PNAS
www.pnas.org

Supplementary information for

MHC heterozygosity is not a condition-dependent male ornament and does not reflect MHC heterozygosity

Authors: Arslan A. Zaidi[1], Julie D. White[1], Brooke C. Mattern, Corey R. Liebowitz, David A. Puts, Peter Claes, Mark D. Shriver
[1]These authors contributed equally

Corresponding authors:
Arslan A. Zaidi (saz5078@psu.edu)
312 Wartik Building
Pennsylvania State University
University Park
Pennsylvania-16802
United States of America

Mark D. Shriver (mds17@psu.edu)
511 Carpenter Building
Pennsylvania State University
University Park
Pennsylvania-16802
United States of America

This PDF includes:
Figs. S1-S14
Table S1

Other supplementary materials for this manuscript include the following:
Dataset S1
Dataset S2

www.pnas.org/cgi/doi/10.1073/pnas.1808659116

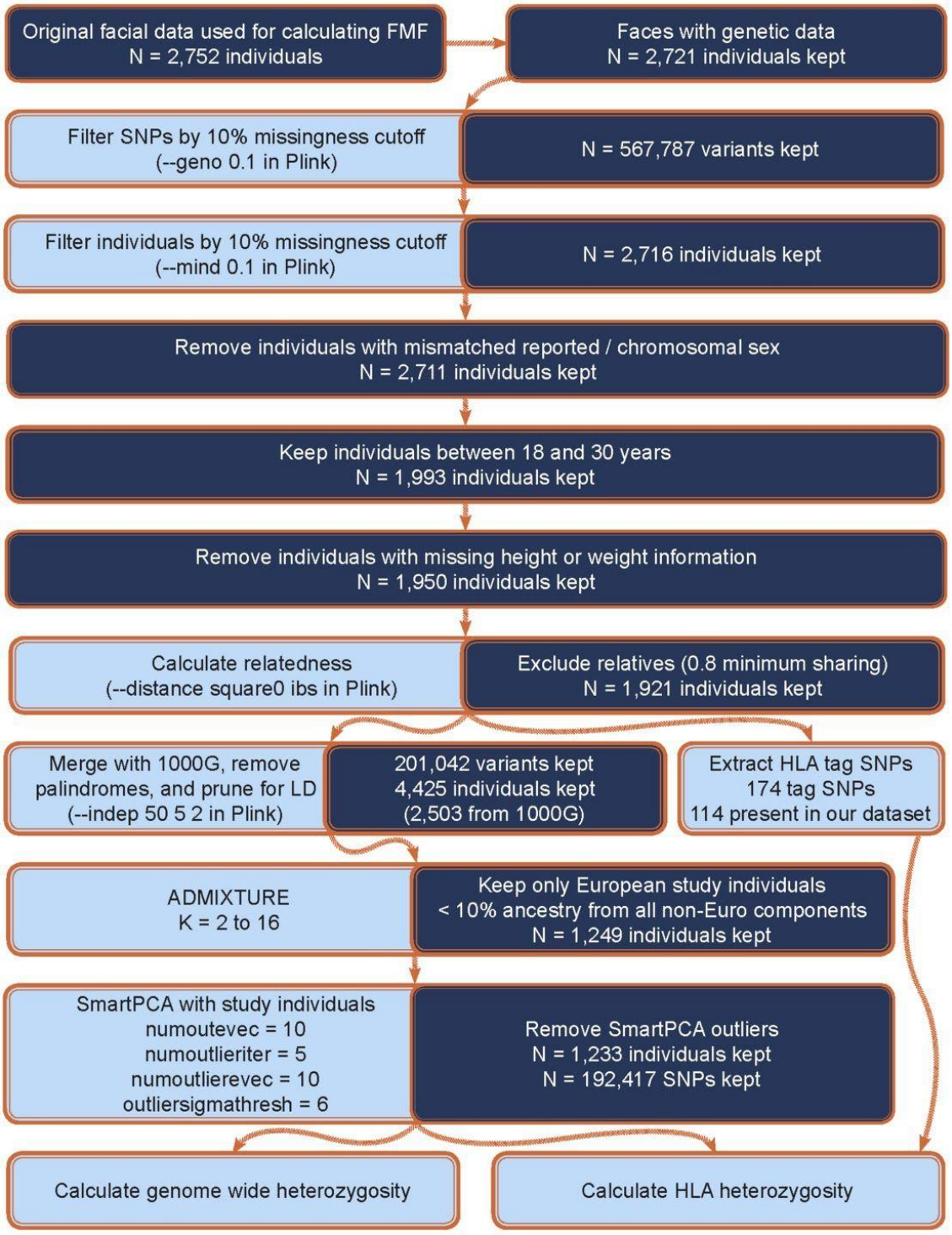

Figure S1: Data curation diagram. Visual depiction of the filtering process used to create the dataset used for analysis.

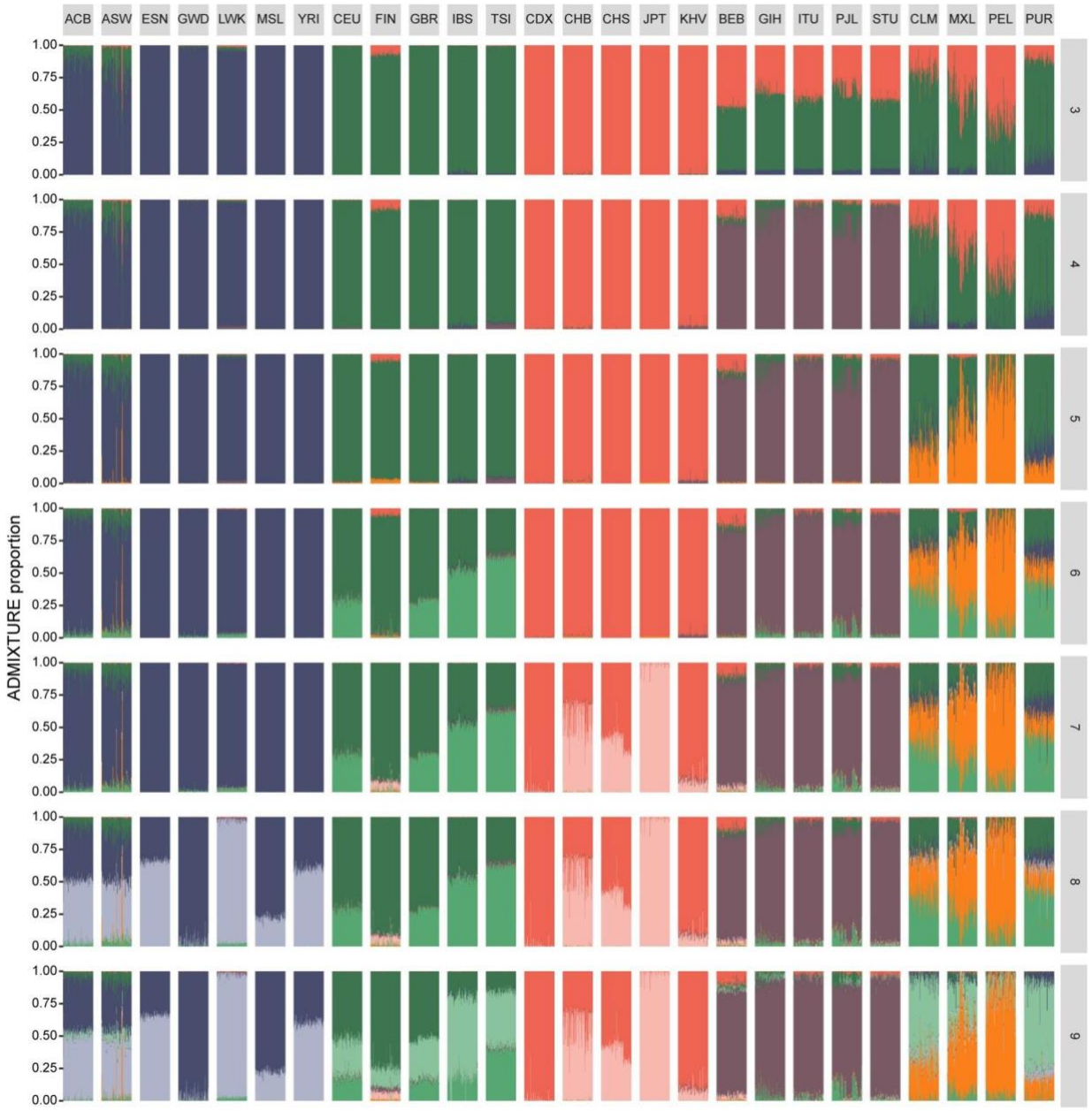

Figure S2: ADMIXTURE plot for 1000 Genomes samples for K from 3 to 9. Scale on the left Y-axis shows the ADMIXTURE fraction from each cluster (represented by different colors). Each vertical bar represents an individual's ancestry fraction and the relative heights of a color in each bar represents the ancestry fraction from that cluster. Individuals are grouped column-wise into known population and continental groupings as shown on top (see Table S1 for description of abbreviations). Results are further grouped row-wise for different values of K indicated on the right.

Table S1: 1000 Genomes Populations. Population descriptions for 1000 Genomes populations used for ancestry comparisons (1). Populations appear in order of appearance of column-wise groupings in Figure S2.

Population Code	Population Description	Super Population Code
ACB	African Caribbeans in Barbados	African
ASW	Americans of African Ancestry in SW USA	African
ESN	Esan in Nigeria	African
GWD	Gambian in Western Divisions in the Gambia	African
LWK	Luhya in Webuye, Kenya	African
MSL	Mende in Sierra Leone	African
YRI	Yoruba in Ibadan, Nigeria	African
CEU	Utah Residents (CEPH) with Northern and Western European Ancestry	European
FIN	Finnish in Finland	European
GBR	British in England and Scotland	European
IBS	Iberian Population in Spain	European
TSI	Toscani in Italia	European
CDX	Chinese Dai in Xishuangbanna, China	East Asian
CHB	Han Chinese in Beijing, China	East Asian
CHS	Southern Han Chinese	East Asian
JPT	Japanese in Tokyo, Japan	East Asian
KHV	Kinh in Ho Chi Minh City, Vietnam	East Asian
BEB	Bengali from Bangladesh	South Asian
GIH	Gujarati Indian from Houston, Texas	South Asian
ITU	Indian Telugu from the UK	South Asian
PJL	Punjabi from Lahore, Pakistan	South Asian
STU	Sri Lankan Tamil from the UK	South Asian
CLM	Colombians from Medellin, Colombia	Admixed American
MXL	Mexican Ancestry from Los Angeles USA	Admixed American
PEL	Peruvians from Lima, Peru	Admixed American
PUR	Puerto Ricans from Puerto Rico	Admixed American

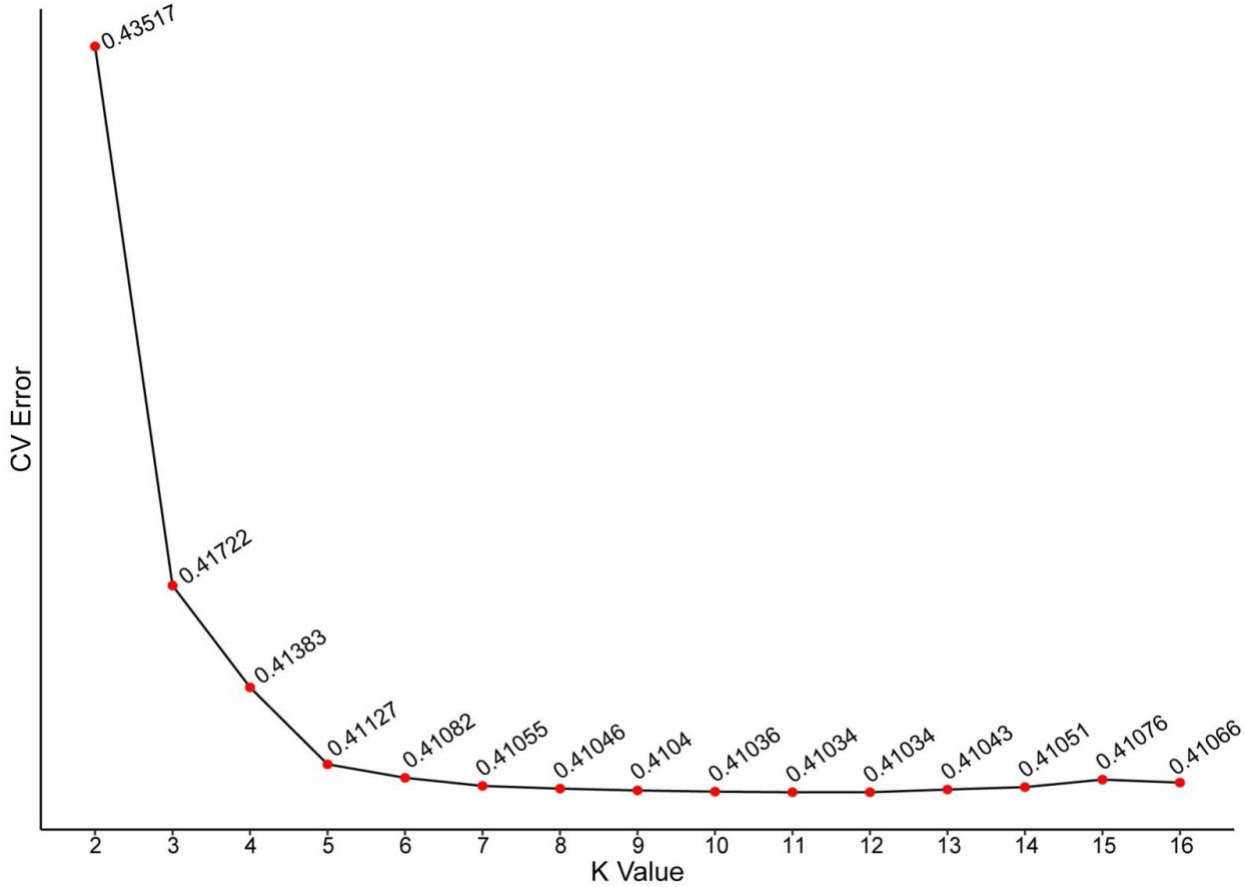

Figure S3: Cross validation error for ADMIXTURE run using K from 2 to 16. A K value of 11 and 12 gave the absolute lowest CV error. However, an elbow in the curve can be seen at K = 6. Therefore, K=6 was chosen to preliminarily select individuals with European-derived ancestry, as it facilitated easy separation of the five 1000 Genomes (1) super-populations (Fig. S1) and had a low CV error.

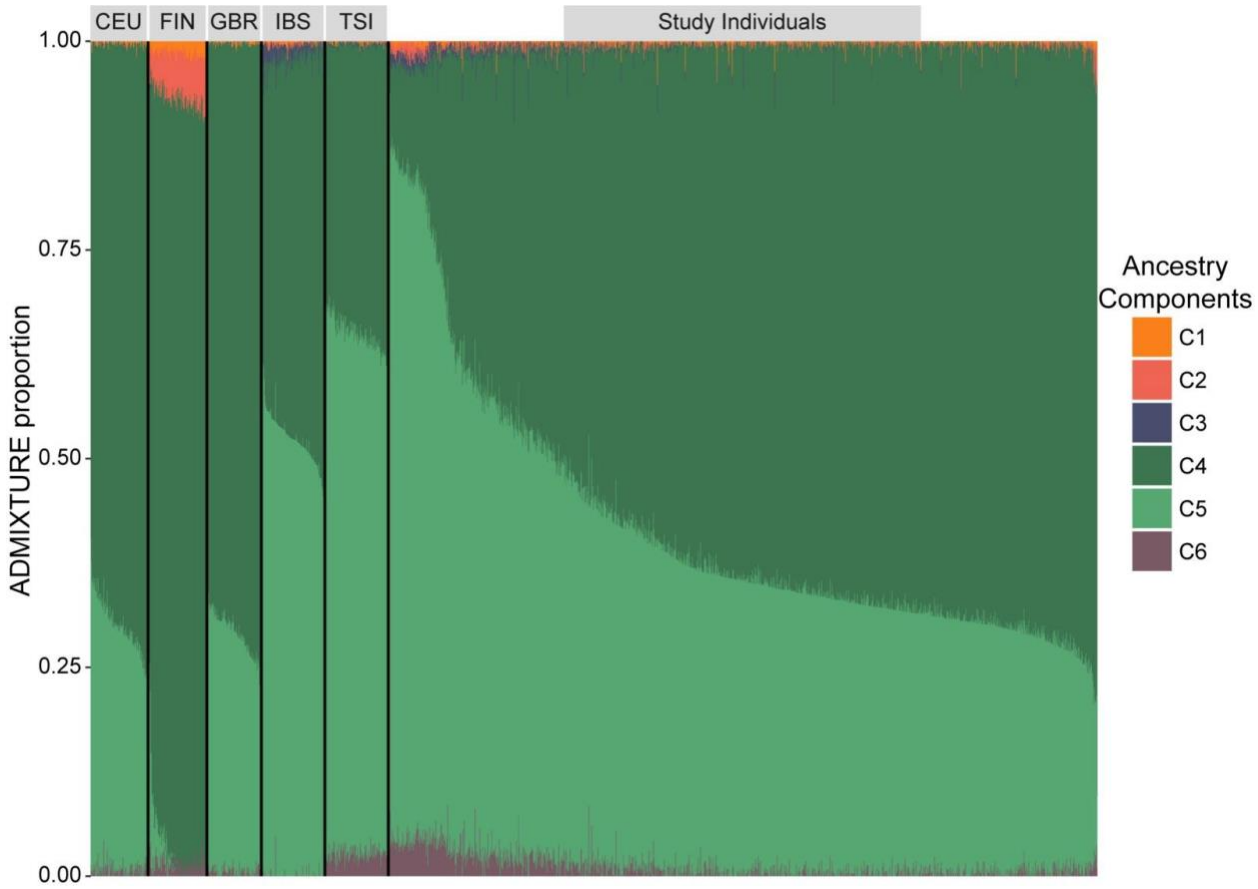

Figure S4: ADMIXTURE (K=6) results of individuals of European ancestry. The 1000 Genomes European individuals (CEU, FIN, GBR, IBS, and TSI) (1) are used here as reference to illustrate that the individuals used for analyses in this paper are primarily of European ancestry.

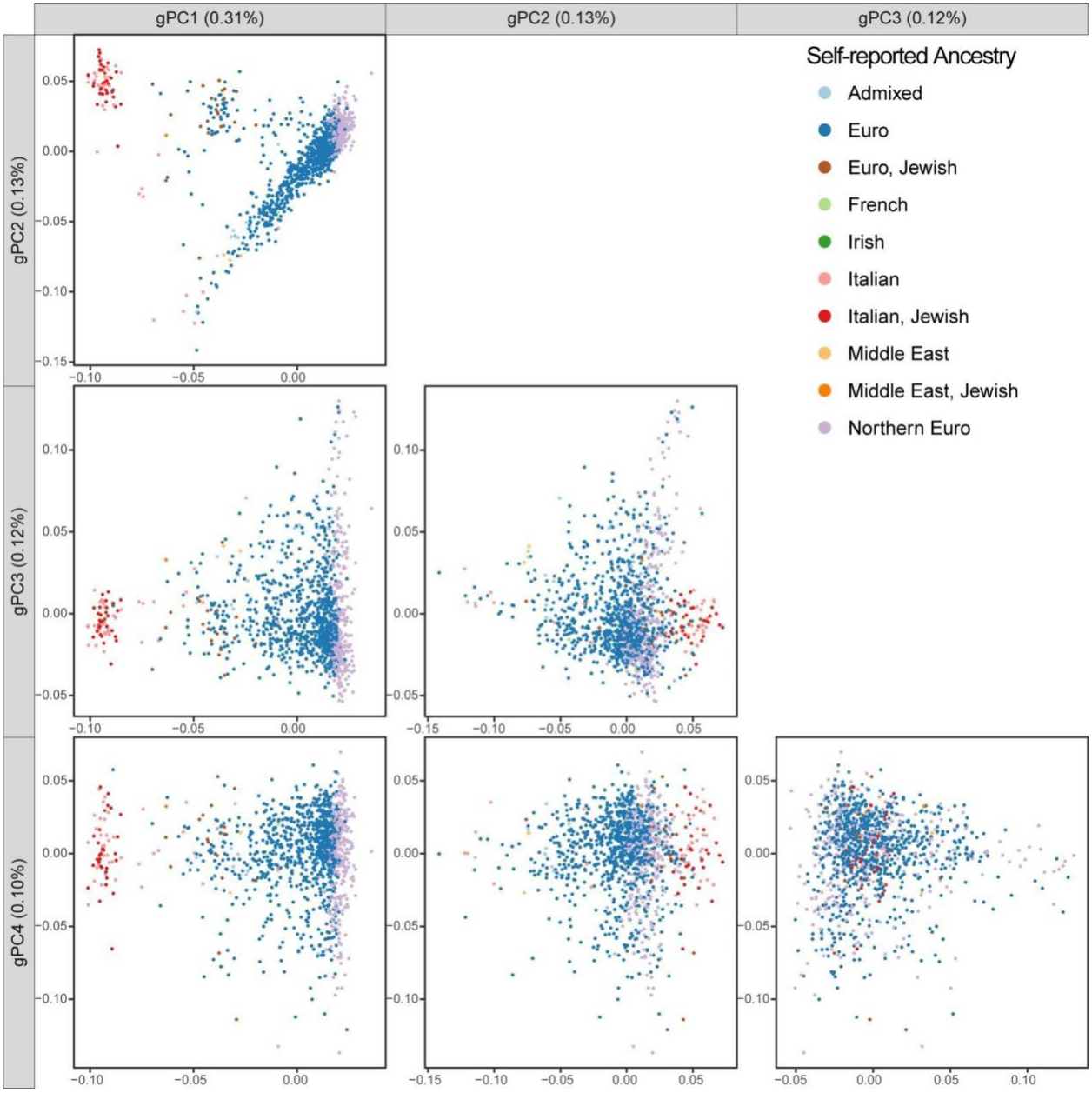

Figure S5: PCA of individuals of European ancestry. PCA on the study individuals was performed using Eigensoft and outliers were removed. Points are colored based on self-reported ancestry.

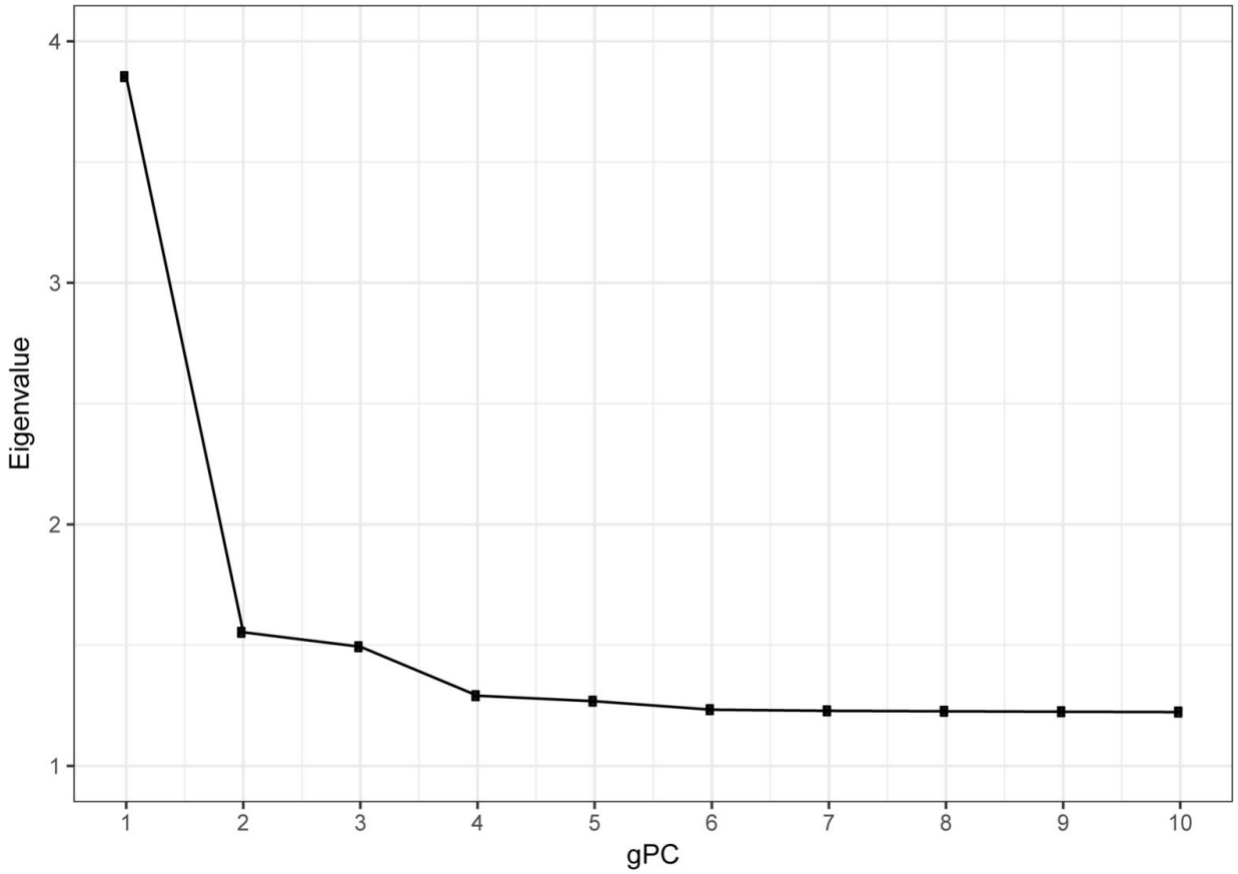

Figure S6: Screeplot showing eigenvalues from a principal components analysis (PCA) carried out on genotype data. There is little change in the eigenvalue after gPC3, suggesting most of the population structure is captured by gPCs1-3.

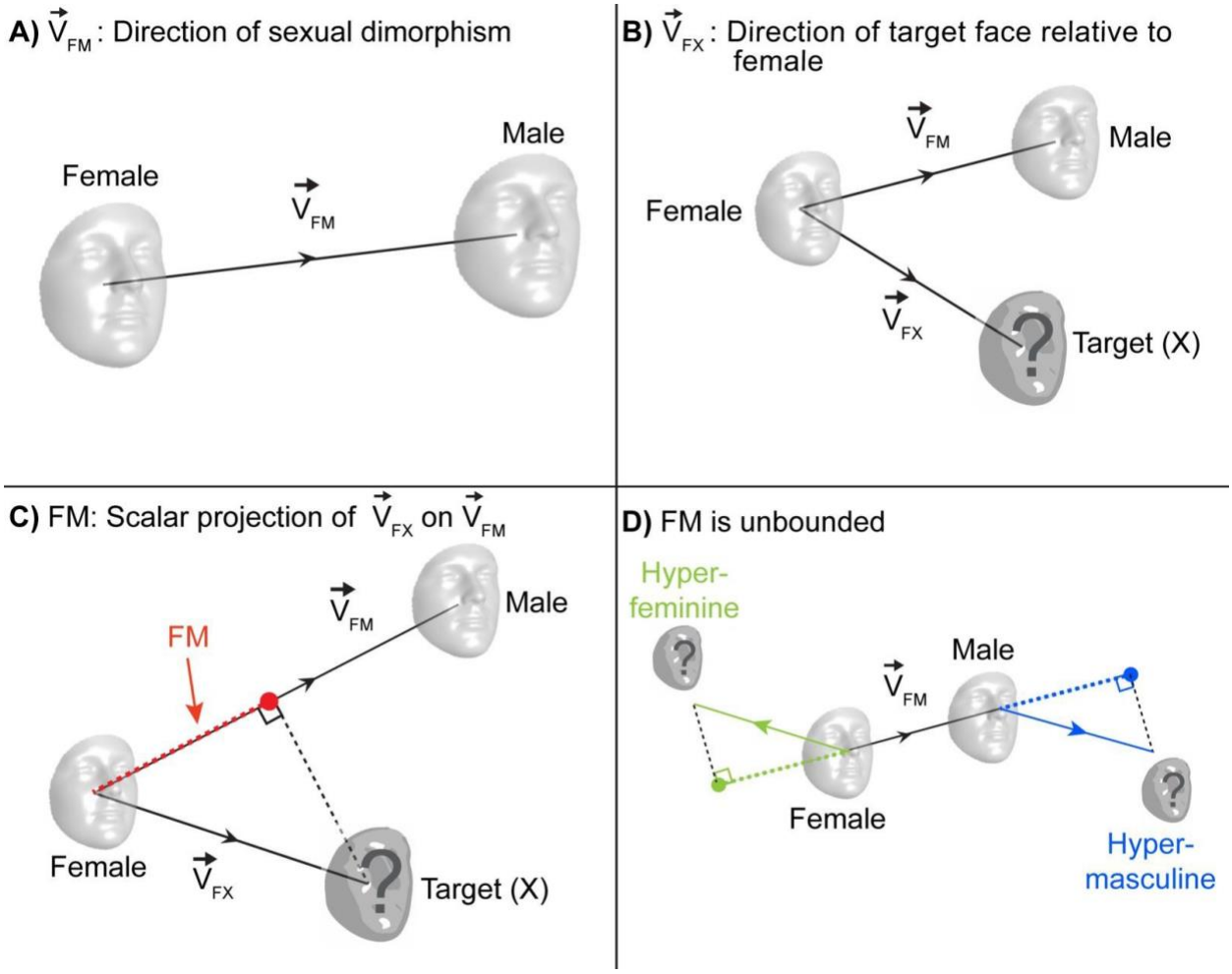

Fig. S7: Calculation of facial masculinity. A) The direction of sexual dimorphism $\overrightarrow{V_{FM}}$ is established by subtracting the female consensus face from the male consensus face. B) $\overrightarrow{V_{FX}}$, the direction of a target face (X) relative to the female face, is calculated. C) The scalar projection of $\overrightarrow{V_{FX}}$ along $\overrightarrow{V_{FM}}$ is the facial masculinity (FM). D) FM can be positive or negative and can be more masculine or more feminine than the consensus male and female face, respectively.

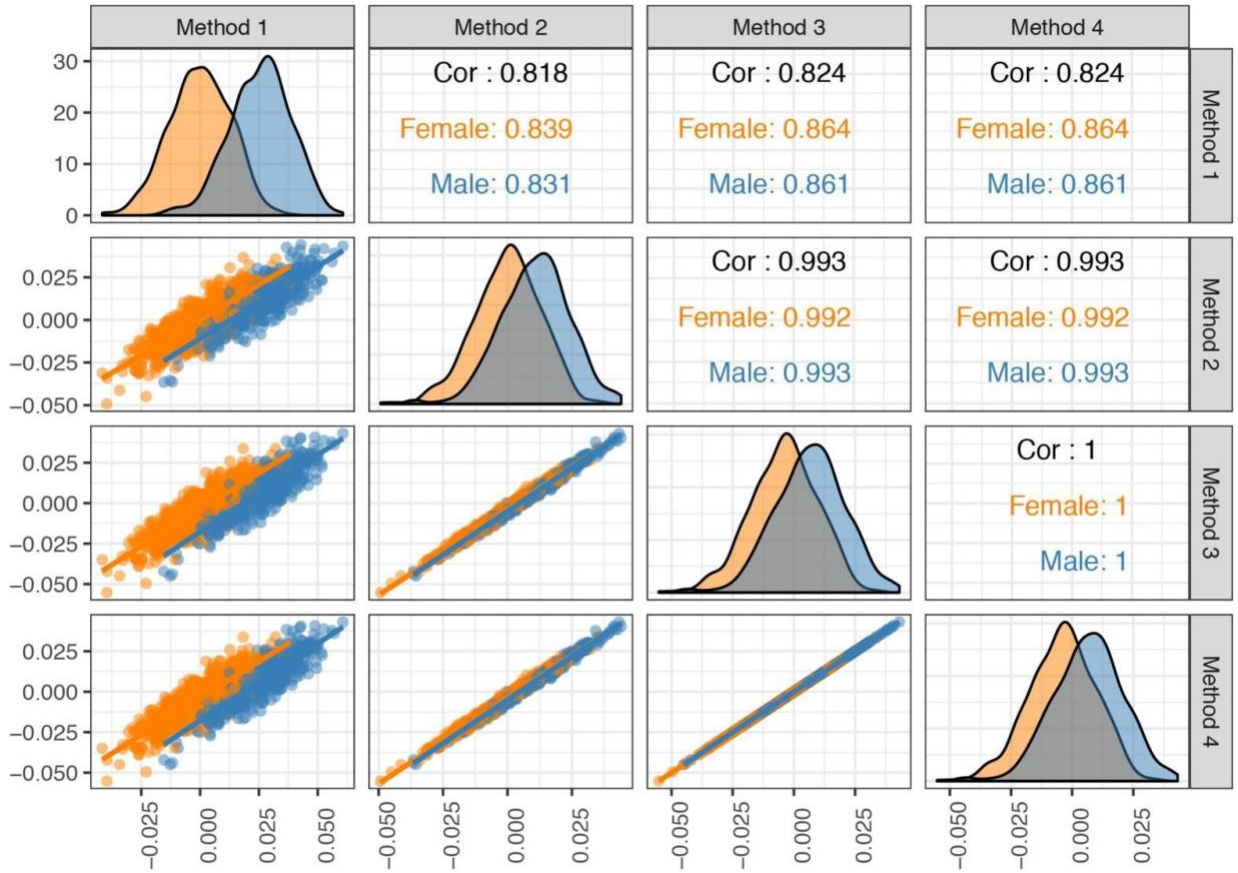

Figure S8: Comparison of overall facial masculinity ($FM_{overall}$) calculated using different methods. In Method 1 we calculated FM_{QL} without correction for allometry as described in the main text and averaged across QLs to get $FM_{overall}$. In Method 2, we obtained residuals from a regression of the original shape coordinates on height, and used these to construct the consensus faces as well as to estimate FM_{QL} for each individual face. In Method 3, we residualize FM_{QL}, estimated using Method 1, for height before averaging to obtain $FM_{overall}$. This approach is equivalent to using height as a covariate in a regression model where the response is FM_{QL}. In Method 4, we estimated $FM_{overall}$using Method 1, and then residualize on height, which is equivalent to using height as a covariate in a regression model where the response is $FM_{overall}$. With the exception of Method 1, in which no correction for height was carried out, all other methods yield extremely similar values of $FM_{overall}$. Therefore, all three approaches provide equivalent corrections for allometry.

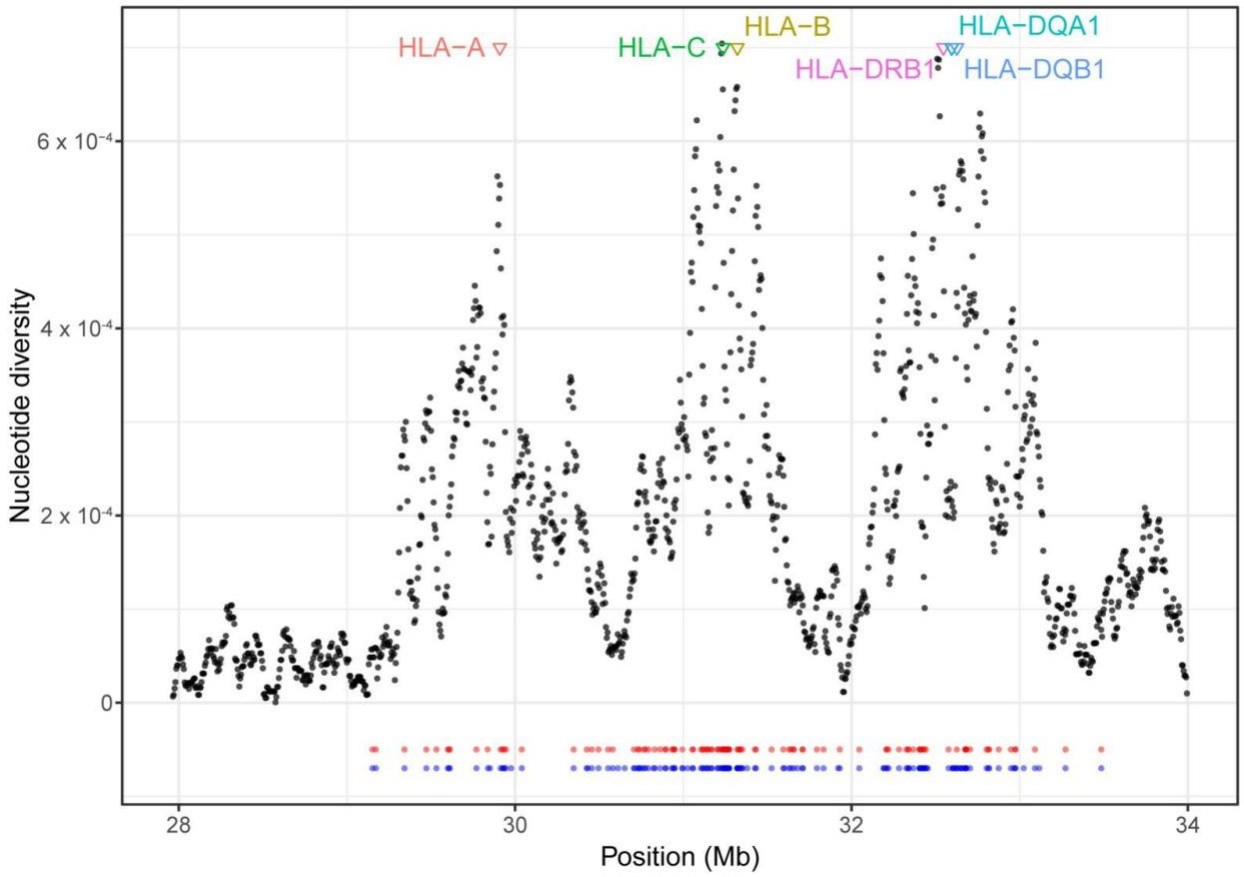

Figure S9: Nucleotide diversity ( ) in our sample at the MHC locus showing high diversity at the HLA genes compared to the genomic background. was calculated in a sliding window of size 50kb with a step size of 5kb with vcftools (2). Location of HLA class-I and class-II genes is indicated on top. Blue points at the bottom indicate position of 195 SNPs tagging HLA alleles in Europeans (3) while red points indicate position of the subset of 114 SNPs genotyped in our sample.

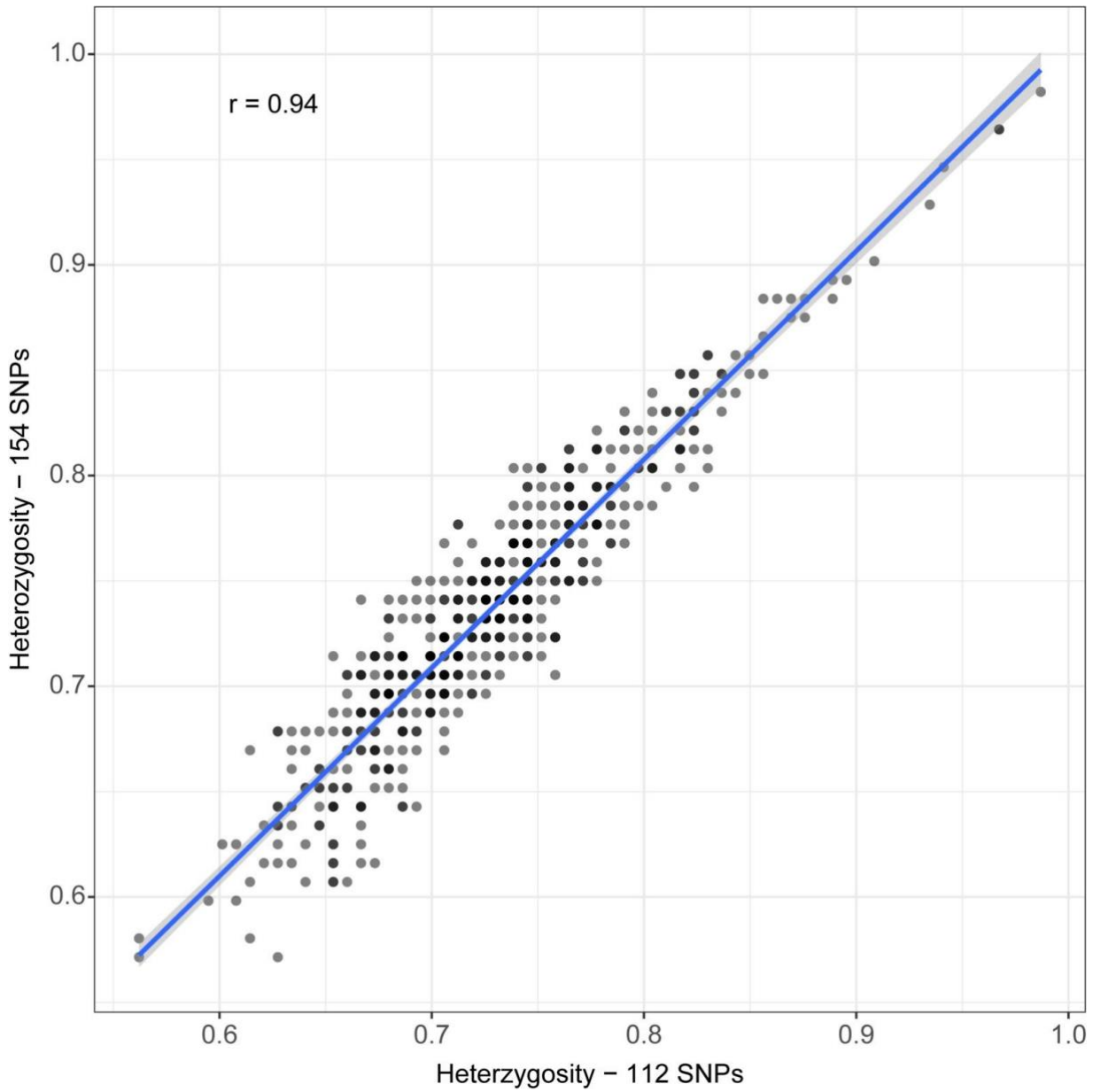

Figure S10: Comparison of Heterozygosity calculated using the full set of 154 SNPs tagging HLA haplotype variation in Europeans (3) and the reduced subset of 112 SNPs for which both 1000G and our sample and were genotyped. Heterozygosity is calculated in 503 individuals of European ancestry from the 1000 Genomes Project Data (1).

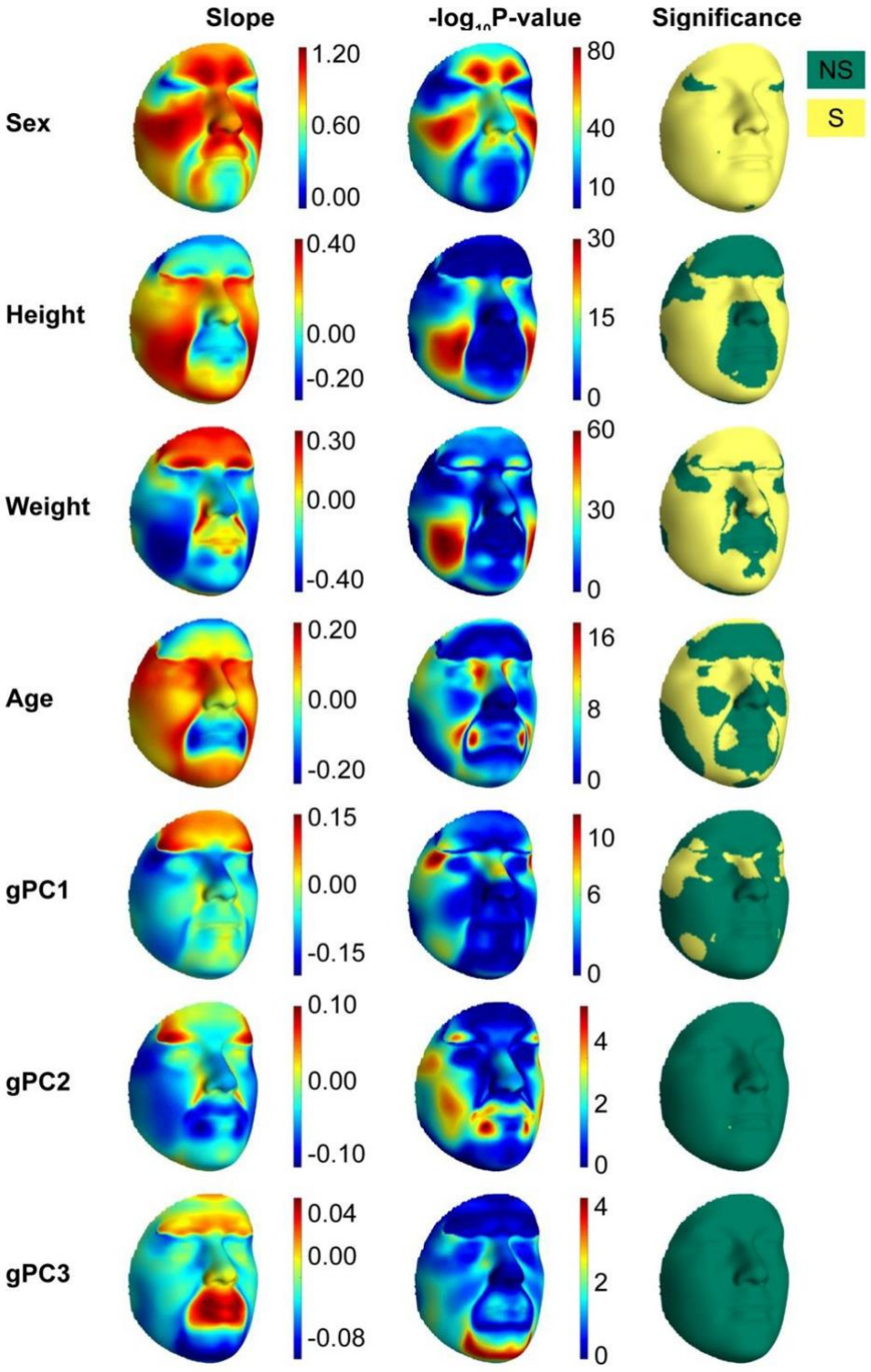

Figure S11: Results of the linear model between FM_{QL} and height, as well as all other covariates included in the model. First column shows a heatmap of the standardized regression coefficients per QL, second column shows the negative $\log_{10}$ of the p-value of the slope, and third column highlights regions that are significant after adjusting for multiple testing. Regions with a Bonferroni corrected P-value of less than 7.00×10^{-06} (0.05/7,150 for 7,150 QLs) are highlighted in yellow.

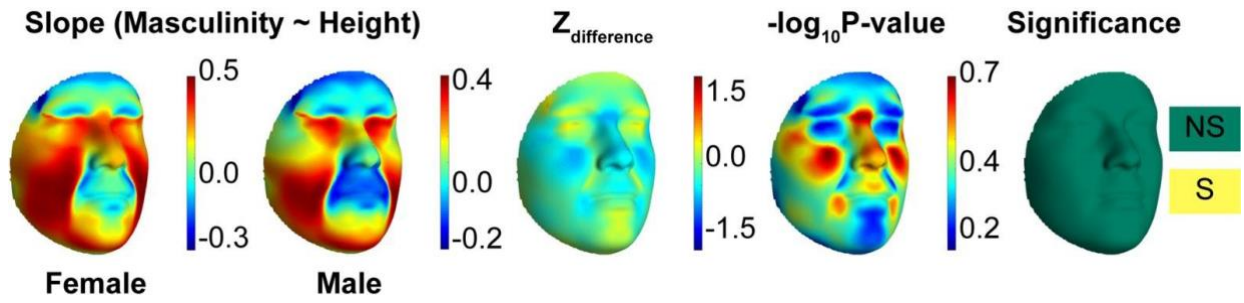

Figure S12: Effect of Height on facial masculinity in females (first column) and males (second column). The Z-score for the difference in slope between males and females is shown in column three. The fourth column shows $-\log_{10}$ of the P-value and the final column shows regions which are significant after correcting for multiple testing (yellow: significant, green: not significant).

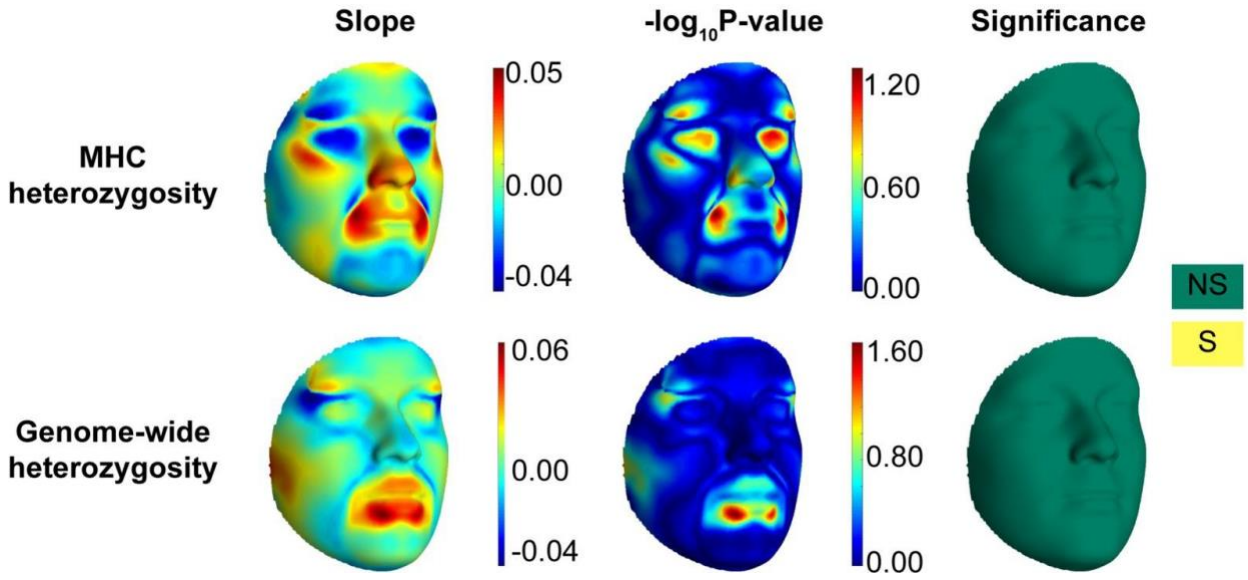

Figure S13: Effect of MHC heterozygosity and genome-wide heterozygosity on facial masculinity. First column shows a heatmap of the slope, second column shows the $-\log_{10}$ of the P-value and third column shows regions that are significant after adjusting for multiple testing (yellow: significant, green: not significant).

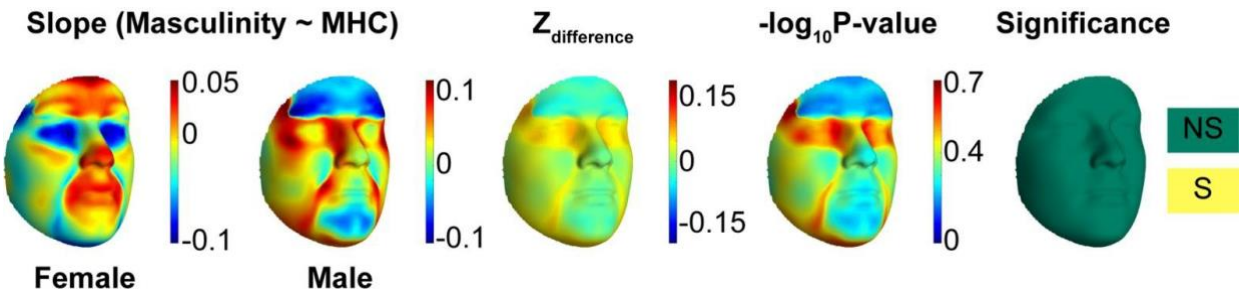

Figure S14: Association between MHC heterozygosity and facial masculinity in females (first column) and males (second column). Third column shows the Z-score for the difference in slope, fourth column shows -$\log_{10}$ of the P-value, and final column shows regions that are significant after adjusting for multiple testing (yellow: significant, green: not significant).

References:

1. 1000 Genomes Project Consortium, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.

2. Danecek P, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

3. de Bakker PIW, et al. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 38(10):1166–1172.