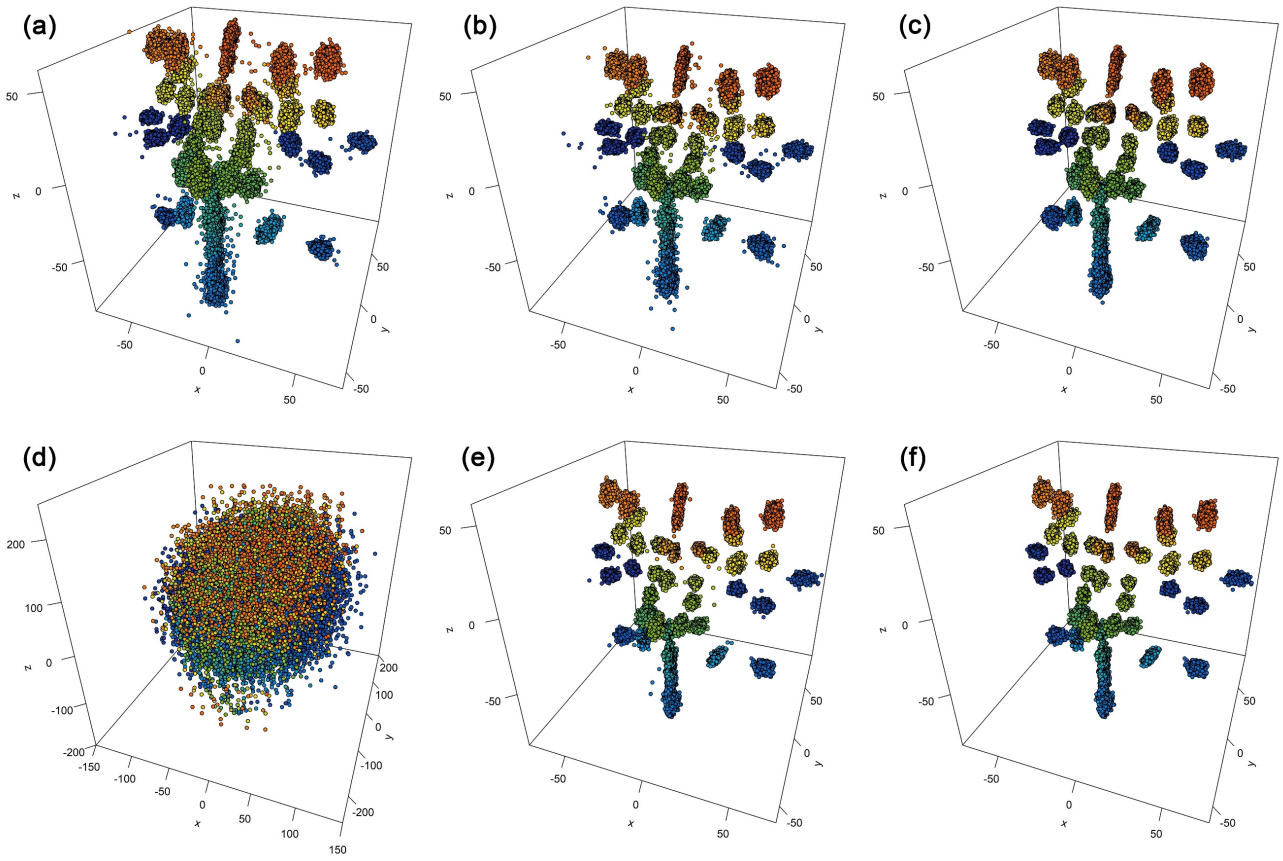
Combined genome-wide association study of facial traits in Europeans increases explained variance and improves prediction

Ziyi Xiong, Yi Li, Xianjing Liu, Haojie Lu, Pirro G. Hysi, Luba M. Pardo, Andre G. Uitterlinden, Fernando Rivadeneira, M. Arfan Ikram, Mohsen Ghanbari, Eppo B. Wolvius, Gennady V. Roshchupkin, Stephen Richmond, Tamar Nijsten, Timothy D. Spector, Sijia Wang, Fan Liu, Manfred Kayser

Contents:

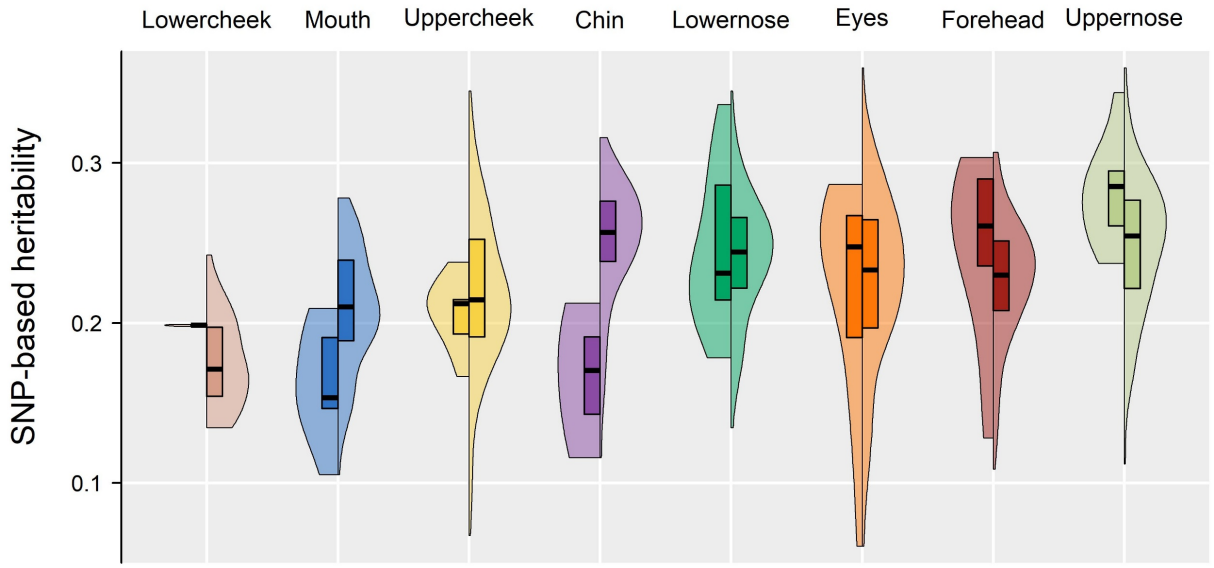
Supplementary Figure 1. Quality control of 44 landmarks in RS and TwinsUK.

(**a**) Original distribution of 44 color-coded landmarks from RS. (**b**) Distribution of 44 color-coded landmarks from RS after GPA. (**c**) Distribution of 44 color-coded landmarks from RS after GPA and outliers excluded. (**d**) Original distribution of 44 color-coded landmarks from TwinsUK. (**e**) Distribution of 44 color-coded landmarks from TwinsUK after GPA. (**f**) Distribution of 44 color-coded landmarks from TwinsUK after GPA and outliers excluded.



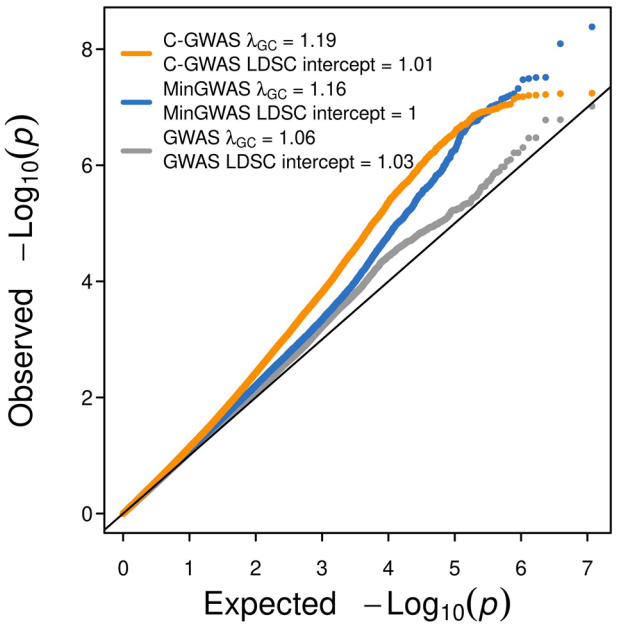
Supplementary Figure 2. Distribution of LDSC heritability for facial traits across eight facial regions.

The violin plots on the left and right for eight color-coded regions represent the increasing heritability of intra-region distance traits and distance traits between this region and other regions, respectively. The band indicates the median, the box indicates the first and third quartiles.



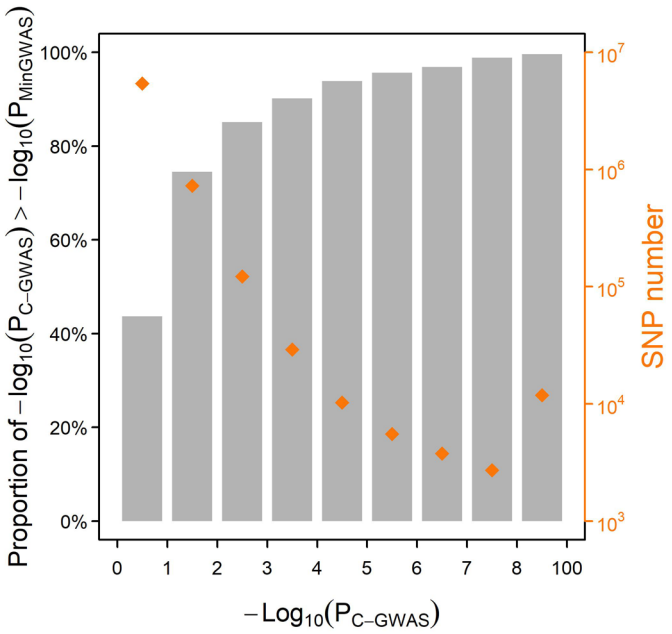
Supplementary Figure 3. QQ plots for C-GWAS, MinGWAS, and the most inflated single-trait GWAS after excluding significant loci.

QQ plots were illustrated for genome-wide p-values from three type of analyses after excluding 188 genomic regions (±500kbp) that are genome-wide significantly associated with facial traits. Grey dots represent results from the GWAS for the facial trait L21_L23, which had the highest inflation among all single trait GWASs. LDSC intercepts and λ_{GC} for the three types of results are indicated.



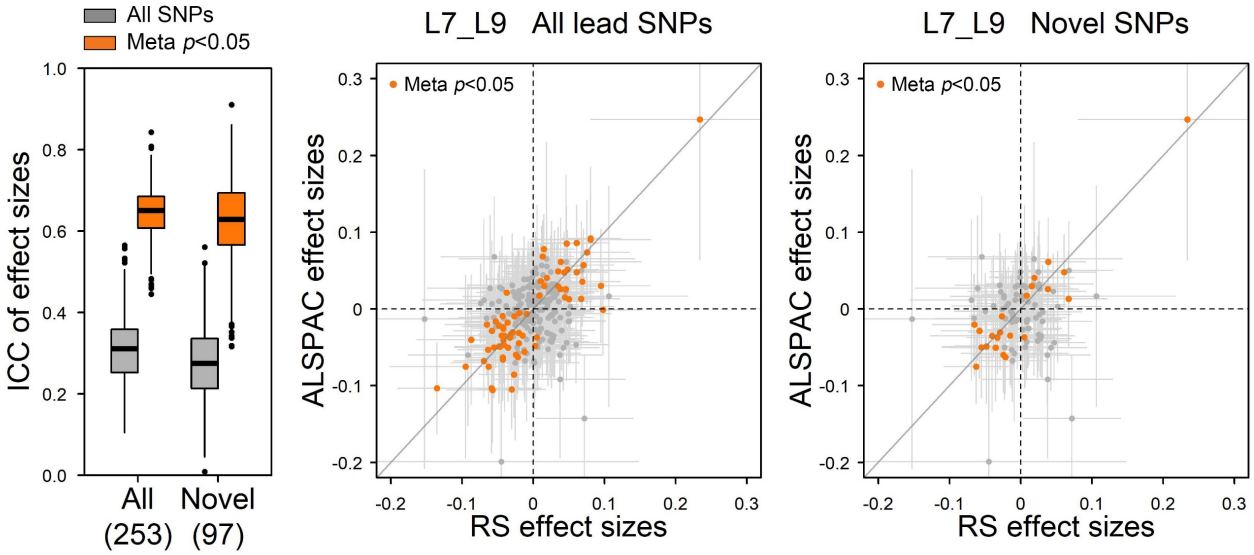
Supplementary Figure 4. Fraction of C-GWAS p-values that surpass MinGWAS p-values in significance.

The proportions (bars) were calculated based on SNP sets classified by different C-GWAS p-value thresholds (x-axis), with the number of SNPs in each set highlighted by orange diamonds.



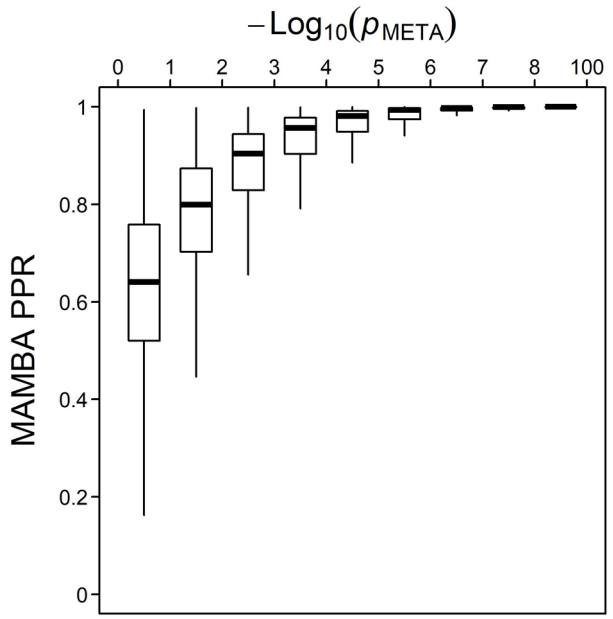
Supplementary Figure 5. Consistency of effect sizes for 253 lead SNPs across cohorts in the meta-analysis, with a representative example.

Left boxplots illustrates the consistency of effect sizes for 253 lead SNPs across all European cohorts involved in the meta-analysis, assessed using the Intraclass Correlation Coefficient (ICC) for each of 514 facial traits. The band indicates the median, the box indicates the first and third quartiles, and the whiskers indicate 1.5 times of interquartile range from box. Results are categorized into two groups (all 253 or 97 novel lead SNPs) and two levels (only SNPs with p < 0.05 in the meta-analysis or not). On the right, the distribution of effect sizes for the all 253 or 97 novel lead SNPs in the two largest cohorts in the meta-analysis (RS and ALSPAC) for nasion width (L7_L9) is taken as an example. The grey cross-shaped error bars indicate the 95% confidence intervals in each cohort, while the orange dots represent SNPs with nominal significance (p < 0.05) in the meta-analysis.



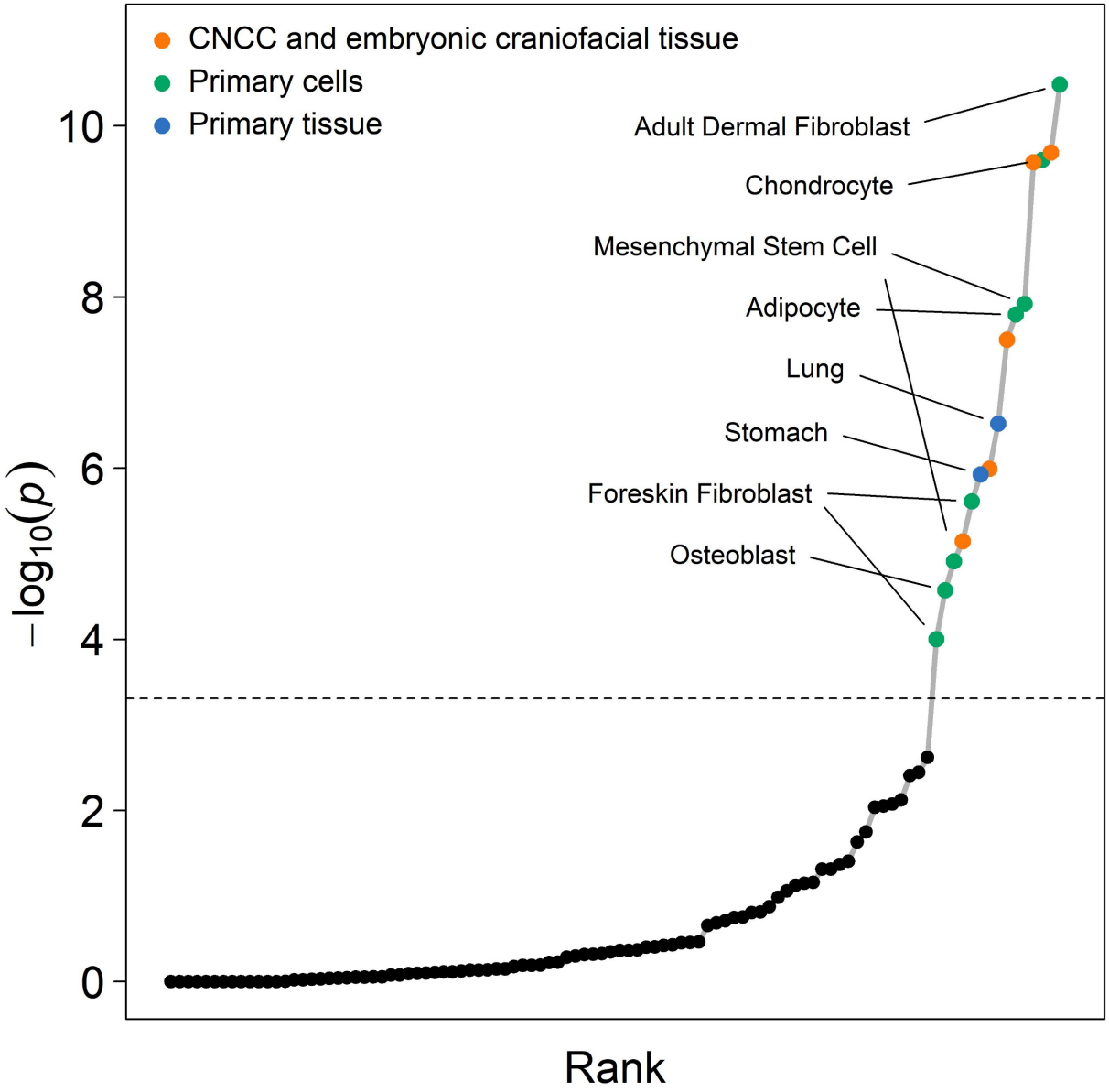
Supplementary Figure 6. Relationship between MAMBA posterior probability of replicability (PPR) and Meta-analysis significance for 253 lead SNPs.

META refers to the meta-analysis of five cohorts. PPRs were illustrated with boxplots based on SNP sets classified by different META p-value thresholds (x-axis). The band indicates the median, the box indicates the first and third quartiles, and the whiskers indicate 1.5 times of interquartile range from box.



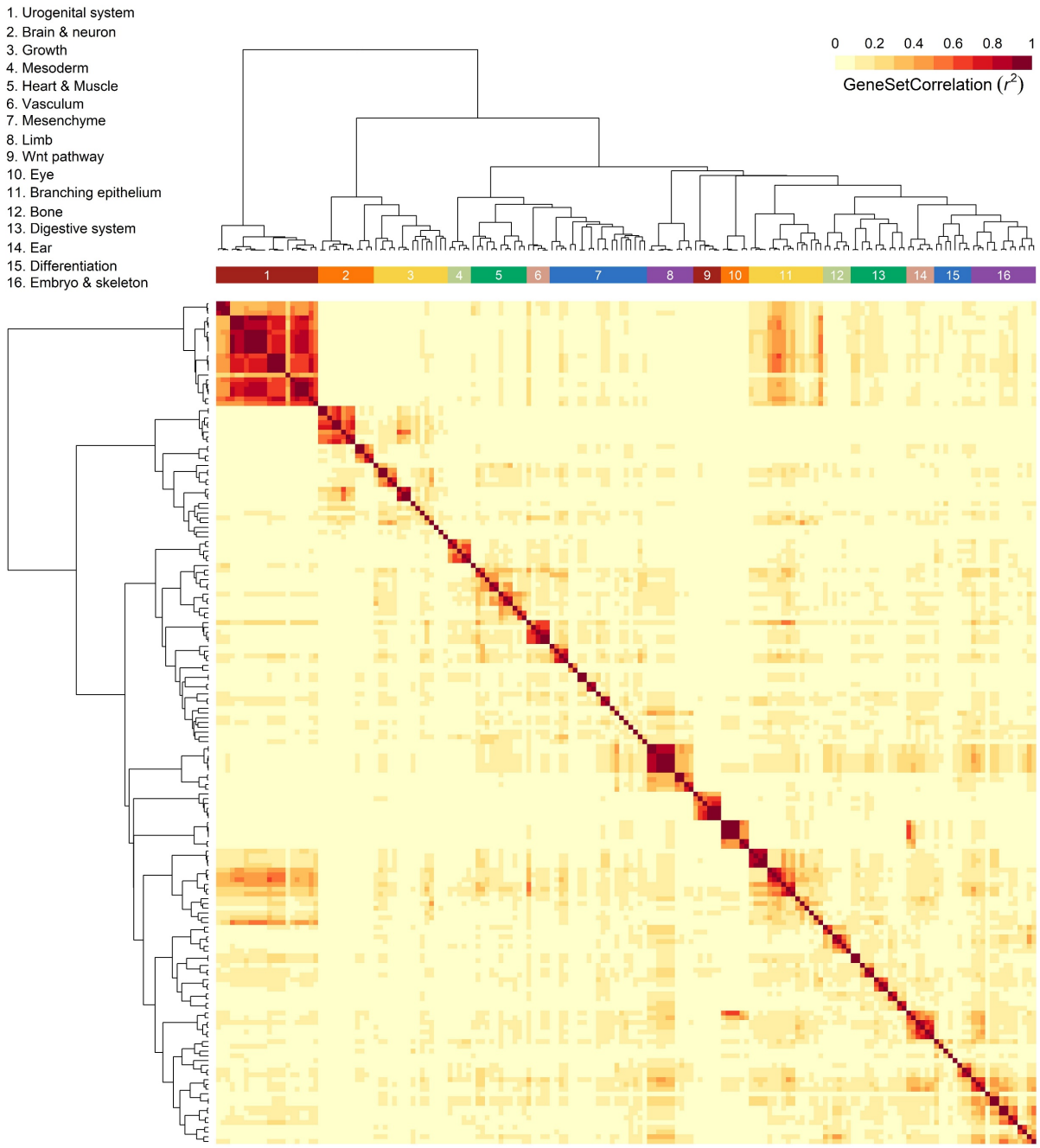
Supplementary Figure 7. Stratified-LDSC of C-GWAS in 102 tissue or cell types.

One-side p-values of z-score from stratified-LDSC results of C-GWAS in 102 tissue or cell types were plotted in decreasing order. Those significant after Bonferroni correction ($p < 4.9\times10^{-4}$) were highlighted with three colour representing different cell or tissue types and were labelled in the upper left corner. The specific types in primacy cells and primacy tissue were annotated besides.



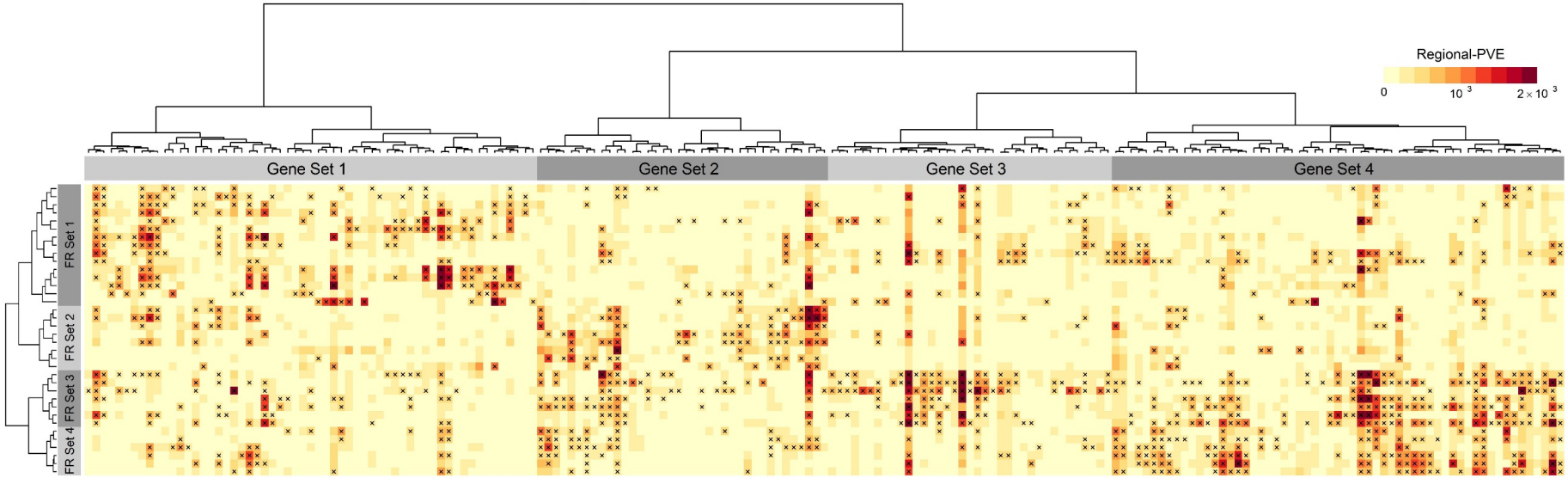
Supplementary Figure 8. Hierarchical clustering of 177 biological processes.

The biological processes were clustering based on their pair-wise gene set correlation (see Methods). The dendrograms of hierarchical clustering were illustrated symmetrically in the left and top. Heatmap showed squared gene set correlation among these biological processes. 16 clusters were indicated in different colour and number labels and annotated in the top left.



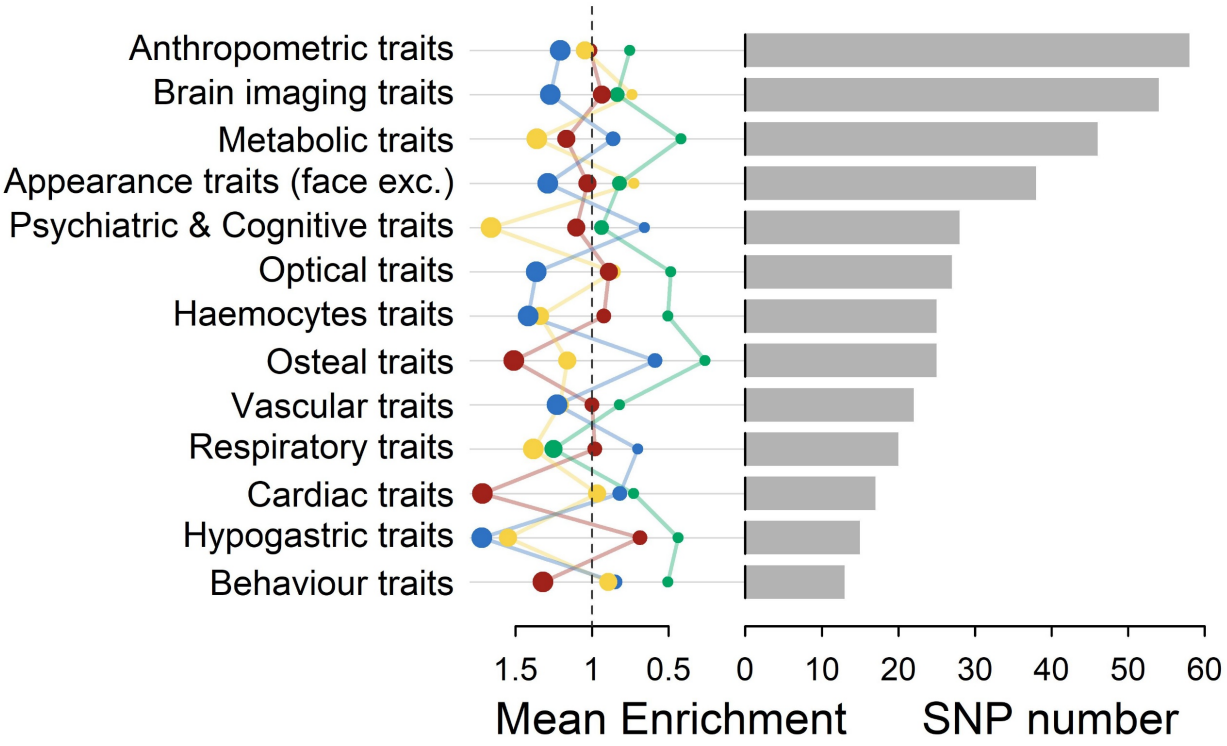
Supplementary Figure 9. Hierarchical clustering of regional-PVE across 8 facial regions and face candidate genes.

Hierarchical clustering was separately applied in 36 regional-PVEs and 193 face candidate genes, where dendrograms of hierarchical clustering and distinction between 4 clusters were illustrated in the left and top, respectively. Heatmap showed the regional-PVE with significant those ($p < 0.001$ after Bonferroni correction) highlighted in crosses.



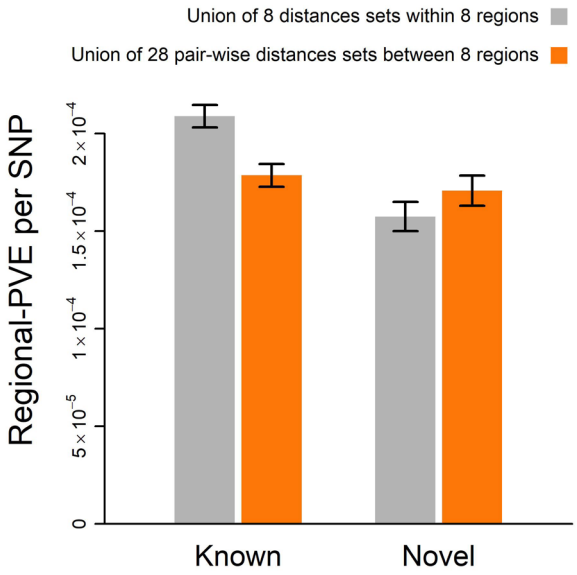
Supplementary Figure 10. GWAS Catalog analysis reveals the pleiotropic effects of 253 lead SNPs.

The pleiotropic traits associated with the 253 lead SNPs are manually categorized into 13 categories, marked on the left. The bar chart on the right indicates the number of lead SNPs in each category with corresponding pleiotropic association. The central plots illustrate the fold enrichment of gene set annotated lead SNPs across the 13 categories, with different colours denoting different gene sets as shown in Figure 3. The dashed line represents the fold enrichment under the null distribution (no enrichment).



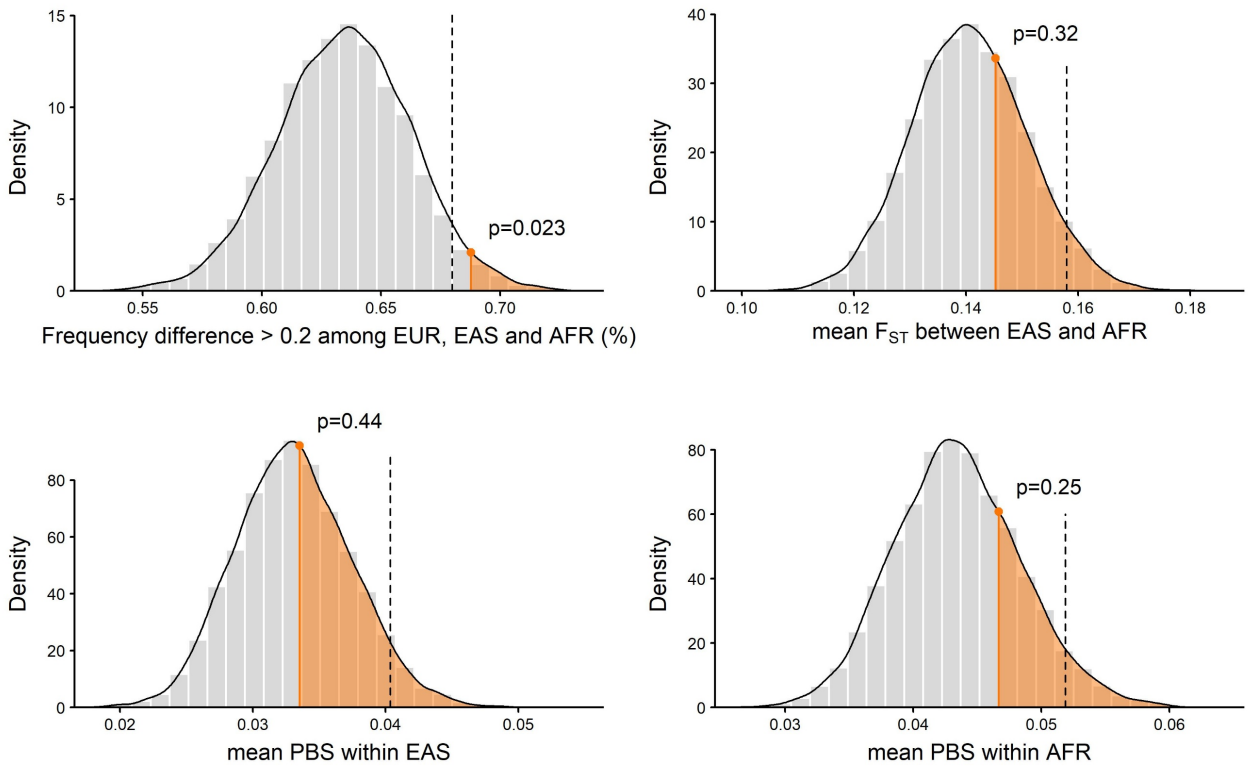
Supplementary Figure 11. Different distributed tendency of known and novel lead SNPs on effects within and between facial regions.

Bar plots displayed the mean regional-PVE of distances within 8 regions (grey, represent specific effect within regions) and between 8 regions (orange, represent widespread effect between regions) in two level: per reported SNP and per novel SNP. Error bar indicated 95% confidence interval estimated from 10,000 times simulation (see Methods).



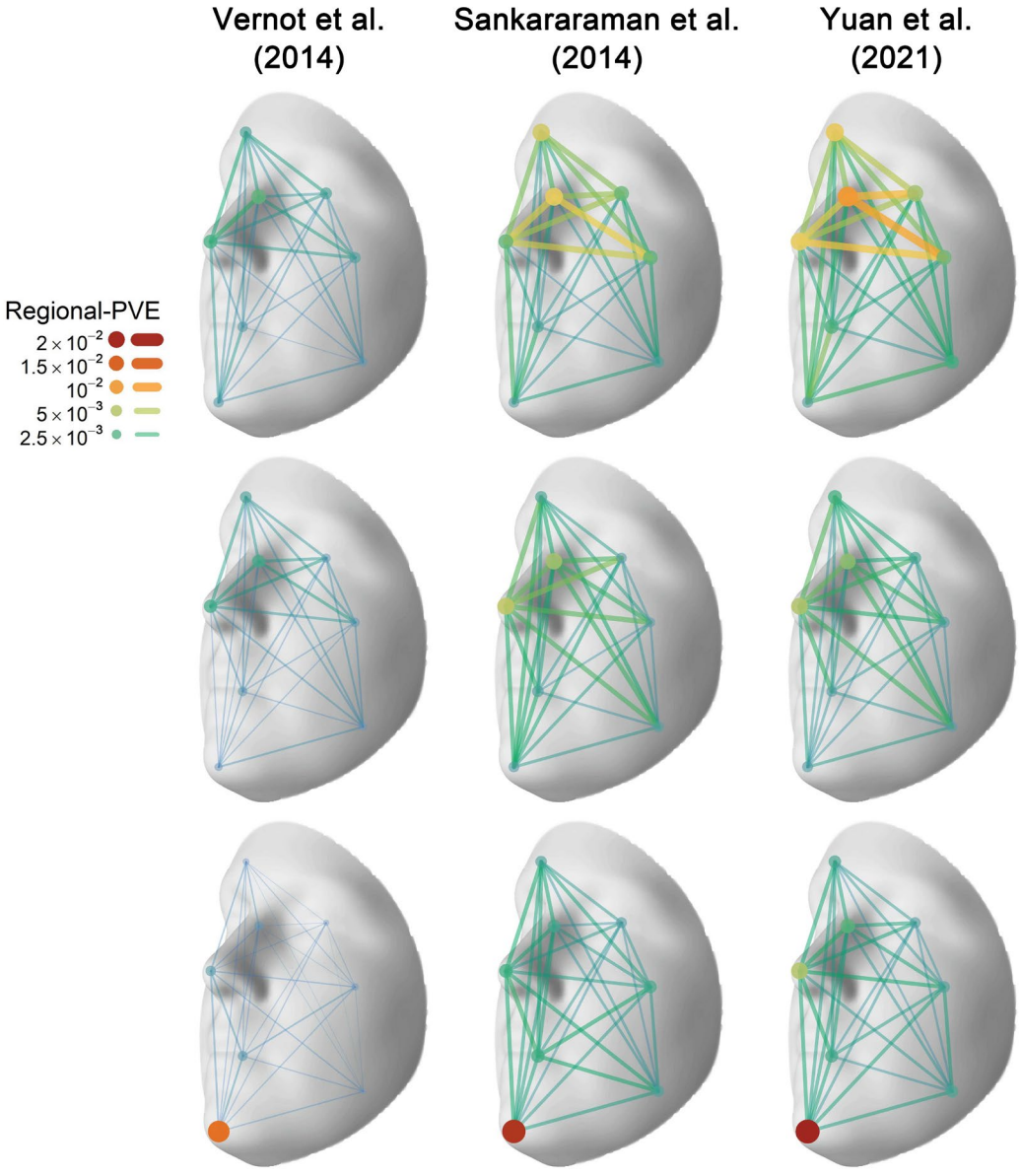
Supplementary Figure 12. Face-associated SNPs show frequency difference and are not under natural selection in EAS and AFR.

Grey bars and the curve illustrate the density of the expected distribution estimated from 10,000 times simulation under the null hypothesis. The dashed line indicates the cumulative density threshold of 0.95. Observed values derived from 253 lead SNPs are highlighted in orange and annotated with their one-sided empirical p-values.



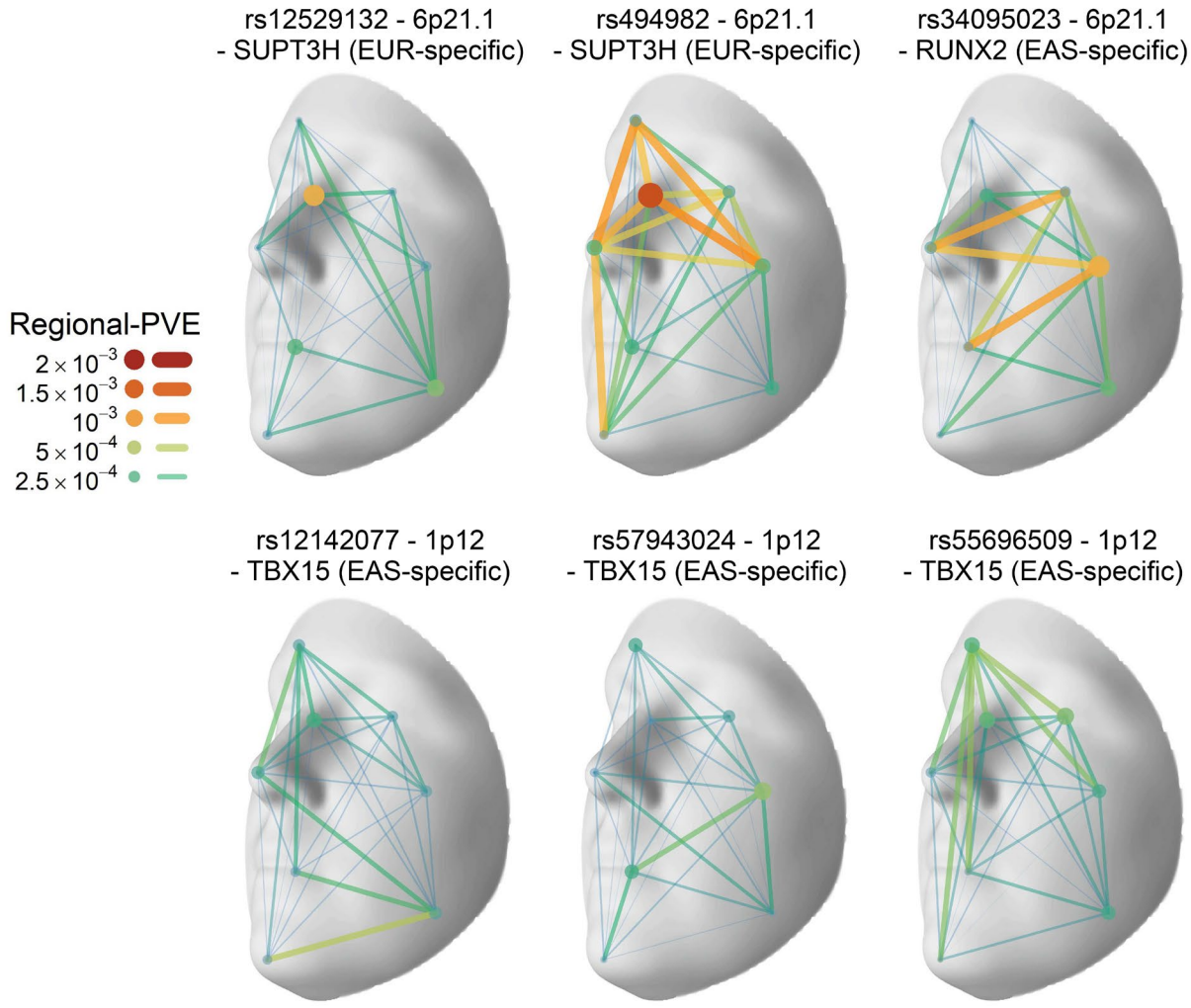
Supplementary Figure 13. Accumulated effect of lead SNPs in Neanderthal introgressed segments identified across three studies.

Accumulated effect of lead SNP sets were illustrated on face using summarized regional-PVE across 8 facial regions. The SNP sets used in first, second, and third column were lead SNPs overlapped with introgressed segments identified in studies of Vernot et al. (DOI: 10.1126/science.1245938), Sankararaman et al. (DOI: 10.1038/nature12961), and Yuan et al. (DOI: 10.1038/s41467-021-26503-5), respectively. The SNP sets used in the top, middle, and bottom rows are lead SNPs overlapped with introgressed fragments 1) specifically identified in EUR, 2) identified in both EUR and EAS, and 3) specifically identified in EAS, respectively. The 3D template facial image in this figure is adapted from White et al.[18] published under an Open Access license (CC BY 4.0), see http://creativecommons.org/licenses/by/4.0/.



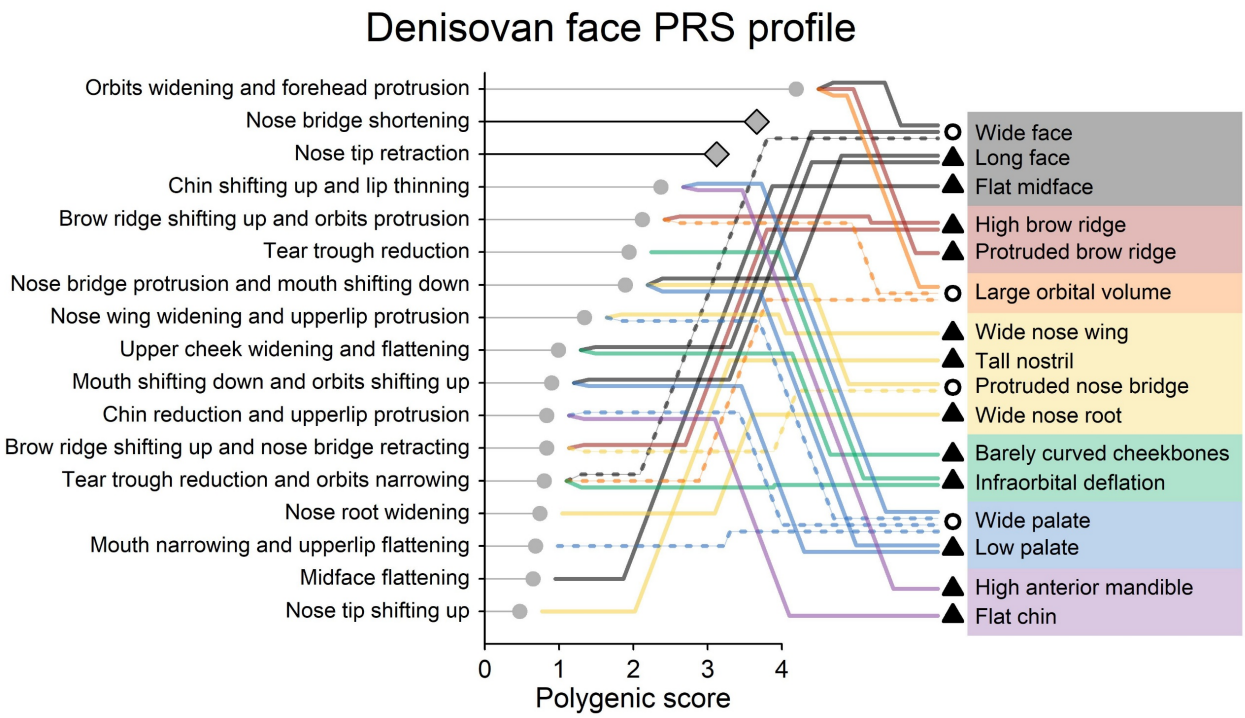
Supplementary Figure 14. Facial effect of some key SNPs involved in Neanderthal introgression.

Facial effect of selected SNPs was illustrated on face using their regional-PVE across 8 facial regions. Annotation of SNPs include rsID, ctyoband, the most prioritized gene, and overlapped introgressed segments types (located in the segment specifically identified in EUR, EUR-specific; located in the segment specifically identified in EAS, EAS-specific). The 3D template facial image in this figure is adapted from White et al.[18] published under an Open Access license (CC BY 4.0), see http://creativecommons.org/licenses/by/4.0/.



Supplementary Figure 15. Facial PRS profiles of Denisovan.

The profiles is especially illustrated for Denisovan 3. The left panel sorts facial PRS averages by their absolute mean values (>0.4), and annotated with summaries of their facial impact (see Methods). The right panel included 16 predicted facial features of Denisovan, categorized by colour. These include those of entire face (grey), forehead (red), eyes (orange), nose (yellow), cheek (green), mouth (blue), and chin (purple). The line connecting PRS to the characteristic indicate whether the observed PRS results in enhancement (solid) or reduction (dashed) of this characteristic. Markers in front of the characteristics summarize the PRS-based predictions: triangles indicate consistent directional predictions across multiple PRS or only one PRS involved, with upward triangles showing that the characteristic is more pronounced, and downward triangles showing it is less pronounced. Circles represent inconsistent directional predictions, suggesting minimal or indeterminate differences for this characteristic. PRS that do not correspond to known characteristics are highlighted with a diamond shape.



Supplementary Table 1. Information of cohorts in discovery phase of GWAS and meta-analysis.

Demographic and phenotyping information:

Cohort	Country	Individual	N	Male%	Average age	SD of age	Covariance
RS	the Netherlands	Unrelated	4242	42.9	68.9	9.5	Sex, Age, First 4 genomic PCs
TwinsUK	the United Kingdom	Twins	1080	0	59.6	9.4	Age, First 4 genomic PCs
US-P	the United States	Unrelated	1990	30.7	33.4	16.8	Sex, Age, Age^2, First 4 genomic PCs, Height, Weight
US-I	the United States	Unrelated	784	31.3	23.7	10.7	Sex, Age, Age^2, First 4 genomic PCs, Height, Weight
ALSPAC	the United Kingdom	Unrelated	3566	47.2	15.3	0.5	Sex, Age, Age^2, First 4 genomic PCs, Height, Weight

Genotyping and quality control information:

Cohort	SNP panel	Reference Data	Imputation	Imputation QC	Missingness QC	MAF QC	HWE QC
RS	Illumina HumanHap 550K	1000G Phase I v.3	MACH	R2 > 0.8	0.03	0.01	10^{-4}
TwinsUK	Illumina HumanHap300 and HumanHap610Q	1000G Phase I v.2	IMPUTE2	INFO > 0.8	0.03	0.01	10^{-4}
US-P	v3 and v4 arrays of 23andMe	1000G Phase III	Sanger Imputation Server	INFO > 0.8	0.5	0.01	10^{-6}
US-I	Illumina's Infinium Multi-Ethnic Global-8 v1	1000G Phase III	Sanger Imputation Server	INFO > 0.8	0.5	0.01	10^{-6}
ALSPAC	Illumina HumanHap 550K	1000G Phase I v.3	IMPUTE2	From database	0.05	0.01	5×10^{-7}

Supplementary Table 2. Descriptions of 44 landmarks

Landmarks	Full name	Descripion
1	Upper glabella	The point of upper glabella
2	Right superciliary arch	The point of right superciliary arch
3	Right top superciliary arch	The point of right top superciliary arch
4	Glabella	The point of glabella
5	Left top superciliary arch	The point of left top superciliary arch
6	Left superciliary arch	The point of left superciliary arch
7	Right nasion	The right point of nose bridge
8	Nasion	The point where bridge of the nose meets forehead
9	Left nasion	The left point of nose bridge
10	Right palpebrale superius	The point of right palpebrale superius
11	Right Exocanthion	The point of right lateral outer canthus
12	Right palpebrale inferius	The point of right palpebrale inferius
13	Right endocanthion	The point of right lateral inner canthus
14	Left palpebrale superius	The point of left palpebrale superius
15	Left exocanthion	The point of left lateral outer canthus
16	Left palpebrale inferius	The point of left palpebrale inferius
17	Left endocanthion	The point of left lateral inner canthus
18	Right nasal bone	The point of right nasal bone
19	Top nasal bone	The point of top nasal bone
20	Left nasal bone	The point of left nasal bone
21	Right top alare	The point of right top lateral nose wing
22	Pronasale	The point of the nose tip
23	Left top alare	The point of left top lateral nose wing
24	Right corner alare	The point of lower corner of right nose wing
25	Right bottom alare	The point of right bottom lateral nose wing
26	Left corner alare	The point of lower corner of left nose wing
27	Left bottom alare	The point of left bottom lateral nose wing
28	Subnasale	The point where base of nasal septum meets philtrum
29	Labial superius	The point of labial superius
30	Labial middle	The point of labial middle
31	Labial inferius	The point of labial inferius
32	Right cheilion	The point of right lateral angulus oris
33	Left cheilion	The point of left lateral angulus oris
34	Up pogonion	The point of up pogonion
35	Top pogonion	The point of top pogonion
36	Menton	The point of menton
37	Right angulus mandibulae	The point of right angulus mandibulae
38	Left angulus mandibulae	The point left right angulus mandibulae
39	Right zygomatic bone	The point of right zygomatic bone
40	Right top zygomatic bone	The point of right top zygomatic bone
41	Right inferior orbital bone	The point of right inferior orbital bone
42	Left zygomatic bone	The point of left zygomatic bone
43	Left top zygomatic bone	The point of left top zygomatic bone
44	Left inferior orbital bone	The point of left inferior orbital bone

Supplementary Table 3. Information of 10 archaic specimens in AADR.

The information of specimens were derived from Allen Ancient DNA Resource (AADR) (Dataverse 8.0). The estimated time period during the archaic individual lived were shown in column: Date. Latitude and longitude coordinates of the discovery site were shown in column: Lat. and Long.. The specific archaic human taxon the specimen belonged to were shown in column: Species.

IID	Date	Lat.	Long.	Sex	Species	Missing Rate
Les_Cottes_noUDG.SG	42846	na	na	F	Neanderthal	0.18
Goyet_noUDG.SG	42436	50.446	5.008	F	Neanderthal	0.18
AltaiNeanderthal.DG	110450	51.409	84.689	F	Neanderthal	0
Chagyrskaya_noUDG.SG	80000	na	na	U	Neanderthal	0.07
Mezmaiskaya2_noUDG.SG	43449	44.167	40	M	Neanderthal	0.28
Spy_noUDG.SG	40922	50.482	4.669	M	Neanderthal	0.53
Vindija_snpAD.DG	50000	46.299	16.071	F	Neanderthal	0.07
VindijaG1_noUDG.SG	50000	46.299	16.071	F	Neanderthal	0.37
Denisova3.DG	63900	51.398	84.676	F	Denisova	0
Denisova11_noUDG.SG	98700	51.398	84.676	F	DenisovaNeanderthalMix	0.23

Supplementary Table 4. The correspondence between facial PRS profile and 16 facial features of Neanderthals.

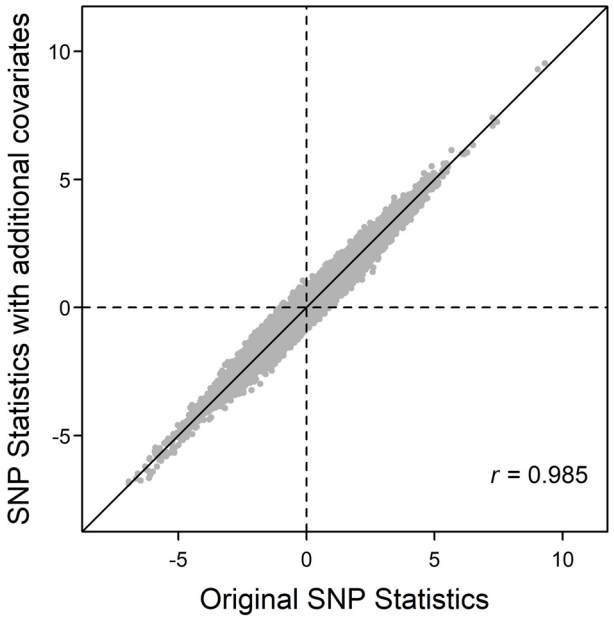
16 facial features of Neanderthals were shown in column: Features, and facial regions they belonged to: Position. Corresponding PRS were shown in column: KeyPhe and Description, which were based on those in Supplementary Data 14. Column "Direction" indicated that the feature was related to the increase (+) or the decrease (-) in PRS of corresponding facial trait.

Position	Features	KeyPhe	Direction	Description
Face	Wide face	13_39+17_42	-	Upper cheek widening and flattening
Face	Wide face	7_12+9_16	+	Orbits widening and forehead protrusion
Face	Wide face	35_39+35_42	+	Chin protrusion and upper cheek widening
Face	Wide face	11_41+15_44	+	Tear trough enlargement and orbits widening
Face	Long face	19_32+19_33	+	Nose bridge protrusion and mouth shifting down
Face	Long face	37_39+38_42	+	Cheek elongation
Face	Long face	13_33+17_32	+	Mouth shifting down and orbits shifting up
Face	Midface protrusion	2_38+6_37	-	Midface protrusion
Forehead	Low brow ridge	1_41+1_44	-	Brow ridge shifting down and orbits retraction
Forehead	Low brow ridge	3_27+5_25	-	Brow ridge shifting down and nose bridge protrusion
Forehead	Projecting brow ridge	7_12+9_16	+	Orbits widening and forehead protrusion
Eye	Large orbital volume	1_41+1_44	-	Brow ridge shifting down and orbits retraction
Eye	Large orbital volume	7_12+9_16	+	Orbits widening and forehead protrusion
Eye	Large orbital volume	11_41+15_44	+	Tear trough enlargement and orbits widening
Nose	Wide nose wing	25_30+27_30	+	Nose wing widening and upperlip protrusion
Nose	Tall nostril	22_28	+	Nose tip shifting up
Nose	Projecting nose bridge	19_32+19_33	+	Nose bridge protrusion and mouth shifting down
Nose	Projecting nose bridge	3_27+5_25	-	Brow ridge shifting down and nose bridge protrusion
Nose	Retracting nose root	7_8+8_9	-	Nose root reduction
Nose	Retracting nose root	13_17	-	Nose root narrowing
Cheek	Barely curved cheekbones	13_39+17_42	-	Upper cheek widening and flattening
Cheek	Barely curved cheekbones	29_40+29_43	-	Upperlip flattening and upper cheek flattening
Cheek	Infraorbital inflation	12_41+16_44	+	Tear trough enlargement
Cheek	Infraorbital inflation	11_41+15_44	+	Tear trough enlargement and orbits widening
Mouth	Wide palate	25_30+27_30	-	Nose wing narrowing and upperlip flattening
Mouth	Wide palate	34_36	+	Chin enlargement and upperlip flattening
Mouth	Wide palate	29_34	-	Chin shifting up and lip thinning
Mouth	Wide palate	29_32+29_33	+	Mouth widening and upperlip protrusion
Mouth	Wide palate	29_40+29_43	-	Upperlip flattening and upper cheek flattening
Mouth	Low palate	19_32+19_33	+	Nose bridge protrusion and mouth shifting down
Mouth	Low palate	13_33+17_32	+	Mouth shifting down and orbits shifting up
Chin	High anterior mandible	29_34	-	Chin shifting up and lip thinning
Chin	Flat chin	34_36	-	Chin reduction and upperlip protrusion
Chin	Flat chin	35_39+35_42	-	Chin flattening and upper cheek narrowing

Supplementary Note 1

To evaluate the influence of varying covariates on our GWAS findings (Supplementary Table 1), we performed a sensitivity analysis in the RS cohort using the 253 lead SNPs. We compared GWAS results derived from models including additional covariates (squared age, height, and weight) with those derived from models without these covariates (Supplementary Fig. 16). The results showed minimal differences, suggesting that these additional covariates had a negligible impact on our GWAS outcomes. Consequently, we anticipate that similar minor effects would apply to other cohorts, even if their covariate sets differ slightly.

Supplementary Figure 16. Correlation between T-statistics with or without covariates in association testing of 253 lead SNPs in RS.



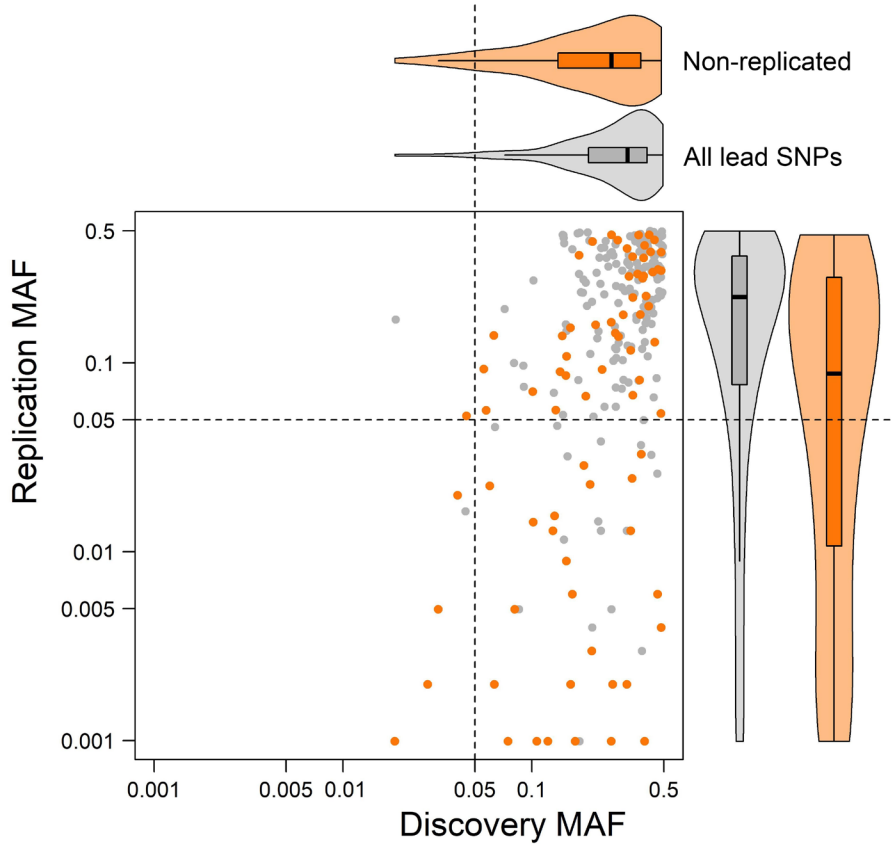
Supplementary Note 2

We downloaded the summary statistics from a recently facial GWAS in East Asians[1], which included 9,674 Chinese individuals from 3 independent cohorts, i.e., the National Survey of Physical Traits (NSPT, N = 3,322), the Northern Han Chinese (NHC, N = 4,767) and the Taizhou Longitudinal Study (TZL, N = 2,881). As reported by Zhang et al.[1], all participants provided written informed consent, and all study protocols were approved by the institutional review boards of the pertinent research institutions. NSPT is a subproject of The National Science & Technology Basic Research Project approved by the Ethics Committee of Human Genetic Resources of School of Life Sciences, Fudan University, Shanghai (14117). NHC was approved by the Ethics Committee of Human Genetic Resources at the Shanghai Institute of Life Sciences, Chinese Academy of Sciences (ER-SIBS-261410-A1801). TZL was approved by the Ethics Committee of Human Genetic Resources at the Shanghai institute of life Sciences, Chinese Academy of Sciences (ER-SIBS-261410). Written informed consent was granted for each participant before enrolment in the study. See Zhang et al.[1] for further details. The data contains 63 sets of GWAS summary statistics corresponding to the 63 facial segments. Similar to the adjustment applied to the minimal p-values of the 514 meta-analyses in the discovery phase, we adjusted the minimal p-values of the GWASs for 63 facial segments using simulations implemented in C-GWAS. This ensured that the resultant p-values followed a uniform distribution under the null hypothesis, making adjusted $p < 0.05$ correspond to nominally significant replicated p-values. Given the different ancestry and phenotyping methods between the discovery study and the study for replication, we extracted adjusted p-value from 253 lead SNPs and their LD counterparts ($r^2 > 0.2$), and regarded the minimal adjusted p-values among LD clusters of each lead SNPs as raw replicated results. To account for multiple testing issue, the raw p-values were first adjusted for the number of SNPs separately in each of LD clusters using the method of effective number of independent tests described before[2] and then for the number of LD clusters using FDR correction.

We cross-referenced the 253 lead SNPs identified by the discovery C-GWAS in Europeans with findings from a recent facial GWAS[1] involving 9,674 East Asian individuals from China. Although reflecting cross-continental replication, an LD-based analysis (see above) revealed high replication rates: 69.8% overall and 39.6% for the novel SNPs (FDR corrected $p < 0.05$, Supplementary Data 5). Notably, as not unexpected, the allele frequencies of these 253 lead SNPs were significantly higher in our European discovery dataset compared to the East Asian replication dataset (18.7% vs. 2.8% MAF < 0.05). This disparity was particularly marked in the 76 non-replicated SNPs, where the differences were even more pronounced (38.2% vs. 6.6% MAF < 0.05). Furthermore, the 76 non-replicated SNPs exhibited much lower allele frequencies in the Chinese population compared to the full set of 253 lead SNPs (χ^2 test $p = 9.7\times10^{-4}$), although no significant difference was observed in the European cohort (χ^2 test $p = 0.16$, Supplementary Fig. 17). These findings suggest conservative replication rates, but also underscore a robust pan-ethnic genetic influence on facial features in Europeans and East Asians.

Supplementary Figure 17. Cross-ancestry replication of 253 lead SNPs.

Scatterplots shows the MAF of 253 lead SNPs in the European samples used for discovery and in East Asian samples used for cross-ancestry replication. Orange dots represent lead SNPs with insufficient cross-ancestry replication evidence (FDR p > 0.05). The distributions of MAF for these SNPs in the European discovery and East Asian replication are shown in violin plots above and to the right of the scatterplots. The band indicates the median, the box indicates the first and third quartiles, and the whiskers indicate 1.5 times of interquartile range from box. The dashed line indicates the threshold that defines low-frequency SNPs (MAF < 0.05).



Supplementary Note 3

We included 5 facial development relevant cell types (CNCC[3] and 4-stages of embryonic craniofacial tissue types[4]) from special studies and 97 general cell types from Roadmap Epigenomics resource[5] in stratified LDSC[6] based heritability enrichments analyses (Supplementary Data 6). The details of annotations used for stratified LDSC analyses have been described in the previous study[7]. In brief, for CNCC, we annotated all enhancer and super-enhancer regions inferred from H3K27ac and ATAC peaks. For embryonic craniofacial tissues, we combined and annotated following regions indicated by 25-state chromHMM model, including 'Enh', 'TxReg', 'PromD1', 'PromD2', 'PromU' and 'TssA'. For other cell types, we combined and annotated following regions indicated by 15-state chromHMM model, including '1_TssA', '2_TssAFlnk', '6_EnhG' and '7_Enh'. Each annotation was individually added to the baseline LD model[6]. We regarded one-side p-value under the standard normal distribution of z-score of enrichment coefficient from stratified LDSC as significance of heritability enrichment. For SNP-based annotation, overlap between promoter or enhancer regions indicated by chromHMM model and 253 lead SNPs and their high LD counterparts (r^2 >0.6) were investigated (Supplementary Data 10). We used the following criteria to classify states in 25-state chromHMM model: '1_TssA', '2_PromU', '3_PromD1' and '4_PromD2' belong to promoters; '9_TxReg', '10_TxEnh5", '11_TxEnh3" and '12_TxEnhW' belong to transcribed enhancers; '13_EnhA1', '14_EnhA2' and '15_EnhAF' belong to active enhancers; '16_EnhW1', '17_EnhW2' and '18_EnhAc' belong to weak enhancers. In 15-state chromHMM model: '1_TssA', and '2_TssAFlnk' belong to promoters; '6_EnhG' and '7_Enh' belong to enhancers.

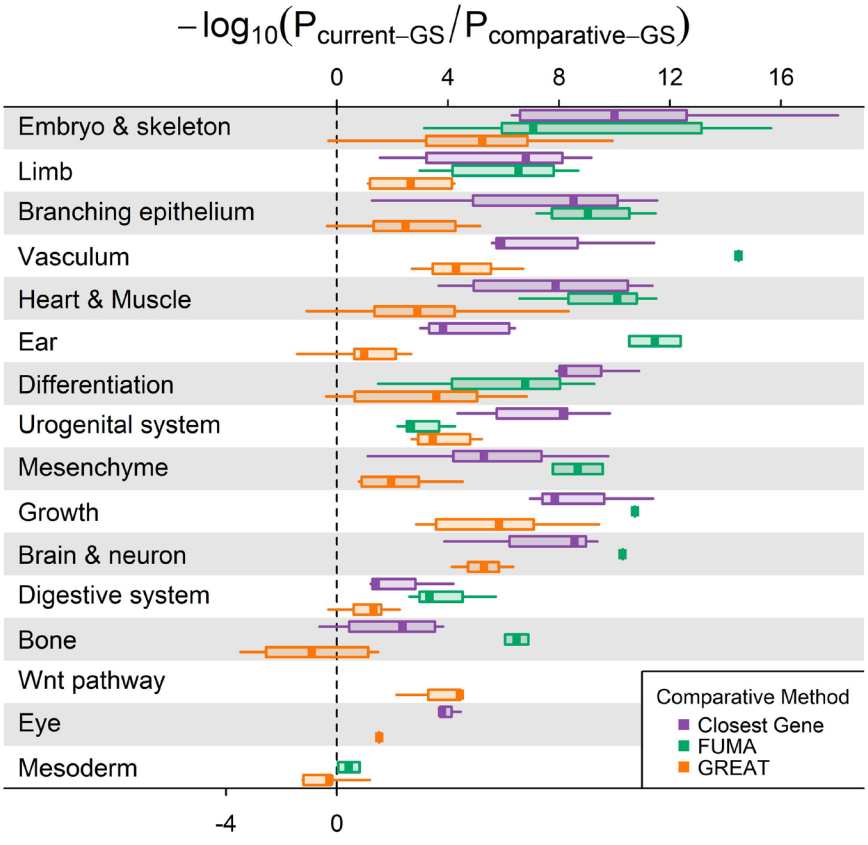
Our facial heritability enrichment analysis, utilizing stratified LDSC[6] across various cell and tissue types, identified 15 cell types with statistically significant enrichment (Bonferroni threshold $p = 4.9\times10^{-4}$, Supplementary Fig. 7 and Supplementary Data 6). The most substantial enrichments ($p < 1\times10^{-9}$) were observed in adult dermal fibroblast primary cells ($p = 3.3\times10^{-11}$), cranial neural crest cells (CNCC, $p = 2.1\times10^{-10}$), mesenchymal stem cell-derived chondrocyte cultured cells ($p = 2.5\times10^{-10}$), and embryonic craniofacial tissue at Carnegie stage 15 (CS15, $p = 2.7\times10^{-10}$). These findings indicate significant overlaps between specific regulatory active regions and face-associated SNPs in these cell and tissue types. Further analysis using 15-state chromHMM on the top four enriched cell and tissue types revealed that a large proportion of our lead SNPs were annotated with enhancer (83.3% of all 253 SNPs, including 79.4% of the 97 novel SNPs) and/or promoter states (37.5% of all 253 SNPs, including 35.1% of the 97 novel SNPs). Notably, these SNPs were not significantly associated with other chromatin states (Supplementary Data 10), underscoring their functional roles in regulatory non-coding elements within cells contributing to facial development. A more detailed 25-state chromHMM analysis specific to CS15 highlighted a predominance annotation of weak enhancers (n = 99), followed by active enhancers (n = 43), aligning with results from a previous facial GWAS studies[8]. This detailed mapping supports the critical role of enhancer elements in influencing facial variation throughout developmental stages.

Supplementary Note 4

We compared our prioritized gene list (365 genes) with three alternative gene lists: the closest gene list (213 genes), the FUMA-selected gene list (718 genes), and the GREAT[9]-selected gene list (362 genes). To evaluate the biological relevance and enrichment of these gene sets, we performed GO enrichment analysis focusing on BP. At a stringent Bonferroni-corrected threshold $p < 0.001$, our gene list identified 177 BPs, while the closest gene list identified 74 BPs, of which 72 (97%) overlapped with the 177 BPs identified by our gene list, the FUMA gene list yielded 66 BPs, all of which were encompassed within our 177 BPs, and the GREAT gene list identified 137 BPs, of which 128 (93%) overlapped with our 177 BPs. The significance of gene enrichment in our method was markedly higher than that observed in the closest gene, FUMA, and GREAT methods (Supplementary Fig. 18), confirming statistical robustness of our gene prioritization approach.

Supplementary Figure 18. Comparative enrichment significance across gene prioritization methods.

This figure shows the -log10-transformed ratio of enriched p-values across 177 identified biological processes classified into 16 categories (same as Fig. 3a), comparing our current candidate gene set with those from three other prioritization methods: Closest Gene (purple), FUMA (green), and GREAT (orange). Positive values indicate higher enrichment significance in our current candidate gene set, while negative values indicate higher significance in the comparative gene sets.



Supplementary Note 5

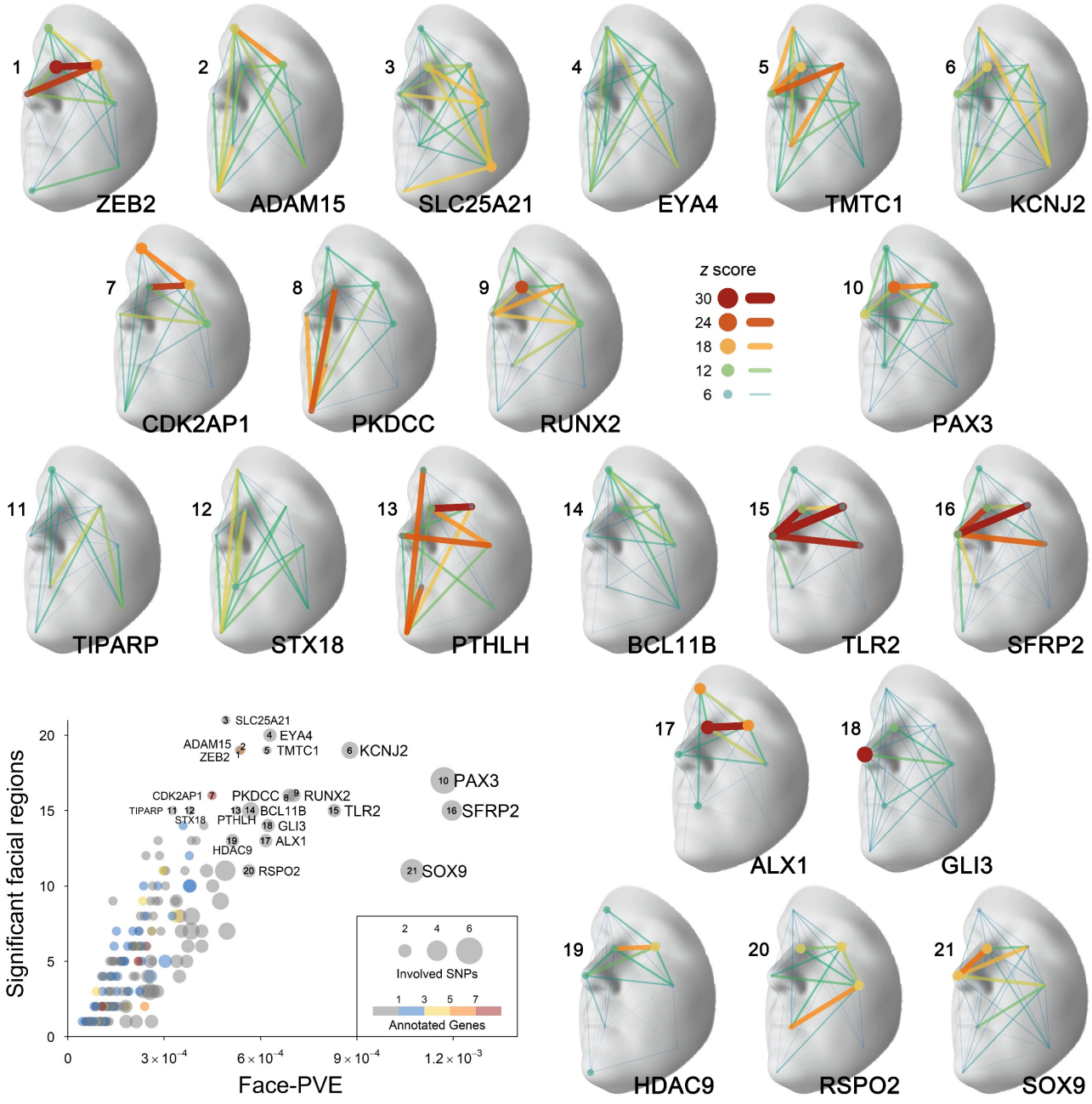
Across the 514 facial traits analysed, the proportion of variance explained (PVE) by individual SNPs ranged between 0% and 1.61%, corroborating the absence of a major gene effect on facial features as consistently observed in previous face GWASs[10]. The combined PVEs from all 253 lead SNPs ranged between 2.25% and 7.89% for different facial traits (Supplementary Data 2), representing a significant increase from the PVEs reported in our earlier study[10], which were based on 31 SNPs and 78 facial traits and ranged from 0.65% to 4.62%. Thus, with an increase of 7.16 times in the number of SNPs identified from C-GWAS based on increased number of facial phenotypes and larger sample sizes, we achieved a 2.25-fold increase in the average genetically explained facial variance in this study. This enhancement is particularly notable given that the current estimates are more conservatively calculated from meta-analysis summary statistics, adjusted for PVE inflation due to the large number of SNPs (see Methods).

To assess the explanatory power of various numbers of SNPs within specific facial regions (regional-PVE, including 1 to 32 traits) and across the entire face (face-PVE), we extended the PVE analysis from single-trait to multiple facial traits. This was done by averaging PVEs across multiple facial traits, which were projected onto an Eigenspace weighted by Eigenvalues (see Methods). The face-PVEs for individual SNPs ranged from 0.004% to 0.064% across SNPs (Supplementary Data 5), with the highest value observed for rs10020603 at 4q31.3 near *SFRP2*. The combined face-PVE for all 253 lead SNPs together was 4.5%, of which more than one third (36%) was contributed by the identified 97 novel SNPs. Notably, the most explanatory loci for face-PVE (>0.1%) included 4q31.3 containing *SFRP2* with 4 SNPs predominantly affecting lower nose and midface, 2q36.1 containing *PAX3* with 6 SNPs influencing upper nose, and 17q24.3 containing *SOX9* with 5 SNPs mainly affecting the entire nose especially lower nose (Supplementary Fig. 19 and Supplementary Data 12).

Different facial regions exhibited varied regional-PVEs (Fig. 4a), with the highest values observed for the upper nose (6.0%) and lower nose (5.6%), and the lowest values for the lower cheek (3.2%) and mouth (3.6%). These variations align with the relative heritability observed between different facial regions. The SNP with the highest regional-PVE, rs2759661 at 1q31.3 near *CRB1*, representing a previously known face-associated locus, had a regional-PVE of 1.45% and predominantly affected chin. Notably, the novel SNPs tended to have more widespread effects across multiple facial regions rather than being confined to a specific area (Supplementary Fig. 11), such as rs8014914 at 14q13.3 containing *SLC25A21* (Supplementary Fig. 19). Our comprehensive mapping of genetic effects on distinct facial dimensions significantly enhanced both the resolution of single-trait-PVE and face-PVE, as well as regional-PVE, thereby highlighting the polygenic and multidimensional nature of facial shape.

Supplementary Figure 19. Facial effect significance map across face candidate gene sets.

In the bottom left, face-PVE of face candidate gene sets (Supplementary Data 12) were plotted against the number of significant regional-PVEs the sets have. The highlighted style of number of SNPs and genes contained in the sets were indicated in the bottom right of the plots. The top sets with large face-PVE (>0.005) or number of significant regional-PVEs (>14) were annotated with the serial number and the most prioritized genes. Their corresponding regional-PVEs significance (z-score) were mapped on face illustrated around. The 3D template facial image in this figure is adapted from White et al.[18] published under an Open Access license (CC BY 4.0), see http://creativecommons.org/licenses/by/4.0/.



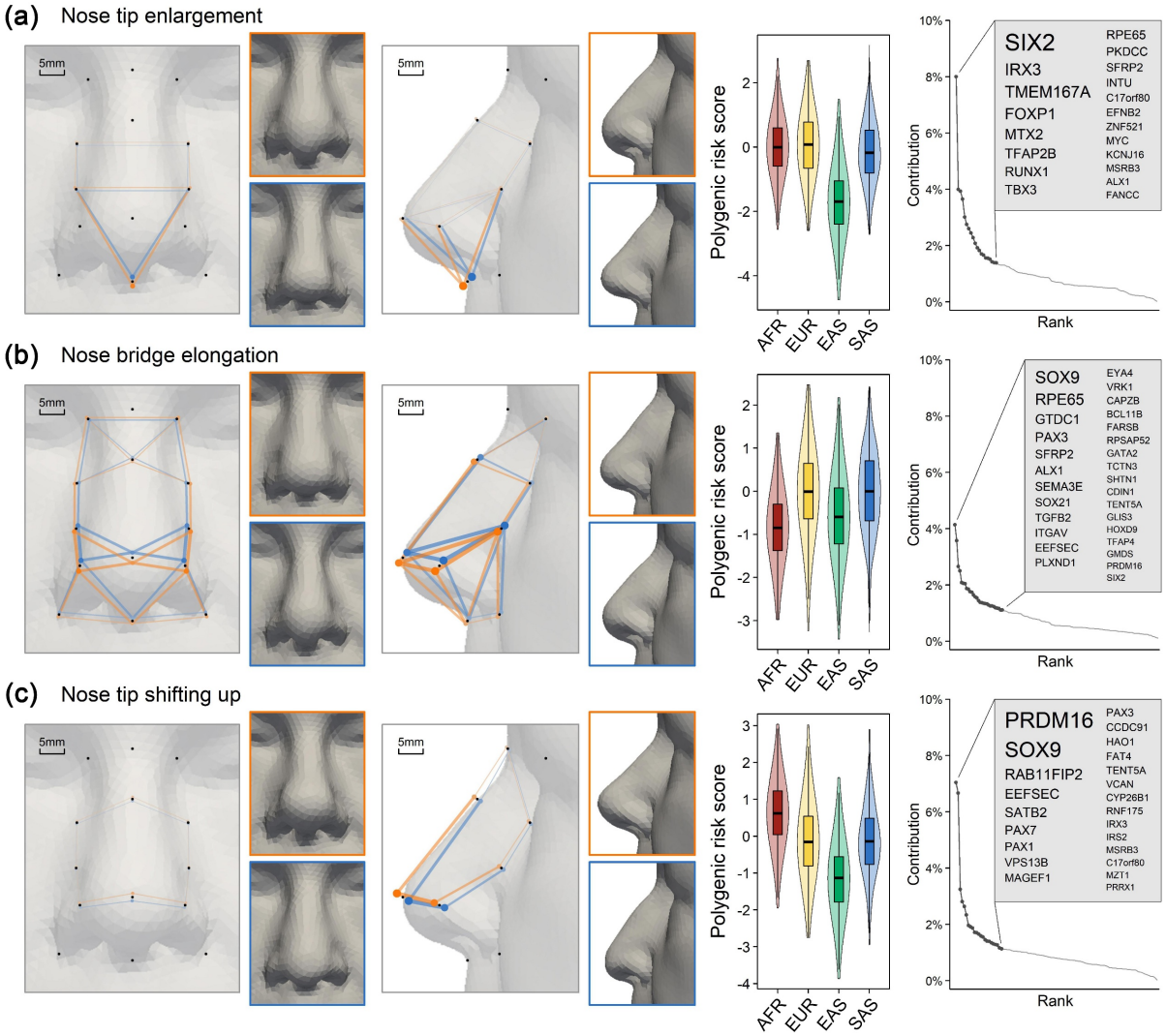
Supplementary Note 6

To map genetic effects to facial shape, the raw 3D facial image data were built into a template-based mesh data set following a registration pipeline[11]. In brief, the pipeline converts each raw 3D facial image into a triangular mesh, comprises a set of triangles that are connected by their common edges or vertices, by using non-rigid registration techniques. Then, we exploited a 3D graph auto-encoder SpiralNet++[12], a type of artificial neural networks to learn compressed representations of the input data in a self-supervised manner, to perform dimensionality reduction on the 3D facial mesh. Specifically, the auto-encoder consists of an encoder and a decoder that can perform feature mapping in a non-linear manner. The encoder compresses the facial mesh into 200 latent representations, while the decoder reconstructs the facial mesh from the representations. After the training of the auto-encoder, each input facial mesh is represented by a latent point in the latent space[13]. The latent space can be considered as a vector subspace with basis vectors $v = (v_1,v_2,..,v_{200})^T$, in which any vector can be represented by $b \circ v$ as a linear combination of 200 latent representations, where $b = (b_1,b_2,..,b_{200})^T$ are weightings of each latent representation. In latent space, most input representation, such as facial PRS, has a vector direction that predominantly captures its variability. We used a recently developed projection-wise disentangling technique[13] to couple a vector direction with an input phenotype by estimating the optimal weightings of b in the a-posteriori manner, aiming for a maximum correlation between the linear combination of latent representations and an input phenotype. Thus, the estimated vector can maximally and accurately capture the variability of the corresponding phenotype. Subsequently, we performed sampling of latent representations along the identified vector direction, and reconstructed a sequence of 3D facial images showing facial deformation related to the input phenotype[13]. The range of sampling was set to be mean ± 3SD of the RS. Finally, we summarized the PRS effect on face from two aspects: 1) we built linear regression models relating PRSs to the 3D coordinates of 44 face landmarks. Variations in extreme risk groups (± 3SD) were illustrated using three times the model effect size; 2) we compare the reconstructed facial image at mean ± 3SD to the average facial images.

Motivated by the higher regional-PVE observed for the nose region compared to all other facial regions, we conducted a polygenic risk scores (PRS) analysis focusing on the nose. This analysis included five distinct nose features summarized from the mapped effects on face of PRS for corresponding five nose traits (see above): nose tip protrusion (landmarks L21-L24 and L23-L26 measurements, Fig. 4c), nose root enlargement (L7-L8 and L8-L9, Fig. 4d), nose tip enlargement (L21-L28 and L23-L28, Supplementary Fig. 20a), nose bridge elongation (L7-L23 and L9-L21, Supplementary Fig. 20b), and the upward shift of the nose tip (L22-L28, Supplementary Fig. 20c).

Supplementary Figure 20. PRS analysis for three nose traits.

PRS analysis and its effects on the nose are illustrated for: (**a**) L7-L8 and L8-L9, (**b**) L7-L23 and L9-L21, and (**c**) L22-L28. Each figure is divided into three sections in the same format with Fig 4c-d: PRS effects on the nose (left, enlarged nasal view), population profiling (middle, violin plots), and gene contributions (right). In the nasal views, landmarks significantly associated with standardized PRS are marked with orange dots (3 times positive effect offset) or blue dots (3 times negative effect offset). Connecting lines visually represent nasal variation, with scaled distances shown in the top left. Nasal images reconstructed at mean PRS ±3SD are framed in orange (+3SD) and blue (-3SD) and generated using a 3D graph auto-encoder (see Methods). The overall effect of increasing PRS is described in text at the top left. The violin plots show standardized PRS results for four major populations from the 1000 Genomes Project (African, AFR, N = 504; European, EUR, N = 504; East Asian, EAS, N = 503; South Asian, SAS, N = 489), with the band indicating the median, the box representing the first and third quartiles, and whiskers extending 1.5 times the interquartile range. Gene contribution plots on the right list the top regions accounting for 50% of PRS variance. The 3D template facial image in this figure was generated from the average covariate-adjusted facial shape of all RS participants (*N* = 4,242).



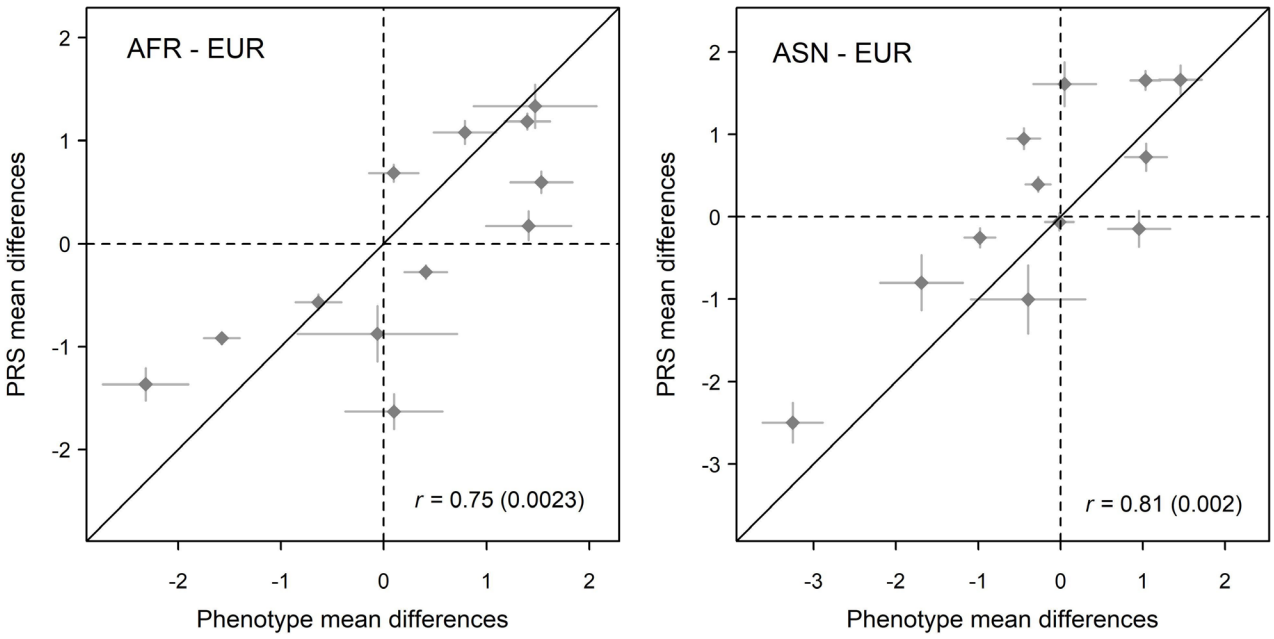
Supplementary Note 7

To evaluate whether our PRS capture facial shape differences across modern human populations, we conducted a comparative analysis using 3D facial data. We focused on the study by Simmons-Ehrhardt et al.[14], which analysed 3D bone and soft tissue parameters from computed tomography scans of 388 adults from AFR, Asian (ASN), EUR, and Hispanic populations. This study was selected because it used 3D data, as we used in our study, and included many facial traits that overlapped with those in our dataset. We identified 42 facial phenotypes common to both studies. For each trait, we calculated the standardized mean differences between non-European populations (AFR, and ASN, excluding Hispanic) and the EUR, using the EUR as the reference. This resulted in 42×2 phenotype mean differences representing the facial shape variations among populations. We constructed PRS for the same 42 traits using 382 SNPs and the same approach described in Methods "Genetic modelling of facial traits". These PRS were applied to individuals from corresponding populations in the 1000 Genomes Project dataset. We calculated the standardized mean differences in PRS between non-European and European populations, yielding 42×2 PRS mean differences. Note that because the study of Simmons-Ehrhardt et al. does not distinguish EAS and SAS, we merged in our study EAS and SAS as ASN samples.

Since many of the 42 traits are highly correlated, we transformed the PRS and phenotype mean differences into an eigenvector space using the same approach described in Methods "Proportion of Facial Variance Genetically Explained". We retained the top 12 transformed mean differences, corresponding to the accumulated Eigen value over 95% of total 42. We assessed the correlation between the 12 transformed PRS mean differences and the phenotype mean differences. A higher correlation would indicate that the PRS effectively reflects the observed facial differences among populations. To determine the significance of this correlation, we generated a null distribution through 10,000 replicates. In each replicate, we constructed PRS using 382 randomly selected SNPs that were matched for MAF. We transformed mean differences and calculated the correlations for these replicates to establish a null distribution of correlations expected by chance. We compared our observed correlation to the null distribution to calculate a p-value. This allowed us to assess whether the PRS captured population-level facial differences beyond what would be expected randomly. The analysis revealed significant correlations between mean differences in facial traits and mean differences in PRS between EUR and AFR (Pearson's $r = 0.75$, $p = 0.0023$), as well as between EUR and ASN ($r = 0.81$, $p = 0.002$) populations (Supplementary Fig. 21).

Supplementary Figure 21. Correlation between phenotype and PRS mean differences in Europeans vs. non-Europeans.

The scatterplot depicts the relationship between standardized mean differences in phenotypes (x-axis) and PRS (y-axis) for top 12 eigenvector-transformed facial traits between AFR and EUR (left) and between Asian (ASN) and EUR (right). Error bars represent 95% confidence intervals. Pearson's correlation coefficient is displayed in the bottom-right corner, where its significance level was derived from a null distribution through 10,000 replicates.



Supplementary Note 8

To address potential winner's curse bias in our PRS analysis due to overlapping samples used for training and testing, we conducted additional simulation analyses explicitly quantifying the extent of the winner's curse bias. We generated synthetic genotype and phenotype data with a known genetic architecture that closely mirrors our study population. Specifically, genotypes were simulated for a discovery cohort (N = 11,662 individuals, representing all cohorts) across one million SNPs, assigning normally distributed effect sizes to causal variants to achieve a targeted heritability of 0.23 on average for facial traits and ensuring approximately 10% variance explained by the PRS. Phenotypes were then derived from these genotypes using the true effect sizes and incorporating random environmental variation.

After conducting GWAS on the synthetic data, we simulated a power boost of C-GWAS to closely match conditions in our original dataset, i.e., we applied a multiplicative factor of 2.77 to the statistics of true causal SNPs using the increase in the mean χ^2 statistic method described in Turley et al.[15]. PRSs were subsequently calculated using significant SNPs ($p < 5\times10^{-8}$) under three distinct scenarios: (1) directly applying GWAS effect estimates to overlapping samples analogous to RS (N = 4,242), (2) employing a cross-validation approach within RS, and (3) applying effect estimates derived from RS to an independent overlapping set analogous to TwinsUK (N = 1,080).

As shown in the results of Supplementary Table 5, our simulations demonstrated clear winner's curse bias under scenario (1), where PRS performance was inflated relative to the true underlying genetic signal. Importantly, however, scenarios (2) and (3) significantly mitigated this bias, providing more conservative estimates of PRS predictive performance. Consequently, we interpret the observed decrease in AUC from RS to TwinsUK as predominantly reflecting genuine genetic and phenotypic differences between these cohorts rather than solely attributable to winner's curse bias. One systematic difference between both cohorts is that TwinsUK reflects a female-only cohort, while RS includes males and females, which could have an impact on the differences in PRS results.

Supplementary Table 5. Difference between explanatory power of PRS using estimates and true effects.

The explanatory power of PRS is presented as the percentage of the squared correlation between PRS and the phenotype. The mean and standard deviation estimates are derived from 200 simulations.

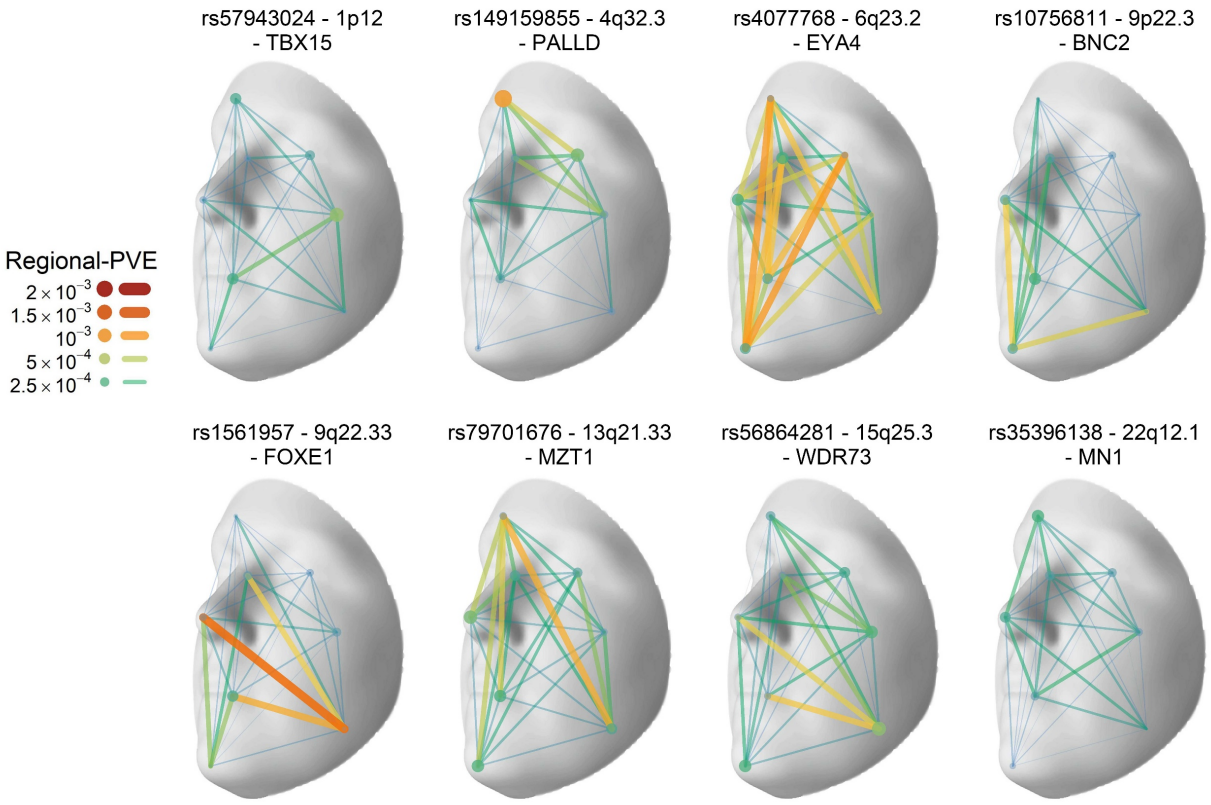
Scenarios	Explained variance (%)			
	PRS using estimates effects		PRS using true effects	
	Mean	SD	Mean	SD
Scenario 1	11.02	1.1	9.86	1.05
Scenario 2	7.15	1	9.84	1.05
Scenario 3	7.26	1.37	9.94	1.56

Supplementary Note 9

For all facial SNPs located within introgressed segments, we assessed whether they are truly archaic SNPs, and incorporated this information into Supplementary Data 17. Specifically, we identified archaic SNPs within introgressed regions based on criteria established by Zeberg et al.[16], and then overlapped these with our 253 lead SNPs and their high-LD counterparts ($r^2 > 0.6$). Lead SNPs identified through this overlap were classified as high-confidence introgressed lead SNPs. This approach yielded only eight high-confidence introgressed lead SNPs among the 92 lead SNPs within introgressed segments in EUR or EAS populations. For instance, rs57943024 on chromosome 1p12 is within a Denisovan introgression segment (99% confidence) and is in strong LD ($r^2 = 0.86$) with rs3790553, a previously reported Denisovan-introgressed facial SNP associated with lip thickness in contemporary Latin Americans[17]. Eight high-confidence introgressed lead SNPs exhibited small overall effects on facial shape, primarily influencing the forehead, cheeks, and chin (Supplementary Fig. 22). However, we acknowledge that the small number of identified high-confidence introgressed facial SNPs limits our ability to draw statistically robust conclusions.

Supplementary Figure 22. Facial effect of eight high-confidence introgressed lead SNPs.

Facial effect of selected SNPs was illustrated on face using their regional-PVE across 8 facial regions. Annotation of SNPs include rsID, ctyoband, and the most prioritized gene. The 3D template facial image in this figure is adapted from White et al.[18] published under an Open Access license (CC BY 4.0), see http://creativecommons.org/licenses/by/4.0/.



Supplementary Note 10

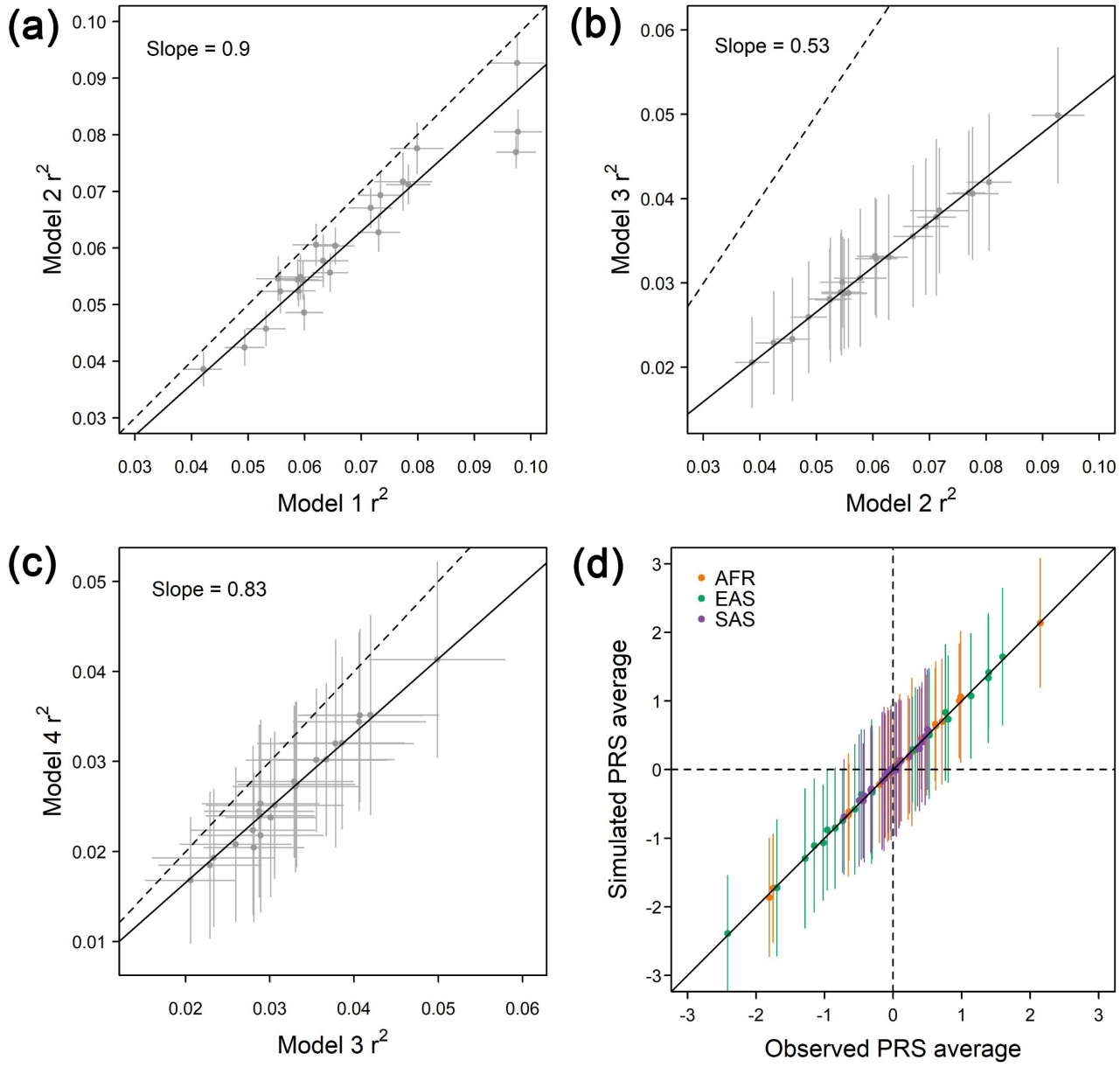
We performed simulations to evaluate how LD and the use of pseudohaploid data, as expected from sequencing a DNA sample of archaic humans, impact on the accuracy of PRS for predicting the 23 facial traits using the RS dataset across four distinct models. In Model 1, we conducted cross-validation with 382 SNPs (same SNPs set described in Methods "Genetic modelling of facial traits"), fitting the PRS to the 23 facial traits using diploid data. Model 2 involved cross-validating 357 SNPs (same SNPs set described in Methods "Archaic introgressed regions and facial PRS profiling"). For Model 3, we maintained the same set of 357 SNPs, but utilized pseudohaploid data for validation by randomly converting heterozygous genotypes into homozygous genotypes of minor or major allele with 50% probability each. Model 4 extended this approach by employing pseudohaploid data for both the training and validation data sets, allowing for a direct comparison with Model 3, which retained diploid data for training. To assess and compare the performance across these models, scatterplots were generated to visualize the fit of the PRS, measured by the coefficient of determination (r^2), with the no-intercept regression slopes of these plots indicating the relative decline in model performance. Finally, to assess the impact of small sample size, e.g., N = 8 for Neanderthals and pseudohaploid data on population-level PRS averages, we applied the Model 1 to each of three populations (AFR, EAS, SAS) from the 1000 Genomes Project for deriving 23 PRS averages, and compared them with those derived from 8 randomly selected samples from each of AFR, EAS, and SAS using fitting approach of Model 3, which employs simulated pseudohaploid data. All simulations and cross validation were repeated for 100 times to derive the mean and 95% confidence intervals.

As a result, Model 2 showed a slight performance decline compared to Model 1 (slope = 0.90, Supplementary Fig. 23a) due to noise from SNP substitution with LD proxies. A more substantial drop occurred in Model 3 (slope = 0.53, Supplementary Fig. 23b) with pseudohaploid data validation, highlighting the loss of heterozygous information. Model 4 showed a moderate decline (slope = 0.83, Supplementary Fig. 23c) when pseudohaploid data were used for both training and validation, reflecting cumulative information loss. These results emphasize the challenges of applying PRS from diploid datasets to pseudohaploid data when dealing with DNA data from archaic humans.

To evaluate the effects of pseudohaploid data and small sample sizes, as confronted with archaic genome data, on population PRS averages, we applied PRS derived from EUR (Model 1) to randomly selected subsets (N = 8, as Neanderthal sample size was 8) drawn from AFR, EAS, and SAS using pseudohaploid-based fitting approach of Model 3 (Supplementary Fig. 23d). Simulated PRS averages were consistent with observed values across populations, with negligible bias but wide 95% confidence intervals. This suggests that pseudohaploid data and small sample size do not bias PRS averages but increase its uncertainty, e.g., absolute PRS < 1 in our case.

Supplementary Figure 23. Impact of LD and pseudohaploid data on the accuracy of PRS for predicting 23 facial traits.

For panels **a**, **b**, and **c**, the grey dots with error bars represent the average r^2 and their 95% confidence intervals across 23 facial traits for two models under comparison. The solid line depicts the no-intercept regression line of the average r^2, with its slope noted in the top left. The dashed line represents a slope of 1. (**a**) Comparison of PRS fit between Models 1 and 2 (diploid data with 382 vs. 357 SNPs). (**b**) Comparison of PRS fit between Models 2 and 3 (diploid vs. pseudohaploid validation). (**c**) Comparison of PRS fit between Models 3 and 4 (pseudohaploid validation vs. training and validation). (**d**) Comparison of observed and simulated PRS averages across populations using small sample size and fitting approach of Model 3 for pseudohaploid data.

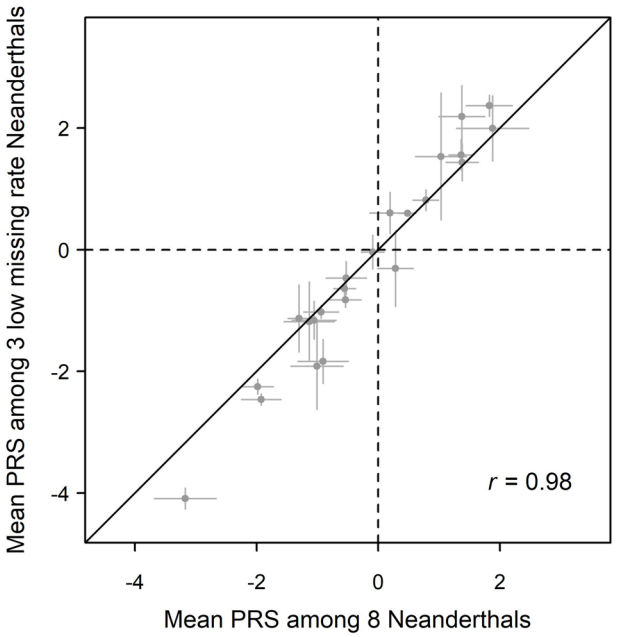


Supplementary Note 11

We analysed the missing data rate of the 357 SNPs in the DNA data from the 10 archaic human individuals (Supplementary Table 3). High-quality, low-missing rate samples (<10%) included three Neanderthals and one Denisovan. For all missing values, we imputed genotypes by replacing them with the mean dosage across all ancient individuals (N = 6,094) from the AADR database. To assess the impact of imputation in our PRS analysis, we calculated the mean PRS for 23 facial traits separately for all 8 Neanderthals and the 3 low-missing rate Neanderthals. A scatterplot was used to compare the PRS means between these two groups, and the correlation was assessed using Pearson's r (Supplementary Fig. 24). The comparison of PRS between all 8 Neanderthals and the 3 low-missing rate Neanderthals revealed a strong correlation (Pearson's r = 0.98), supporting the reliability of our imputation approach in reconstructing the PRS for archaic human individuals in the presence of missing, but imputed, SNP data.

Supplementary Figure 24. Comparison of PRS between full and high-coverage Neanderthals after genotype imputation.

The scatterplot shows the relationship between the mean PRS for 23 facial traits among 8 Neanderthals (x-axis) and the subset of 3 low-missing rate Neanderthals with <10% missing SNPs (y-axis). Error bars represent 95% confidence intervals of mean PRS.



Supplementary References

1. Zhang M*, et al.* Genetic variants underlying differences in facial morphology in East Asian and European populations. *Nature genetics* **54**, 403-411 (2022).
2. Galwey NW. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. *Genetic epidemiology* **33**, 559-568 (2009).
3. Prescott SL*, et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68-83 (2015).
4. Wilderman A, VanOudenhove J, Kron J, Noonan JP, Cotney J. High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep* **23**, 1581-1597 (2018).
5. Roadmap Epigenomics C*, et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
6. Finucane HK*, et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228-1235 (2015).
7. Li Y*, et al.* Combined genome-wide association study of 136 quantitative ear morphology traits in multiple populations reveal 8 novel loci. *PLoS genetics* **19**, e1010786 (2023).
8. White JD*, et al.* Insights into the genetic architecture of the human face. *Nature genetics* **53**, 45-53 (2021).
9. McLean CY*, et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495-501 (2010).
10. Xiong Z*, et al.* Novel genetic loci affecting facial shape variation in humans. *Elife* **8**, (2019).
11. Liu X*, et al.* Association between prenatal alcohol exposure and children's facial shape: a prospective population-based cohort study. *Hum Reprod* **38**, 961-972 (2023).
12. Gong S, Chen L, Bronstein M, Zafeiriou S. SpiralNet++: A Fast and Highly Efficient Mesh Convolution Operator. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* (ed^(eds) (2019).
13. Liu X, Li B, Bron EE, Niessen WJ, Wolvius EB, Roshchupkin GV. Projection-Wise Disentangling for Fair and Interpretable Representation Learning: Application to 3D Facial Shape Analysis. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (ed^(eds de Bruijne M*, et al.*). Springer International Publishing (2021).
14. Simmons-Ehrhardt TL, Monson KL, Flint T, Saunders CP. Quantitative accuracy and 3D biometric matching of 388 statistically estimated facial approximations of live subjects. *Forensic Imaging* **21**, 200377 (2020).
15. Turley P*, et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics* **50**, 229-237 (2018).
16. Zeberg H*, et al.* A Neanderthal Sodium Channel Increases Pain Sensitivity in Present-Day Humans. *Current biology : CB* **30**, 3465-3469 e3464 (2020).
17. Bonfante B*, et al.* A GWAS in Latin Americans identifies novel face shape loci, implicating VPS13B and a Denisovan introgressed region in facial variation. *Sci Adv* **7**, (2021).
18. White JD*, et al.* MeshMonk: Open-source large-scale intensive 3D phenotyping. *Scientific reports* **9**, 6085 (2019).