

# Identification of Individuals by Trait Prediction using Whole Genome Sequencing Data

Lippert et al. 10.1073/pnas.1711125114

## Supporting Information (SI)

### Materials and Methods

**Study Design.** Participants for our project were from the greater San Diego area (Protocol No. HLI 2015-001, Western Institutional Review Board, Puyallup, WA); they were recruited by ads, social media, posting signs on university campuses, and by word-of-mouth. Inclusion criteria included both male and female and  $\geq 18$  years of age; exclusion criteria included intravenous drug usage, positive for Hepatitis A, Hepatitis B, HIV-1, and/or HIV-2; mustache and/or beard; and pregnant at the time of participation. The participants were provided with copies of the study protocol, consent, and the California Experimental Subject's Bill of Rights. Participants were asked to review these materials and to ask any questions. Individuals who met the inclusion criteria and agreed to the terms of the study were allowed to participate. In addition, participants agreeing to the use of their image and/or likeness in publications and presentations signed a photo release.

Participants reported sex, age or date of birth, eye color, ancestry, and approximate hours since last shave on an iPad Mini. Weight was measured in kilograms (*kg*) and height in centimeters (*cm*), both without shoes, using the MedVue Digital Eye-Level Physician Scale with attached height rod (DETECTO Scale Company, Webb City, MO).

The face was photographed using the 3dMDtrio System with Acquisition software (3dMD LLC, Atlanta, GA); this is a high-resolution three-dimensional (3D) system equipped with nine machine vision cameras and an industrial-grade synchronized flash system; the 3D 200-degree face was captured in approximately 1.5 milliseconds. If necessary, the participants' hair was pulled away from the face by the use of hairbands and hairpins in order to expose significant facial landmarks. Further, the participants were asked to remove all makeup and facial jewelry, *e.g.*, earrings and nose studs. Each participant sat directly in front of the camera system on a manually controlled height stool; they were asked to focus their eyes on a marking 6" above the center camera and maintain a neutral expression.

In addition, the participants' voices were recorded with both scripted text and a 2-minute minimum non-scripted free speech using The Olympus Digital Voice Recorder WS-822 (Olympus Imaging Corp., Tokyo, Japan) with attached RadioShack Unidirectional Dynamic Microphone (RadioShack, Ft. Worth, TX).

A minimum of 5ml EDTA-anticoagulated blood was collected for all 1,061 participants. The blood was stored at room temperature during the day and at the end of each collection session, they were placed in 4°C storage until extraction.

**Genome Sequencing.** The genome was extracted, quantified, normalized, sheared, clustered, and sequenced. TruSeq Nano DNA HT Library Preparation Kit (Illumina, Inc., San Diego, CA) for next generation sequencing library preparation was

used following the manufacturer's recommendations. Deoxyribonucleic acid (DNA) libraries were normalized and clustered using the HiSeq SBS Kit v4 (Illumina, Inc.) and HiSeq PE Cluster Kit v4 cBot (Illumina, Inc.) and sequenced on HiSeq X Ten System sequencers (Illumina, Inc.) using a 150 base paired-end single index read format following the manufacturer's recommendations. The whole-genome sequencing workflow requires 100ng of DNA. In practice, we have collected 1.25μg per sample. We sequenced the full genome of each participant at an average depth of 41X.

**Estimating Ancestry Information from Genome.** We constructed a reference panel from the Human Genome Diversity Project (HGDP) (1) and the 1000Genome Project (2), and used it for ancestry admixture analysis. We used the two references to estimate ancestry proportions for each individual in our data. Genotype data from HGDP (52 populations) and the 1000Genome Project (26 populations) were merged based on dbSNP rsid positions. SNPs with discordant forward strand alleles between genome build 36 and 38 were removed to avoid assembly inconsistency because they are likely to be assembly errors. In total, 3,444 individuals with 636,698 SNPs were compiled; 116,990 SNPs were then pruned due to linkage disequilibrium (LD) using PLINK 1.9 with the parameters `--indep-pairwise 50 10 0.8`; that is, 50k base pair (bp) window with 10k bp step, and all variants with pairwise  $R^2$  greater than 0.8 were pruned. Allele frequencies were calculated for each population, and the 3,000 most informative SNPs for each population (ranked based on the absolute  $Z$ -score for each allele's frequency against the whole panel of populations) were extracted. The resulting collection of 57,214 unique ancestry informative SNPs were used for ancestry admixture analysis using ADMIXTURE 1.23 (3). We predicted five ancestry components, European (EUR), African (AFR), East Asian (EAS), Central South Asian (CSA), and native American (AMR).

**Quantitative Genotyping.** For the 1,069 individuals from the study cohort, we extracted a set of SNPs from the genome-VCF (variant call format) files of high quality full sequencing data. We accepted the calls for the SNPs that passed the standard quality score threshold (PASS variants) of the Isaac variant caller. All other variants were treated as missing. From this initial set of variants, we filtered to a smaller set of SNPs which we used to compute genomic principal components (PCs) by excluding non-autosomal SNPs, SNPs with a minor allele frequency (MAF)  $f < 5\%$ , SNPs with a missing rate  $\geq 10\%$ , or SNPs found to be in Hardy-Weinberg disequilibrium ( $P < 10^{-4}$ ) on the 1,061 individuals from our cohort. The final set of variants used for computation of PCs consists of 6,147,486 SNPs.

We then constructed the SNP matrix of minor allele dosage values (represented as minor allele counts of 0, 1, or 2). In this matrix, rows represented the individuals and columns represented the SNPs. Missing variants were imputed to

the mean dosage. Each SNP column was scaled by the probability density function of a symmetric *Beta* distribution evaluated at the MAF  $f$  (4).

$$Beta(f|\alpha) = \frac{f^{\alpha-1}(1-f)^{\alpha-1}}{\int_0^1 t^{\alpha-1}(1-t)^{\alpha-1} dt}$$

We chose a shape parameter  $\alpha$  of 0.8 for the symmetric *Beta* distribution. This choice of  $\alpha$  was made to yield a U-shaped distribution that up-weights low frequency variants, but in a less extreme manner than the common practice of standardizing the weights (*i.e.*, division by the standard deviation). Intuitively, such a weighting derives from the fact that, for well-studied traits, associated low-frequency variants have been found to have larger effect sizes than common variants (4). After imputation and scaling, the genomic PCs were computed from the matrix of dosages of the 1,061 individuals in the study cohort.

**Estimating Telomere Length from WGS Data.** We estimated telomere length from WGS data as the product of the size of the human genome and the putative proportion of the telomeric read counts out of total read counts. We considered a read to be telomeric if it contained  $k$  or more telomere patterns (CCCTAA or its complement TTAGGG), where  $k$  is the telomere enrichment level. Thus, the estimated telomere length of individual  $x$ , denoted as  $t_k(x)$  was computed as:

$$t_k(x) = \frac{r_k(x)S}{R(x)N_t}, \quad [1]$$

where  $r_k(x)$  is the count of putative telomeric reads obtained for telomere enrichment level  $k$ ,  $S$  is the size of the human genome (gaps included),  $R(x)$  is the individual's total read count, and  $N_t$  is fixed at 46 for human, the number of telomeres in the genome.

To identify an optimal telomere enrichment level  $k$ , we performed measurement error analysis on 512 WGS runs of the reference individual NA12878 (5). These 512 WGS runs used the same reagent chemistry and were made around the same dates as our CV dataset. We estimated telomere lengths with expression 1 for all runs and enrichment levels. For the measurement error analysis, we compare repeatability between different values of  $k$ . Repeatability was estimated as the variance derived from genetic and environmental effects divided by the total phenotypic variance, or  $1 - v_i/v_p$ , where  $v_p$  is the telomere length variance over our dataset and  $v_i$  is the length variance computed on the NA12878 individual only. In general, repeatability can also be interpreted as the proportion of total variance attributable to among-individual variation. We considered the most repeatable of these runs as our best solution based on the assumption that the true telomere length was constant across all the runs of NA12878. Fig. S1A shows repeatability index curves versus  $k$  over all NA12878 sequencing runs. We found that the curve reached its maximum value of 0.73 for  $k = 4$ . Fig. S1B shows the Pearson correlation coefficient between telomere length estimates and annotated age for our CV set and for all values of  $k$ . As highest correlation was also obtained at  $k = 4$ , evidencing that repeatability provides a good criterion for selecting  $k$ .

We tried to assess the cause for the increased absolute correlations between estimated telomere length and age observed on our study cohort, compared to the  $R = -0.24$  that had

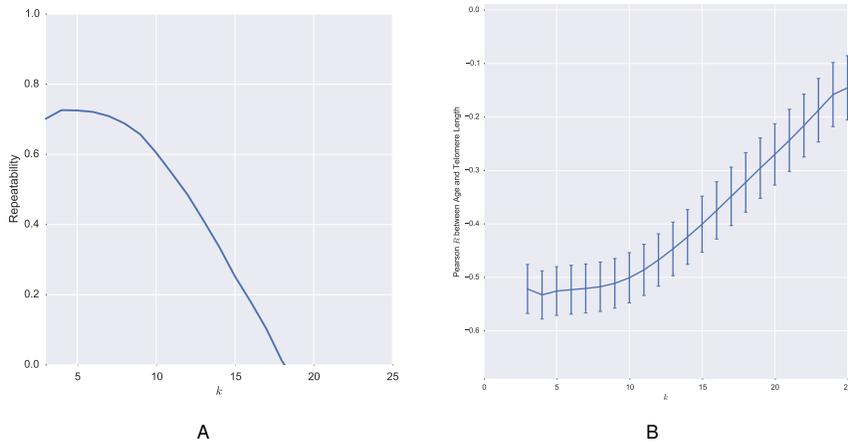
been reported by Ding et al. (6) for 240 leukocyte samples from the TwinsUK cohort. First, we checked if coverage affected estimation of telomere length. To test this, we performed a simulation by uniformly subsampling the reads in the .bam files of the study cohort to achieve various average depths. Measuring quality of the estimate by correlation with age, we confirmed the observation by Ding et al. (6) that the correlation between the telomere length estimate and age plateaus at a read depth of about 2.5X (see Fig. S2). From this analysis we conclude that differences in read depth did not explain the difference in correlation.

Other hypotheses are that the deviation in results may be due to differences in the cohorts, including differences in sample preparation, storage and treatment. We used a set of 1,950 females from the TwinsUK cohort that recently had been sequenced using the same sequencing pipeline as our study cohort (7, 8) to check these hypotheses. If the sequencing pipeline had an influence, then on these individuals we would expect to obtain a higher absolute correlation between estimated telomere length and age, than the correlation reported in Ding et al. (6). However, as shown in Fig. S3, on these individuals we obtain an  $R = -0.19$  between estimated telomere length and age, a value that is comparably low, and even slightly lower, in absolute value than the  $R = -0.24$  that had been reported in Ding et al. (6). This result indicates that the results are dependent on the used cohort and its representativeness.

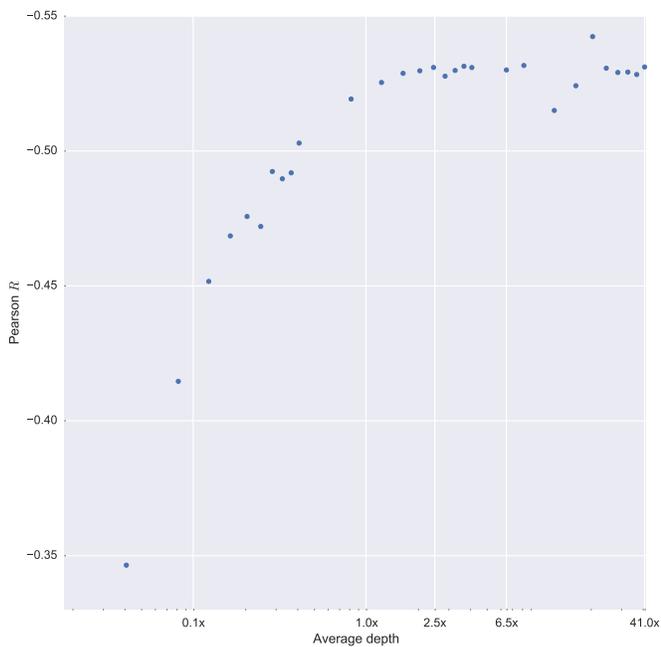
**Computing Chromosome Copy Number Variation from WGS Data.** We used chromosomal copy number (CCN) to quantify the mosaic loss of chromosomes. Naturally, read depth at a chromosome could be used to compute the CCN. However, a large proportion of ChrY is paralogous to some autosomal regions. Many of the reads that mapped to ChrY originate from autosomes. For this reason, prior to computing the copy number of ChrY, we filtered the reads to those that mapped uniquely to ChrY.

More generally, given the HG38 reference genome (RG), we produced a set of uniquely mappable regions, *i.e.*, regions where any 150-mer can be mapped only once throughout the RG. We first simulated 150bp-long reads from the RG at each base position of the genome, and then mapped them to the RG using BWA-mem. Next, we collected the source regions from where the reads originated and mapped only once. Lastly, we removed some repetitive regions annotated by RepeatMasker as `low_complexity`, `retroposon`, `satellite` and `SINE` due to lower region coverage as these regions are more difficult to align. Our approach of generating uniquely mappable regions would work for any version of the RG.

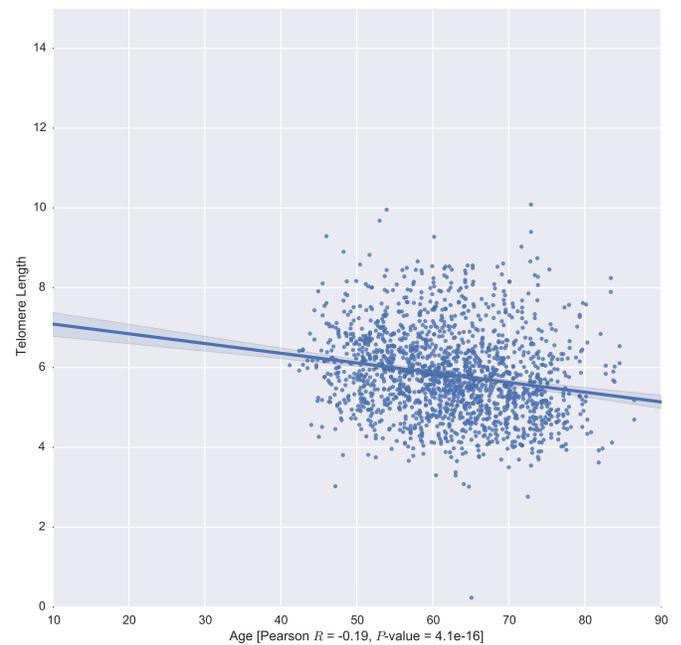
We then selected uniquely mappable regions with length  $> 5$ kb. The length threshold was determined so that each chromosome contained at least 200 bp of each region. We computed read depth of each region using the `samtools mpileup` command, and grouped the regions by their Guanine and Cytosine (GC) content. As can be seen in Fig. S4, genomic regions that have an extreme value in their GC content tend to achieve lower read depth during sequencing. We accounted for this effect, called GC bias, which is known to affect coverage substantially (9) as follows: For a particular GC content group, the median value of the read depth at autosomal regions was used as the baseline value denoted as  $rd_{GC}$ . Here, we assumed a healthy person to have a diploid genome and no detectable



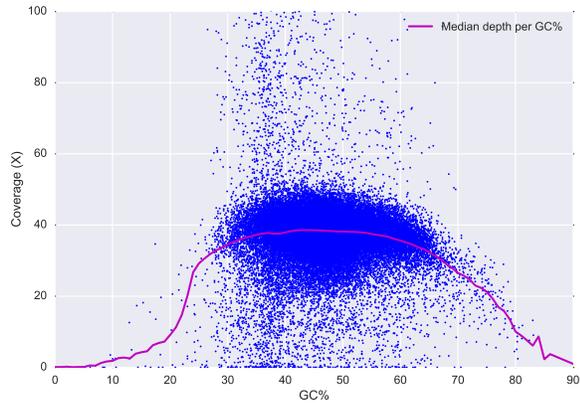
**Fig. S1.** Repeatability and held-out age correlation of telomere length estimates obtained over different values of telomere enrichment level  $k$ . Both NA12878 and the study cohort used the same reagent chemistry for sample preparation. We found that the optimal value of  $k$  depends on the chemistry used (data not shown). (A) For this chemistry, the highest repeatability is achieved at  $k = 4$ . For values of  $k > 18$  the estimates for telomere length become unreliable to the extent that the telomere length variance among individuals is smaller than the variance for NA12878. (B) Pearson correlation coefficient between telomere length and age estimates for our CV set; the best correlation is also obtained at  $k = 4$ .



**Fig. S2.** Evaluation of telomere length estimate as function of read depth. We used `samttools` to randomly sub-sample the reads in our study population to simulate telomere length estimation at varying read depth. As a measure of quality, we used the Pearson correlation coefficient ( $R$ ) between telomere length and age on our study cohort. Lower  $R$  is better. Our cohort  $R$  plateaued above a coverage of 2.5X, similar to previously reported results by Ding et al. (6) on a cohort that had been sequenced at an average coverage of 6.5X.



**Fig. S3.** Correlation between age and telomere length on 1,950 deeply sequenced females from the TwinsUK cohort. The correlation is in line with the correlation of  $R = -0.24$  reported by Ding et al. (6) on individuals from the same cohort that had been sequenced at an average coverage of 6.5X.

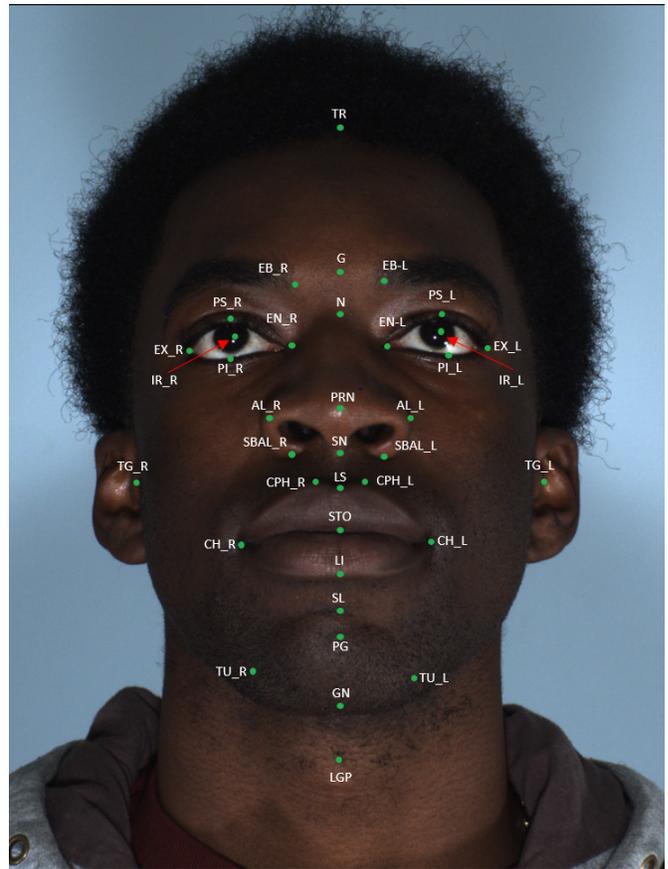


**Fig. S4.** Read depth vs. GC content. Each dot is the average read depth for a  $1kb$  window in chromosome 22, as a function of the GC content in the window. The median line is computed over each one percent bin in GC content. Data is taken from one NA12878 reference sample that had been sequenced at an average read depth of 27.7X and median read depth of 34.5X.

mosaic loss of autosomes. For a region in this GC group, CCN was computed as twice the observed read depth divided by  $rd_{GC}$ . For a given chromosome  $c$ , the CCN was computed as the median CCN of all the regions contained within  $c$ .

**Extracting Single Joint T-cell Receptor Excision Circles.** Several studies have reported that specific somatic DNA rearrangements, called single joint T-cell receptor excision circles (sjTRECs) in T lymphocytes were correlated with age (10, 11). Existing assays based on qPCR on a specific sjTREC ( $\delta Rec-\psi J\alpha$  site) demonstrated that a log-linear relationship between individual age and normalized sjTREC abundance explains a significant portion of the total age variance (12). Therefore, we investigated whether sequences from the sjTRECs could be reliably detected in our genome sequencing data and used as a marker for age. In our investigation, sjTRECs did not show significant signal for age discrimination and we did not use it for our age prediction model. Zubakov et al. (10) reported the signal to be between 2-17 to 2-4 compared to the level of single-copy genes, suggesting that the TRECs are only seen in trace amounts, even in young individuals. This is consistent with our observation that most individuals show no TREC reads at all. Consequently, our read depth at  $\sim 30X$  WGS does not provide the power to detect significant correlation of sjTRECs with age.

We extracted specific structural signatures derived from the somatic excision events at the  $\delta Rec-\psi J\alpha$  site. Specifically, we identified the reads that aligned across the junction of the circular sjTREC, as well as the reads that aligned across the junction of the site of deletion. These junction reads were mapped to two genomic locations on chr14 at a distance of  $\sim 88Kb$  apart. For better sensitivity, the junction reads included both “split reads” as well as the “discordant read pairs” (12) with two paired ends mapped to the two distinct locations of interest. The number of junction reads ranged from zero to three across the individuals that we selected from different age groups. Due to the relatively weak signal that we observed in these selected individuals, the sjTREC signatures



**Fig. S5.** 36 landmarks overlaid on the face of a participant. Landmarks were manually annotated for all participants in our dataset. The landmarks were placed in order from top, going downward in the center, to the right, then left, and bottom. All landmarks in this study were identified visually, *i.e.*, no palpation; the analyst relied upon the 3dMDvultus™ Software v2.3.02 to turn the image and applied the wireframe render mesh of triangles features to annotate each landmark.

identified from our whole genome sequencing did not provide sufficient discriminative power for age prediction. Instead, this particular marker worked well in qPCR assays, perhaps due to the amplification step that exponentially increased the abundance of non-replicated circular sjTRECs which are serially diluted with each cellular division.

## Quantitative Phenotyping

### Landmarking 3D Images and Extracting Landmark Distances.

Facial landmarking is an important basic step in our face modeling procedure as they are used to align face images, and to compute landmark distances (*e.g.*, distance between the inner edges of left and right eyes and width of the nose). A total of 36 landmarks for each 3D image was measured using 3dMDvultus™ Software v2.3.02 (3dMD LLC). Each measurement is precise to 750 microns. The landmarks and their definitions were adopted from [www.facebase.org](http://www.facebase.org) (13), with the addition of the laryngeal prominence (Table S1). Fig. S5 illustrates facial landmarks overlaid on a face image.

We manually annotated the landmarks in the 3D images using 3dMDvultus™ Software (3dMD LLC). The landmark annotations were carefully determined; some of the landmark positions required careful examination at different angles. For example, pronasale is the most protrusive point on the tip

**Table S1. Facial landmarks. 36 landmarks were manually placed on each of the 3D images.**

Landmark	Label	Definition
Trichion	TR	Point where the normal hairline and middle line of the forehead intersect.
Glabella	G	Mid-point between the eyebrows and above the nose; underlying bone which is slightly depressed, and joins the two superciliary ridges; cephalometric landmark that is just superior to the nasion.
Nasion	N	Intersection of the frontal bone and two nasal bones of the human skull; distinctly depressed area directly between the eyes, just superior to the bridge of the nose; just inferior to the glabella.
Eyebrow Right (or Left)	EB_R or L	Lower corner of where the eyebrow begins.
Endocanthus Right (or Left)	EN_R or L	Innermost corner of the eye where where tear duct and the skin meet
Palpebrale Superious Right (or Left)	PS_R or L	Highest point of the upper eyelid, directly above the iris landmark and on the eyelash line.
Ectocanthus Right (or Left)	EX_R or L	Outermost corner of the eye
Palpebrale Inferious Right (or Left)	PI_R or L	Highest point of the lower eyelid, directly below the iris landmark and on the eyelash line.
Iris Right (or Left)	IR_R or L	Center of the pupil.
Pronasale	PRN	Tip of the nose.
Alar Right (or Left)	AL_R or L	Midpoint of the outer flaring cartilaginous wall of the outer side of each nostril. The ala of the nose (wing of the nose) is the lateral surface of the external nose.
Subalar Right (or Left)	SBAL_R or L	Lowest point where the nostril and the skin on the face intersect; located inferior to the "alar" landmark.
Subnasale	SN	Lowest point of the nasal septum intersects with the skin of the upper lip.
Labiale Superious	LS	Midline, between the philtral ridges, along the vermillion border of the upper lip; uppermost point in the center of the upper lip where the lip and skin intersect.
Crista Philtri Right (or Left)	CPH_R or L	Highest point of the philtral ridges, or crests that intersect with the vermillion border of the upper lip.
Chelion Right (or Left)	CH_R or L	Outermost corner, commissure, of the mouth where the upper and the lower lips meet.
Labiale Inferius	LI	Midline along the vermillion border of the lower lip; lowermost point in the center of the lower lip where the lip and skin intersect; midline along the inferior vermillion border of the lower lip.
Stomion	STO	Center point where upper and lower lips meet in the middle; easily identified when lips are closed, point can still be identified when the lips are apart by placing the landmark along the inferior free margin of the upper lip.
Sublabial	SL	Most superior point of the chin, above the pogonion; verify with lateral view.
Pogonion	PG	Most projecting median point on the anterior surface of the chin; verify with lateral view.
Gnathion	GN	Inferior surface of the chin/mandible; immediately adjacent to the corresponding bony landmark on the mandible.
Tuberculare Right (or Left)	TU_R or L	The slight depression of the jawline somewhere between the gnathion and the gonion.
Tragion Right (or Left)	TG_R or L	Small superior notch of the tragus (cartilaginous projection just anterior to the auditory meatus).
Gnathion	GN	Inferior surface of the chin/mandible; immediately adjacent to the corresponding bony landmark on the mandible.
Tuberculare Right (or Left)	TU_R or L	The slight depression of the jawline somewhere between the gnathion and the gonion.
Tragion Right (or Left)	TG_R or L	Small superior notch of the tragus (cartilaginous projection just anterior to the auditory meatus).

The name of the landmark, the label used in our studies, and the definition of each landmark are provided. \_R and \_L signify the same landmark on the right and left side of the face.

of nose; the image must be turned 90° to accurately place this landmark. Given the annotated landmarks, we defined 27 facial landmark distances between a pair of landmarks (Table S2 for the details of the landmark distance definitions).

**Extracting Facial Embedding.** To predict facial structure from genome effectively, we used a low dimensional numerical representation of face, which adequately represents intra-individual variation. For this purpose various algorithms have been used including PC analysis (14), linear discriminant analysis (15), neural networks (16), and others (17, 18). In this study, we used PC analysis because it allows us to discriminate different faces, and importantly, to reconstruct predicted faces.

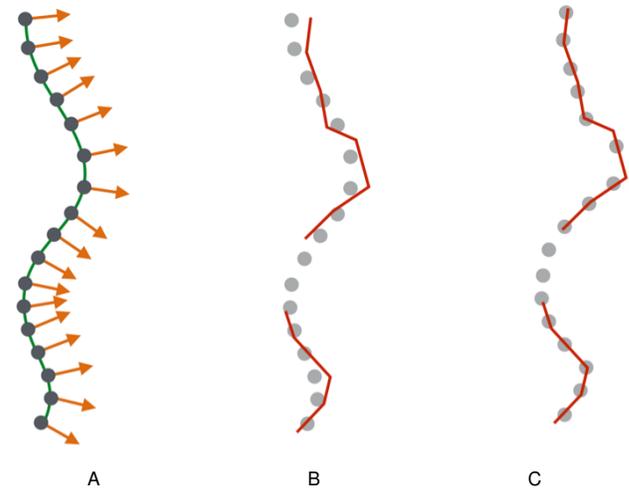
Our pipeline for 3D face phenotyping is similar to the work of Turk et al. (19). We start from a neutral 3D face template and align this template in a non-rigid manner to the 3D scans using an expectation maximization (EM) algorithm. At each iteration, we approximate correspondences between the 3D scan and the deformed version of the template mesh (E-step) and optimize deformation parameters to bring the established correspondences as close to each other as possible (M-step). Because the deformation is a global operation and it applies to the entire face images, the set of correspondences might change after the M step. We iterate until the error (*i.e.*, distance between the template face and the 3D scan) is minimized. Our deformation model is 3D thin plate splines where the degrees of freedom are the weights of knots manually placed on the template mesh (19).

Once the template model was deformed to match the 3D scan, we computed a displacement over the template mesh to capture the fine scale surface details in our 3D scans. Specifically, we traced rays along the normal vectors of the template mesh and displaced template vertices to the intersection points of these rays with the 3D scans, as illustrated in Fig. S6. Alignment of 3D scan of face images to the template face model. This also allowed us to copy the colors from 3D scans onto the template mesh. The areas on the template mesh where the rays do not intersect the scan (either due to noise or scanning problems) were filled using Poisson image editing (20).

Using these procedures, we obtained the deformed template mesh and aligned it to every 3D scan. Because the purpose of facial embedding is not to capture variations in position and orientation of the head at the time of the scan, we aligned the deformed version of the template to the original template mesh. This final alignment was performed using a rigid body transform computed using the work of Horn (21).

Methods similar to the ones used in the paper have previously been employed by Belongie et al. (22) in computer vision, Amberg et al. (23) in computer graphics, and by Guo et al. (24) in bioinformatics. Similar to our method, these methods warp a template mesh onto a 3D scan surface using nonlinear deformations (similar to Amberg et al. (23) and Guo et al. (24), we use thin plate splines). The result of the warping process is a displacement vector over the set of template mesh which is suitable for quantitative phenotyping.

The observed color of the face is a product of the skin reflectivity and the incident lighting from the environment. Skin reflectivity is a measurement we attempted phenotype; however, we do not have the precise measurement of incident illumination. Thus, we created a first order approximation by assuming that skin reflectivity is diffuse (incident light at a point is scattered equally in all outgoing directions) which is



**Fig. S6.** Alignment of 3D scan of face images to the template face model. To minimize the noise due to face image misalignment between different faces of individuals, we aligned face 3D images by matching the vertex of the average template face and each individual face. (A) The vertices of the average template face and their normal vectors. (B) Gray vertices represent the vertex in the average template. Red solid lines represent the scanned face surface for the observed individuals. (C) Average face template vertices are displaced along their normal vectors to the closest observed scanned surface. If there is no scanned surface near the template vertices, the closest scanned surfaces are estimated using the Poisson method.

approximated by albedo, or reflection coefficient. Albedo (25), which models face under different lighting conditions, yields a bilinear form used similarly in the work by Yu and Ahuja (26), and we solved it by iterating the following steps alternatively until convergence: (1) estimate the albedo term while keeping incident lighting fixed; (2) estimate incident lighting which we assume to be constant across the face images while keeping the albedo constant. Finally, we obtained our face embedding that consists of PCs from all vertex positions on the deformed template (shape PCs) and the solved surface albedo at every vertex (color PCs).

**Extracting Eye Color.** To extract eye color, we used the 2D face images. To speed up manual annotation of eyes in the images, we employed a LeNet convolutional neural network (CNN) (27) to locate eyes in facial images and extracted the left and right eyes. First, we manually extracted an initial set of 500 eye patches from 250 facial images. An example of an extracted eye position is shown in Fig. S7A. Additionally, we extracted 500 random patches of the same size as negative examples. Using these eye patches and negative examples, we trained a CNN with the following LeNet architecture:

- Input patch  $260 \times 120$  pixels
- Convolution layer  $11 \times 11$ , stride 1, 10 channels
- max pooling layer  $2 \times 2$
- Convolution layer  $11 \times 11$ , stride 1, 20 channels
- max pooling layer  $2 \times 2$
- fully connected hidden layer size 50
- 1-hot encoded output variables with logistics variables

**Table S2. Calculated facial landmark distances.**

Landmark Distance Symbol	Definition
TGL_TGRpa	Measurement from TG_R through the PRN to TG_L; it is actually a diarc with two arcs combined; curved line from left ear to right ear through the pronasale.
TR_GNpa	Measurement from TR through the PRN to the GN; diarc of two arc measurements combined; curved line from the hairline to just underneath the chin through the pronasale.
EXR_ENR	Width of the right eye.
PSR_PIR	Height of the right eye.
ENR_ENL	Distance from inner left eye to inner right eye.
EXL_ENL	Width of the left eye.
EXR_EXL	Distance from outer left eye to outer right eye.
PSL_PIL	Height of the left eye.
ALL_ALR	Width of the nose.
N_SN	Height of the nose.
N_LS	Distance from top of the nose to top of upper lip.
N_ST	Distance from top of the nose to center point between lips.
TGL_TGR	Straight distance from left ear to right ear.
EBR_EBL	Distance from inner right eyebrow to inner left eyebrow.
IRR_IRL	Distance from right iris to left iris.
SBALL_SBALR	Width of the bottom of the nose.
PRN_IRR	Distance from the tip of the nose to right iris.
PRN_IRL	Distance from the tip of the nose to left iris.
CPHR_CPHL	Distance separating the crests of the upper lip.
CHR_CHL	Width of the mouth.
LS_LI	Height of lips.
LS_ST	Height of upper lip.
LI_ST	Height of lower lip.
TR_G	Height of forehead.
SN_LS	Distance from bottom of the nose to top of upper lip.
LI_PG	Distance from bottom of the lower lip to the chin.

Of the 36 landmarks in Table S1, distances could be calculated from any two landmarks; using three landmarks, a polyarc (pa) curved line distance across three landmarks was calculated. Above is a partial list of the distances measured.

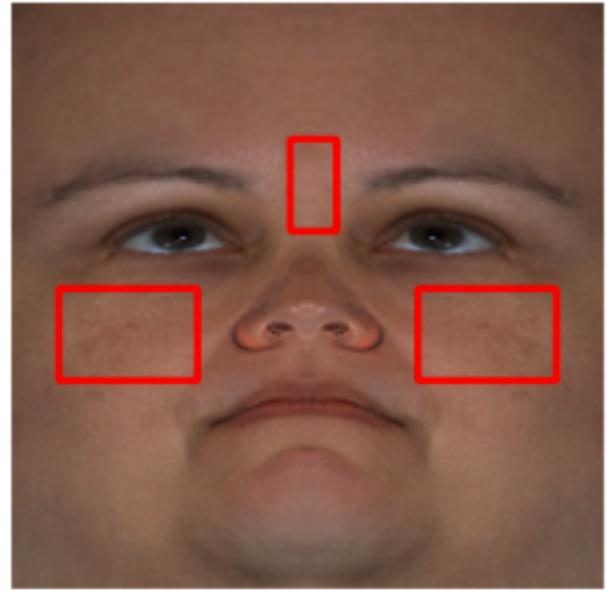


**Fig. S7.** Automatic extraction of the iris area from 2D eye images. (A) An eye image extracted from a face image. (B) Blue area showing the identified iris by the proposed iris extraction method.

We repeated the following process twice, until all eyes were located correctly: After the model had been trained with a batch size of 10, we predicted each image patch in a sliding window fashion with a stride of  $2 \times 2$  pixels. After visual inspection of the top two predictions per image, we generated a larger training set by adding correctly predicted eye locations to the positive examples and false positives to the negative examples. After augmenting the negative examples by random patches to match the number of positive examples, we retrained the classifier. We performed the following procedure to extract iris pixels: (1) converted each eye image to gray scale and performed OpenCV histogram normalization to improve the contrast of the image; (2) detected edges using radial edge detector based on the Sobel operator (28) and chose the iris circle by finding the locations that best match the detected edge signal; (3) located the convex hull of the iris circle; (4) eliminated the pupil area by blocking a fixed radius of size around the center of the circle; and, (5) calculated the brightness histogram for the points in the iris circle, and retained the points in the middle 80% of the histogram, which eliminated reflections and any remaining black pupil points. The result is a set of identified iris pixels (Fig. S7B). We represent these pixels in the RGB color space and calculate the mean value for each R, G, and B parameter to obtain an overall iris color for the eye. We found that the measured eye color for the two eyes was very close, as expected; thus, we used an average of both eyes as the raw color values for the subject. As a subset of our images used three flashes, while the remaining data used two flashes, we performed brightness correction by converting the extracted eye colors to hue (H), saturation (S), and value (V) color space and used 40 individuals that had been photographed under both illumination conditions to estimate a linear function to adjust the V channel such that they matched between conditions. All extracted eye colors underwent this adjustment and were converted back to RGB color space.

**Extracting Skin Color.** To obtain skin color from the 2D image scan, we extracted three skin patches (patches from the forehead and two from the cheek just below each eye) from albedo-normalized and aligned face photos (Fig. S8). To remove the outliers in the skin color, we used  $k$ -medoid clustering ( $k = 3$ ) and chose the RGB values for the cluster center with the medium lightness to account for non-uniform light reflection from the skin surface.

**Extracting Voice Embedding.** We used the Spear open-source speaker recognition toolkit to create low-dimensional voice feature vectors, referred to as identify-vectors or  $i$ -vectors, obtained by a joint factor analysis (29). The Spear toolkit is a state-of-the-art open source toolbox developed for speaker



**Fig. S8.** Three skin patches used for skin color estimation. The three rectangular skin patches are superimposed onto an albedo normalized face image.



**Fig. S9.** Pipeline for  $i$ -vector generation. The Spear toolbox transforms voice samples into  $i$ -vectors through a multi-step pipeline process. After a voice sample is collected, it uses an activity detector based on audio energy to trim out silence from the sample. Next, the Spear toolbox applies a Mel-Frequency Cepstrum Coefficient feature extractor (30) that converts successive windows of the sample to Mel-Frequency Cepstrums. Finally, it projects out the universal background model (UBM) to account for speaker- and channel-independent effects in the sample, and computes the  $i$ -vector corresponding to the original sample.

recognition. We chose  $i$ -vector for our voice representation because it models both speaker and channel variability simultaneously, and can extract the speaker's voice in a compact representation. Full pipeline for generating  $i$ -vectors is illustrated in Fig. S9.

## Experimental Setup

In the following, we provide details on the experimental setup used to train and evaluate predictive models, as well as matching algorithms. We generally employed nested  $l$ -fold CV procedures, where the outer loop was used to evaluate models out-of-sample, whereas nested inner loops were used to obtain any tuning parameters.

**$l$ -Fold Cross-Validation Scheme.** Individuals were deterministically assigned to one of  $l$  CV sets based on a hash function computed on an anonymized subject identifier. This assignment procedure was deterministic, assuring that the assignment of individuals to sets is the same for each experiment. The process is equivalent to uniform assignment of a set and, while it did not guarantee that all sets have exactly the same size, the expected size per fold is the same.

Closely related relatives may share not only a large portion of the genome but also environmental factors. To avoid

splitting related individuals between training and test sets and thus to avoid environmental confounding between training and test sets, we post-processed the cross-validation sets as follows: We first computed the transitive closure for all groups of related individuals and then re-assigned all individuals in such a group to the CV set that had initially been assigned to the individual in the group with the lexicographically first identifier. For details on how we determined relatedness, see the section on “Identifying Related Individuals”.

**Training predictive models using cross validation.** We evaluated each predictive model using  $l$ -fold CV. Unless noted otherwise, ridge regression was used as a predictive model for each trait, and  $l$  was set to be 10. For each of  $l$  CV folds, we used the union of  $l - 1$  CV sets as the training set and the remaining set as the test set such that each individual was used for training exactly nine times and used for testing exactly once. For each repetition, we chose any tuning parameters for the predictive models by performing a nested 5-fold CV within the training set only (*i.e.*, within the union of the nine CV sets of individuals used for training). After the tuning parameters had been determined, the model was retrained using the whole training set. The trained model is then used to predict on the test set. Finally, for each fold the predictions are evaluated using a quality metric that is evaluated on the test set.

The result of this procedure is  $l$  trained models and a set of predictions for each of the  $l$  CV sets, such that each individual had been predicted exactly once. Unless otherwise noted, the final evaluation score was returned as the average of the  $l$  scores obtained on each CV set.

**Setup for Evaluating Re-Identification Performance.** The following is the experimental setup for training and evaluating both  $s_N$  and  $m_N$  algorithms. For details on how we learn a similarity function, as well as definitions, see Supplemental Section “Learning a Similarity Function”.

We perform the following procedure to learn and evaluate similarity metrics for a lineup size  $N$  using 3-fold CV, to ensure that CV sets contain more individuals than the lineup sizes.

For each of 3 folds:

- We define the test data to be the current CV set and the training data to be the remaining nine CV sets.
- For the training data, we obtain a set of embedded phenotypes  $\psi_{\mathcal{P}}(p_{n_{tr}})$  and genomes  $\phi_{\mathcal{P}}(g_{n_{tr}})$  using the CV scheme for training predictive models ( $l = 10$ ) described in the previous section on the training data and taking the union of all predicted values. This way, we obtain a set of predictions that may be considered out-of-sample for each individual  $n_{tr}$  in the training data, without involving any individuals in the test data.
- We randomly assign each individual in the training data to exactly one of  $\text{floor}(N_{tr}/N)$  training lineups of size  $N$ . This implies that  $N_{tr} \bmod N$  individuals are discarded.
- We train an optimal distance metric on the training lineups using YASMET, a maximum entropy model (31).
- We generate test lineups, by forming all possible lineup combinations in the test data.

- We generate the embedded phenotypes  $\psi_{\mathcal{P}}(p_{n_{te}})$  and genomes  $\phi_{\mathcal{P}}(g_{n_{te}})$ , for all individuals  $n_{te}$  that belong to the current test data, re-learning the models using 10-fold CV on the current training data as described above.
- We compute the similarity function that had been estimated on the current training data for all pairs of embedded phenotypes  $\psi_{\mathcal{P}}(p_{n_{te}})$  and genomes  $\phi_{\mathcal{P}}(g_{n'_{te}})$  within each test lineup and evaluate  $s_N$  and  $m_N$ . In  $s_N$ , we simply pick the entity in the pool that has the highest probability of matching the probe. For  $m_N$ , we choose all pairs so as to maximize the total probability of matching within the set of  $N$  pairs. This is performed using the blossom method (32), as implemented by the “max weighted matching” function from the Python package NetworkX.

## Models and Evaluation Metrics

In the following, we provide details on the various models we employed and the evaluation metrics used in the manuscript.

**Ridge Regression for Trait Prediction.** Unless stated otherwise, we fit a ridge regression on the training data set, where the regularized sum of squares was minimized over an offset  $c$  and a set of regression coefficients  $\beta_d$ . For a given individual with index  $n$  out of  $N_{train}$  training individuals, the residual  $r_n$  is defined as the difference between the phenotype value  $y_n$  and a linear regression in the covariates  $x_{nd}$ :

$$r_n = y_n - \left( c + \sum_{d=1}^D x_{nd} \beta_d \right).$$

The optimal coefficients are given by

$$\arg \min_{\beta, c} \left( \sum_{n=1}^{N_{train}} r_n^2 \right) + \alpha \left( \sum_{d=1}^D \beta_d^2 \right).$$

For each repetition, an optimal regularization parameter  $\alpha$  was estimated by a standard nested 5-fold CV over the training data. Given the  $\alpha$ , we predicted the phenotype on the remaining set of test individuals.

We measured prediction accuracy using the out-of-sample measure

$$R_{CV}^2 = 1 - \frac{\sum_{n=1}^{N_{test}} r_n^2}{\left( \sum_{n=1}^{N_{test}} y_n - \bar{y}_{test} \right)^2},$$

where  $\bar{y}_{test}$  is the in-sample mean of the test phenotypes. Note that this measure has a negative expectation for random predictions. In fact, it is negative whenever the in-sample mean of the test data  $\bar{y}_{test}$  has a smaller squared error than the model that had been trained on the training data. Also, in contrast to an in-sample  $R^2$ ,  $R_{CV}^2$  is not expected to increase by adding more covariates to the model, because the model has been fit to the training data set.

**Identifying Related Individuals.** The realized relationship matrix (33) is an unbiased estimator of the kinship between any two individuals  $n$  and  $n'$ . The realized relationship between two individuals is defined as the inner product of their standardized SNP dosages divided by the number of SNPs  $S$ .

$$k(n, n') = \frac{1}{S} \sum_{s=1}^S z_{ns} z_{n's},$$

where  $z_{ns}$  denotes the dosage for the  $n$ -th individual and the  $s$ -th SNP and standardized to have zero mean and unit variance over all individuals.

We computed the realized relationship matrix for each of the three different genetic ancestries (European, African, East Asian). Fig. S10 displays the empirical distribution for realized relatedness between all pairs of individuals in each population. We used 0.3 as the relatedness threshold by evaluating at the empirical distribution of the coefficients in each population.

**Evaluation Metrics for Individual Re-identification.** To assess the effectiveness of our models for the individual re-identification task, we evaluated our predictions via two performance metrics, referred to as select at  $N$  ( $s_N$ ) and match at  $N$  ( $m_N$ ).  $s_N$  is defined as the accuracy in picking a genomic query's corresponding phenotype entity out of a pool of size  $N$ .  $m_N$  represents the task of uniquely pairing  $N$  queries to  $N$  corresponding phenotype entities.

In both settings,  $s_N$  and  $m_N$ , it is assumed that the individual is already known to be part of a lineup. The related problem of determining whether an individual is part of the lineup, or not, (in contrast to our task of identifying a given individual within a lineup) is beyond the scope of this paper. This related setting is considered, for example, in Craig et al. (34), where summary statistics are used to derive a statistical test for whether an individual is part of a cohort (*i.e.*, the lineup), or not. In the context of phenotypic prediction, a corresponding statistical test could in principle be derived based on the same integrated similarity measure proposed in this paper, but would require further research.

The features for  $s_N$  and  $m_N$  are the average absolute differences between each observed trait set and each predicted traits set generated by the predictive models. Between feature sets (*e.g.*, face shape, eye color, etc.) the number of individual variables may be quite different. Residuals are averaged across the variables of a feature set to ensure that the influence of a feature set was not correlated with the number of variables within it.

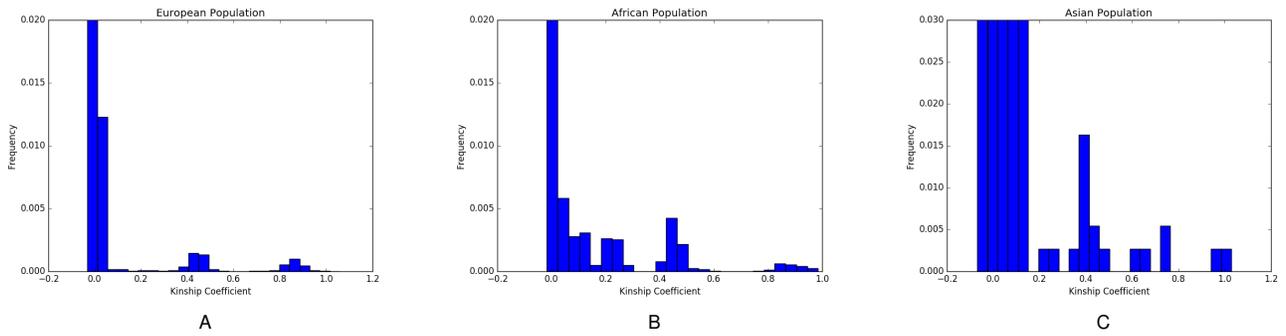
## Supplementary Results

**Face Prediction from Genome.** The shape of the human face is largely genetically determined as evident from the facial similarities between monozygotic twins or closely related individuals. The heritability estimates of craniofacial morphology range from 0.4 to 0.8 in families and twins (35, 36). Despite that, the genetic features responsible for craniofacial morphology remain largely unexplored. Liu et al. (37) reported 12 SNPs influencing facial morphology in Europeans. Claes et al. (38) employed a partial least squares (PLS) regression method, called "bootstrapped response-based imputation modeling", to model variation of the face, and found 24 SNPs from 20 craniofacial genes in individuals from three West African/European admixed populations correlated with face shape. However, in a subsequent study (39) the authors reported that most of the signal came from the genetic ancestry and sex provided, while the effect of the SNPs was marginal.

In this section, we present the results of our prediction of the face from the genome, and discussion on the SNPs associated with facial shapes. We developed a predictive model for face prediction (*i.e.*, predicting face shape and color PCs) using genetic features, sex, and predicted age from genome and BMI on a set of unrelated individuals of ethnically diverse background. In principle, any regression method could be used to predict face shape and color PCs. However, after a preliminary comparison of various regression models, including ridge regression, lasso, ridge regression with stability selection, extreme boosted trees, support vector regression, feed-forward neural networks, and  $k$ -nearest neighbor regression, we concluded that on our data set the choice would have little impact on the results. We chose ridge regression for our face prediction since it is simple and computationally efficient. The cross-validated results for different combinations of covariates that either had been predicted from the genome, or ground truth covariates, are given in Table S3 (Shape) and Table S4 (Color). Sex, genomic ancestry, and age provide the largest contributions to the accuracy of the models. We report both as well as  $s_{10}$  numbers. True faces next to predicted faces by both Ridge and  $k$ -Nearest Neighbor methods for 24 consented individuals that were assigned to the holdout set are given in Fig. S11. 3D faces of three selected individuals from the holdout set scanned and predicted using Ridge regression are provided in Fig. S12.

To examine the effect of ancestry on variability in face shape predictions, we created a group of individuals with  $> 80\%$  AFR ancestry and  $> 80\%$  EUR ancestry. Table S5 presents the AFR:EUR ratio of the standard deviation for each of the first ten face shape PCs. Taken together, these ten PCs explain 91.5% of the variance of face shape in our study cohort. For individuals with  $> 80\%$  AFR ancestry, these explain 88.9% variance in face shape and for individuals with  $> 80\%$  EUR ancestry these explain 90.0% of the variance in face shape. Even though the features describe comparable amounts of phenotypic variance in Africans and in Europeans, the predictions were more variable for those with high African ancestry than those with high European ancestry. To visualize the regions in the face that are most differentiated between individuals of African and European descent, we trained a PLS model that included five regions of ancestry (EUR, AFR, EAS, CSA, and AMR) and sex as PLS dimensions. Fig. S13 shows the pixel-wise absolute differences between individuals that are of 100% AFR ancestry and individuals that are of 100% EUR ancestry (Fig. S13A for males, Fig. S13B for females).

To investigate SNPs associated with the face shape and color, we have performed association testing between the top ten PCs from our face shape and color embedding and the reported SNPs. When we tested for the associations having sex, BMI, and age as covariates, the genomic control inflation factor  $\lambda_{gc}$  on this set of tests was 5.96, which indicates strong confounding effects in the tests. The  $\lambda_{gc}$  statistic is defined as the ratio of the median of observed statistic to the median of the expected statistic under null distribution, and  $\lambda_{gc} > 1$  indicates an inflation of statistics due to confounding. In our analysis, we found strong indication for confounding by population structure. After adding five ancestry proportions as covariates,  $\lambda_{gc}$  dropped to 1.15. At an alpha level of 0.05, none of the 36 candidate SNPs were significant after Bonferroni correction ( $P < 7 \times 10^{-5}$ ). The corresponding Quantile-



**Fig. S10.** The distribution of coefficients in realized relatedness matrix for three different populations. (A) European, (B) African, (C) East Asian. The coefficients in realized relatedness matrix represents the amounts of the relatedness between pair of individuals. The histogram is truncated in order to show the tail of the distribution. Based on this empirical distribution, we decided to use the relatedness threshold of 0.3.

**Table S3. Evaluation of face shape prediction using predicted and observed covariates.**

Covariate Set	Predicted Covariates						Observed Covariates					
	Male	$s_{10}$		$R^2$			Male	$s_{10}$		$R^2$		
		Female	All	Male	Female	All		Female	All	Male	Female	All
PCs	0.26	0.25	0.35	0.17	0.13	0.29	0.26	0.25	0.35	0.17	0.13	0.29
PCs + Age	0.26	0.26	0.36	0.18	0.14	0.30	0.30	0.29	0.39	0.20	0.15	0.31
PCs + Age + BMI	0.26	0.26	0.36	0.18	0.14	0.30	0.35	0.35	0.44	0.25	0.23	0.36
PCs + Age + Height	0.28	0.26	0.37	0.19	0.14	0.30	0.32	0.30	0.41	0.25	0.18	0.35
PCs + Age + BMI + Height	0.28	0.26	0.37	0.18	0.14	0.30	0.36	0.36	0.46	0.30	0.26	0.40

Cross-validated results for different combinations of covariates (age, sex, BMI and height are phenotyped) for ten face shape PCs for Ridge Regression. “PCs” are the top 1,000 genomic PCs. PCs and sex are responsible for most of the performance gain, age (observed), BMI (observed) and height (observed) added small improvement in performance.

**Table S4. Evaluation of face color prediction using predicted and observed covariates.**

Covariate Set	Predicted Covariates						Observed Covariates					
	Male	$s_{10}$		$R^2$			Male	$s_{10}$		$R^2$		
		Female	All	Male	Female	All		Female	All	Male	Female	All
PCs	0.34	0.34	0.34	0.81	0.70	0.74	0.34	0.33	0.34	0.17	0.13	0.74
PCs + Age	0.34	0.35	0.35	0.80	0.70	0.74	0.38	0.36	0.37	0.20	0.15	0.75
PCs + Age + BMI	0.34	0.35	0.35	0.80	0.70	0.74	0.38	0.37	0.37	0.80	0.70	0.75
PCs + Age + Height	0.37	0.34	0.36	0.81	0.81	0.75	0.38	0.36	0.37	0.80	0.70	0.75
PCs + Age + BMI + Height	0.37	0.34	0.36	0.81	0.70	0.75	0.38	0.37	0.37	0.80	0.70	0.75

Cross-validated results for different combinations of covariates (age, sex, BMI and height are phenotyped) for ten face shape PCs for Ridge Regression. “PCs” are the top 1,000 genomic PCs. PCs have the largest contribution to the model performance, sex and then age (observed) add incremental gains.

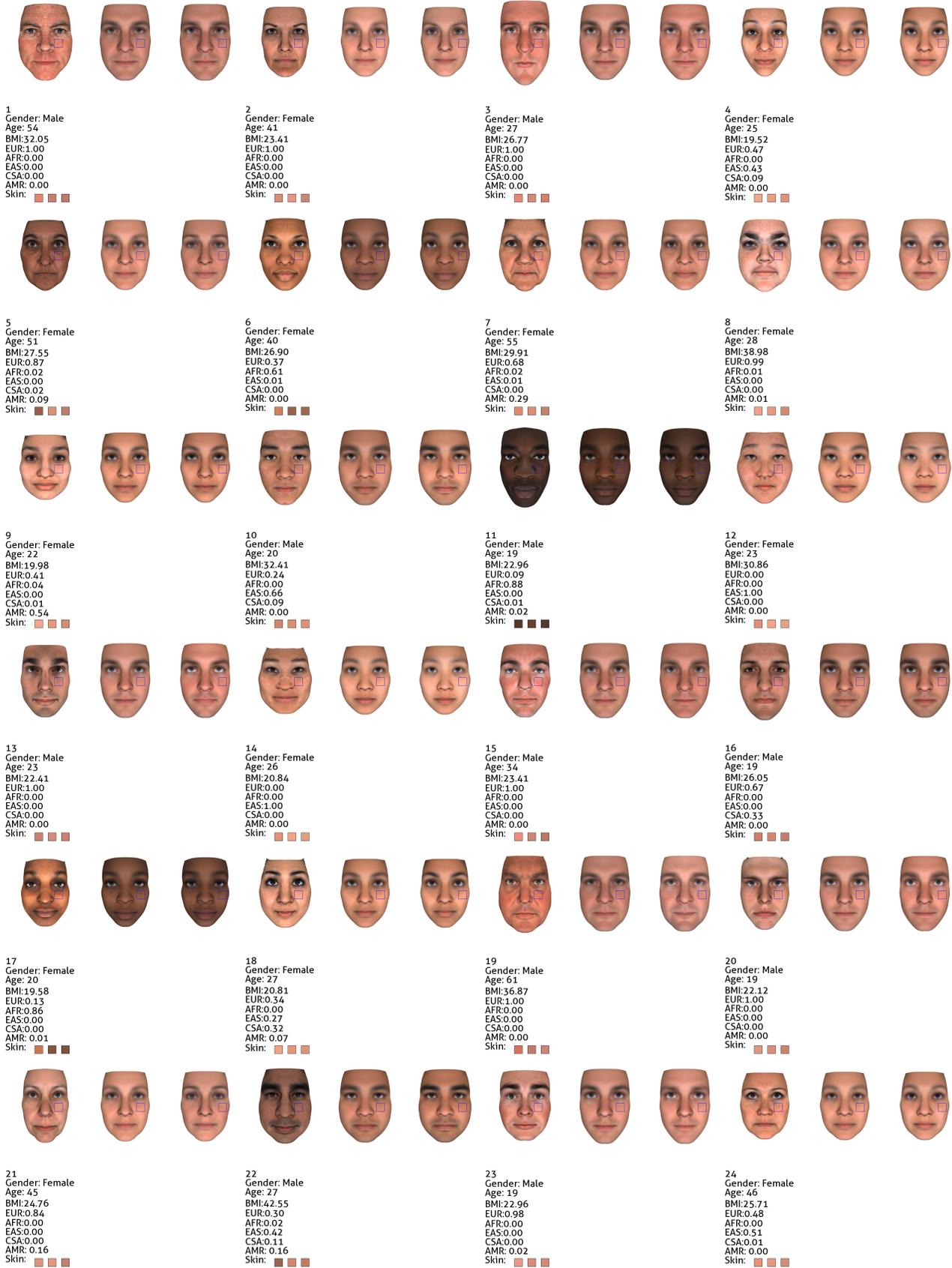
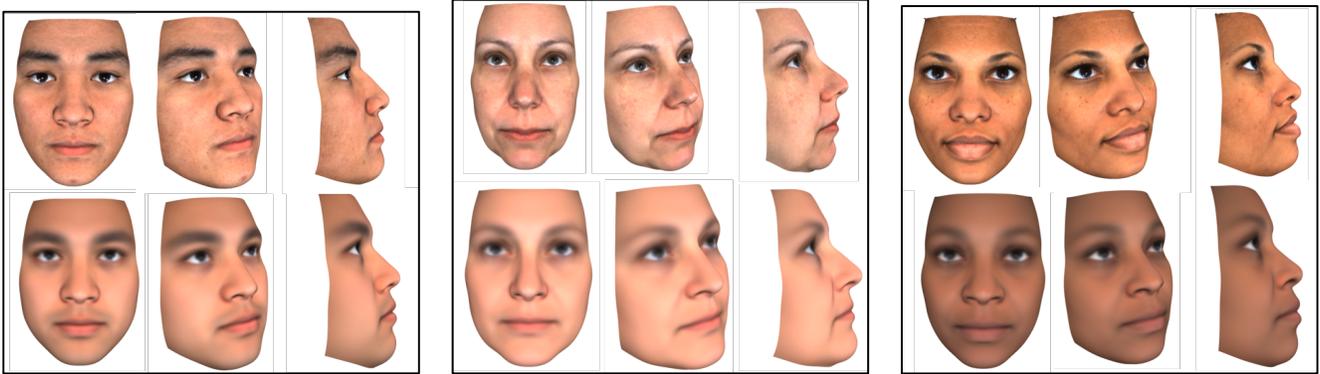
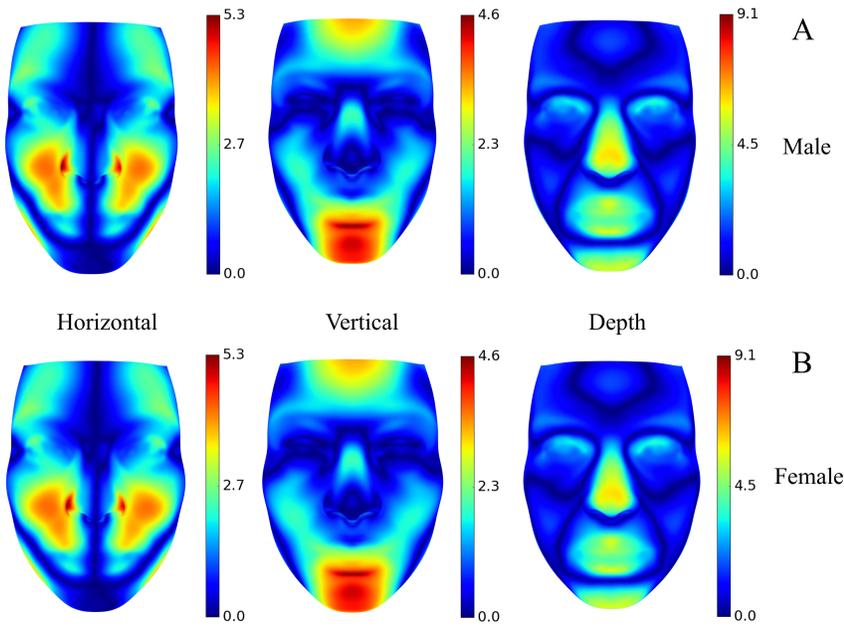


Fig. S11. Holdout set of 24 individuals who consented public view for their face image. Left most face: true face, middle face: Ridge regression predicted face, right most face: Ridge for Shape PCs,  $k$ -Nearest Neighbor for Color PCs.



**Fig. S12.** 3D scan vs. 3D prediction for three selected individuals from holdout set. Top row in each panel represents observed face (0 degrees, 45 degrees and 90 degrees rotated), and bottom row in each panel represents predicted face (0 degrees, 45 degrees and 90 degrees rotated).



**Fig. S13.** Difference between European and African face predictions. Absolute values of the differences (in *cm*) along the horizontal, vertical, and depth dimensions of predicted face shape between an individual of 100% European ancestry, and an individual of 100% African ancestry. Predictions were performed using a PLS model trained on five ancestry components (AFR, AMR, CSA, EAS, EUR) and sex. (A) Differences for male individuals (B) Differences for female individuals.

**Table S5. The ratio of standard deviation of African ancestry ( $\sigma_{AFR}$ ) to standard deviation of European ancestry ( $\sigma_{EUR}$ ) for ten face shape PCs.**

Predicted face shape PC	$\sigma_{AFR}/\sigma_{EUR}$
PC 5	2.80
PC 2	1.61
PC 8	1.57
PC 9	1.46
PC 10	1.27
PC 3	1.26
PC 7	1.22
PC 4	1.12
PC 1	1.01
PC 6	0.89

Among ten PCs, nine of the  $\sigma_{AFR}/\sigma_{EUR}$  are  $> 1.00$ , which indicates a larger facial variability in African ancestry than European ancestry.

Quantile (Q-Q) plots are shown in Fig. S14.

**Landmark Distance Prediction from Genome.** Researchers have studied landmark distances for various purposes including craniofacial anomaly detection and facial growth analysis, and have attempted to relate landmark distances to the genome (37, 40, 41). Paternoster et al. (41) have found an association between the nasion position and a SNP in the PAX3 gene on 2,185 adolescents, which has been replicated on another set of cohorts comprising 1,622 individuals. Furthermore, genome-wide association studies (GWAS) have identified five candidate genes affecting normal facial shape variation in landmark distances for Europeans, PRDM16, PAX3, TP63, C5orf50, and COL17A1 combined 12 SNPs, were identified as genome-wide significant (41, 42). However, the SNP explains only 1.3% of the variance of nasion position, and associations between diverse landmark distances and genome are largely unknown.

We have evaluated the performance of prediction of 26 landmark distances from genomic information, predicted sex from the genome, predicted age from the genome, and the top 1,000 genome PCs, in terms of  $R^2_{CV}$  between observed and predicted landmark distances. In Fig. S15, ALL\_ALR (width of nose) and LS\_LI (height of lip) are the most predictable, while TGL\_TGR (straight distance from left ear to right ear) and PSL\_PIL/PSR\_PIR (height of the left/right eye) are the least predictable. The results agree with our observation that the width of the nose and the height of the lip are excellent features to distinguish between different ethnicities (compare also Fig. S13). However, the length of the nose and the height of the eyes vary greatly within ethnicities. Thus, it is difficult to predict them from genome given our limited sample size (Fig. S15).

**Age Prediction from the Genome.** Age is a critical phenotypic trait for forensic identification. Accurate genomic prediction of age is especially important in our context, as age was used as a covariate for the prediction of other phenotypes. The maximum depth of the tree and the minimum number of individuals per leaf were tuned by cross-validation (CV) within each training fold. Since we aim to evaluate this model for forensic casework using only genomic information, we substituted genome predicted age for actual age in every applicable phenotype model.

**Table S6. Prediction quality for age.**

Covariate Set	$R^2_{CV}$
Telomere length (TL)	0.29
Chr[X Y] copy number	0.31
TL + Chr[X Y] copy number	0.44

Ridge regression was used with different combinations of covariates.

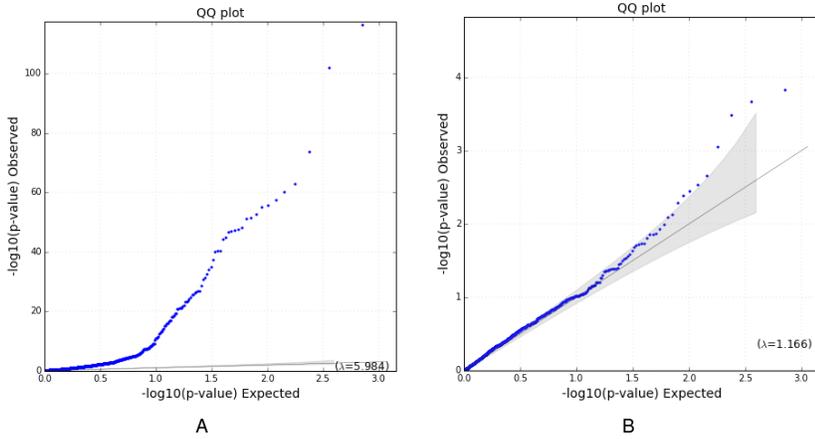
To predict age from the genome, we fit a random forest regression model that used a person’s average telomere length estimate and estimates of chromosome X and Y copy numbers as covariates for predicting age. During training, we removed individuals that were considered outliers. For our purposes, an outlier was defined as any male individual with an estimated Y copy number below 0.95 or above 1.05 or any female individual with an estimated X copy number below 1.95 or above 2.05.

Table S6 shows the held out variance explained ( $R^2_{CV}$ ) for the regression against age of telomere length and chr[X|Y] CCN, both individually and combined. Fig. S16D shows the predicted versus expected age for all our individuals using our final model.

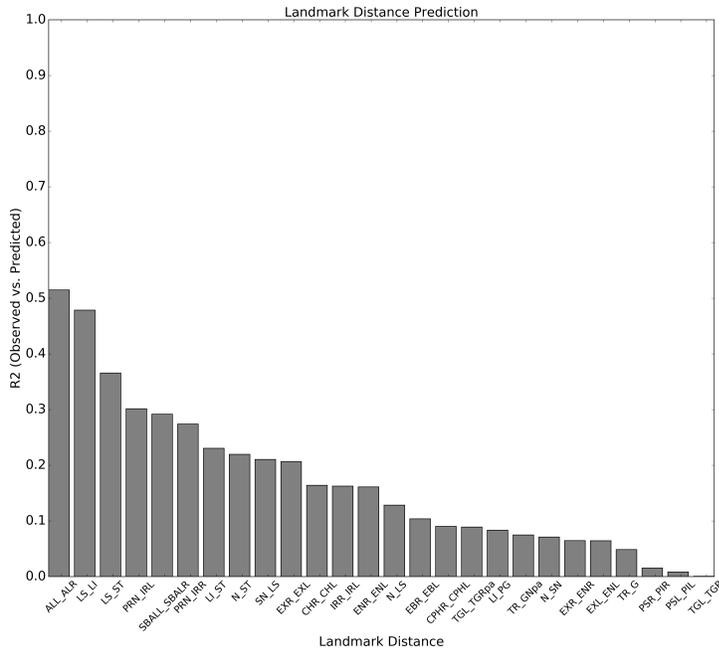
**Continuous Eye Color Prediction from the Genome.** For eye color prediction, we divided our experiments into two separate analyses: 0/1/2 SNP encoding and two-variable SNP encoding using the ridge regression model based on different covariates. First, we applied conventional SNP encoding of the minor allele dosage as 0/1/2. However, some variants associated with eye color exhibit significant dominance effects (43). If a set of SNPs has dominance effects on eye color, the prediction was improved when we modeled the SNPs with two different features: one representing the heterozygous SNP and another representing the homozygous alternate. This model is known as the two-variable SNP encoding. We observed that two-variable SNP encoding representations improve the prediction accuracy (Table S7).

We built three independent prediction models for the red (R), green (G), and blue (B) channels from the RGB color space for the two different encodings. Table S7 shows our prediction accuracy results for each R, G, and B with different covariates, including literature reported SNPs for eye color (Table S8). We also performed a GWAS to discover additional significantly associated variants beyond these published results. We did not test for additional variants other than those previously reported.

We initially considered age, sex, genomic PCs, and SNPs as predictive features in our model. While a correlation between age and eye color has been found for younger subjects in a specific population (46), our study includes only subjects  $\geq 18$  years of age, and we did not find that age was a significant determinate. Thus, we dropped age as a feature from our model. Since eye color clearly varies between different ethnic groups, we included 1,000 genomic PCs in our prediction model as covariates to capture ancestry differences and genome-wide SNP variation. The “Self-reported eye color” covariate represents the average eye color for each self-reported eye color category. Our results suggest that eye color is more accurately predicted from genomic data than by asking people to report their own eye color (see Table S7). This difference in accuracy may be explained by the fact that predictions are performed on an expressive continuous scale, compared to the limited



**Fig. S14.** Quantile-quantile (Q-Q) plots for association tests between all tests of 36 candidate SNPs vs. top 10 PCs for face color data and top 10 PCs for face shape data. (A) Association statistics are computed using age, sex and BMI as covariates, and (B) Association statistics are computed using age, sex, BMI, and five ancestry proportions (AFR, EUR, EAS, CSA, AMR) as covariates. Comparison of these QQ plots shows that these 36 previously identified SNPs are highly correlated with ancestry.

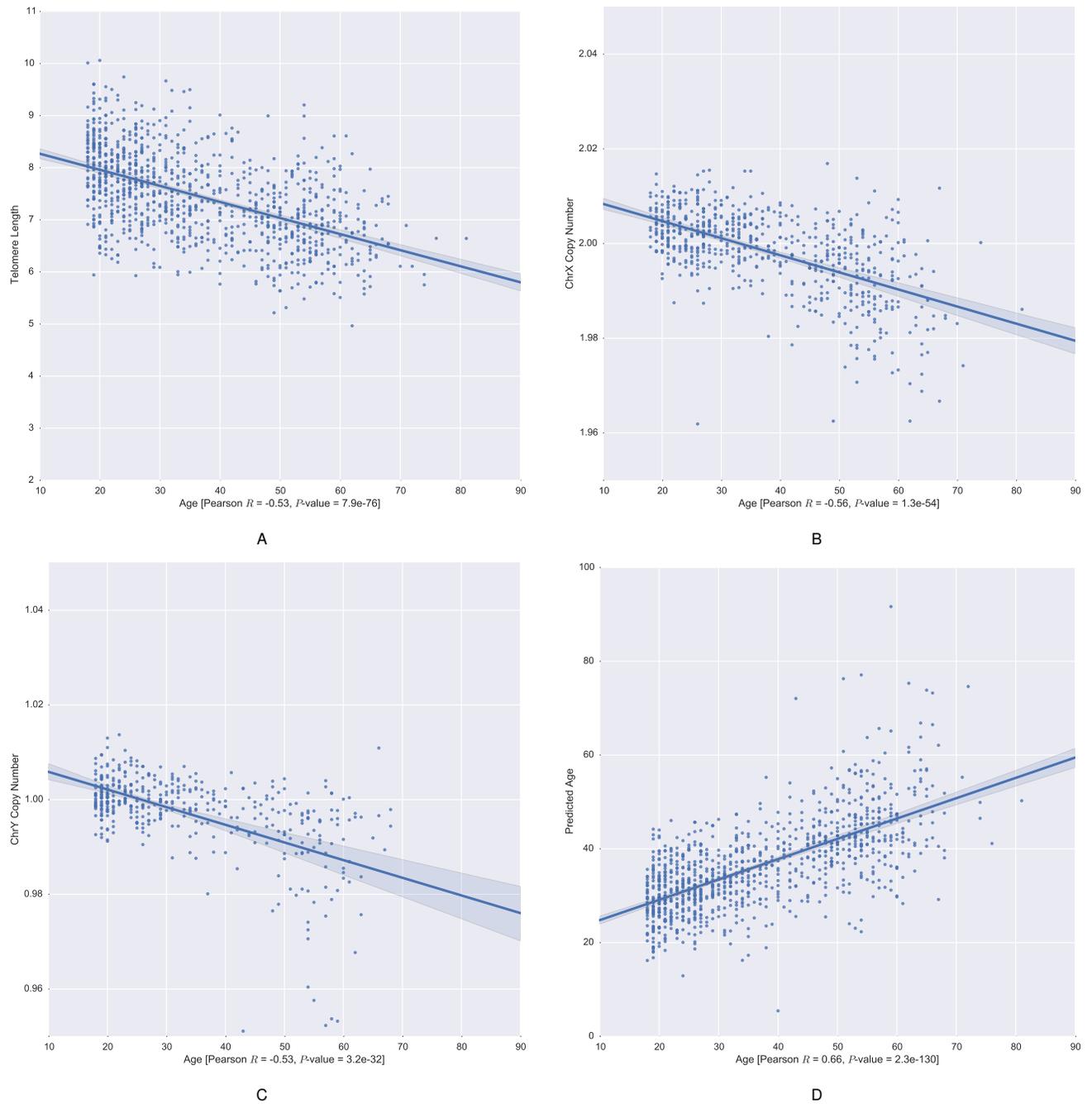


**Fig. S15.** Evaluation of landmark distance predictions. The measured performance in  $R^2_{CV}$  (observed vs. predicted) of predicted landmark distances using sex, predicted age, and top 1,000 genome PCs. ALL\_ALR (the width of the flaring of the nostril) is the highest performing landmark in our study. See Table S2 for the descriptions of each of these distances.

**Table S7. Prediction quality ( $R^2_{CV}$ ) for RGB Eye Color in ridge regression.**

Covariate Set	0/1/2 SNP encoding			Two-variable SNP encoding		
	R	G	B	R	G	B
Age	0.04	0.03	0.02	0.04	0.03	0.02
Sex	0.02	0.02	0.01	0.02	0.02	0.01
Self-reported eye color	0.55	0.72	0.76	0.55	0.72	0.76
Ancestry (1,000 PCs)	0.74	0.67	0.58	0.73	0.67	0.58
5 SNPs (44)	0.73	0.73	0.67	0.74	0.79	0.78
21 SNPs (45)	0.74	0.74	0.67	0.75	0.79	0.78
List A (65 SNPs)	0.63	0.56	0.49	0.63	0.58	0.52
List B (98 SNPs)	0.76	0.74	0.67	0.77	0.79	0.77
List C (241 SNPs)	0.74	0.69	0.60	0.71	0.74	0.70
1,000 genomic PCs, 5 SNPs, 9 interactions (44)	0.79	0.79	0.74	0.80	0.82	0.80

Using the two-variable SNP encoding and 1,000 genomic PCs, five known eye color SNPs, and nine interactions was the best model for eye color prediction in the RGB space. The lists of the SNPs used in these models are provided in Table S8.



**Fig. S16.** Correlation plot between true age and other variables including telomere length, chromosome X copy number, chromosome Y copy number, and predicted age. (A to C): Regression plots for telomere length and X or Y chromosomal copy number against age; (D): Held out predictions vs real age for all individuals.

**Table S8. Lists of literature reported SNPs for Eye Color that have been tested in the prediction models.**

List A	rs10765198 rs10852218 rs11074304 rs11568820 rs11572177 rs11631195 rs11636232 rs12324648 rs12520016 rs12592307 rs1375164 rs1448481 rs1448490 rs1470608 rs1498509 rs1498519 rs1498521 rs1562592 rs1603784 rs17084733 rs17673969 rs17674017 rs1800404 rs1800410 rs1800411 rs1800416 rs1800419 rs1874835 rs1973448 rs2015343 rs2254913 rs2290100 rs2311843 rs2594902 rs2594938 rs2681092 rs2689229 rs2689230 rs2689234 rs2703922 rs2703969 rs2871875 rs3002288 rs3782974 rs4253231 rs4278697 rs4778137 rs4778177 rs4778185 rs4778190 rs4778220 rs4810147 rs6785780 rs7170989 rs7173419 rs7175046 rs7176632 rs7176759 rs728404 rs7643410 rs7975232 rs9476886 rs9584233 rs977588 rs977589
List B	rs1042602 rs10765198 rs10852218 rs11074304 rs1126809 rs1129038 rs11568820 rs11572177 rs11631195 rs11636232 rs12203592 rs12324648 rs12520016 rs12592307 rs12896399 rs12913832 rs1375164 rs1393350 rs1408799 rs1448481 rs1448485 rs1448490 rs1470608 rs1498509 rs1498519 rs1498521 rs1540771 rs1562592 rs1597196 rs1603784 rs1667394 rs16891982 rs17084733 rs17673969 rs17674017 rs1800401 rs1800404 rs1800407 rs1800410 rs1800411 rs1800414 rs1800416 rs1800419 rs1805005 rs1874835 rs1973448 rs2015343 rs2238289 rs2254913 rs2290100 rs2311843 rs2594902 rs2594938 rs26722 rs2681092 rs2689229 rs2689230 rs2689234 rs2703922 rs2703969 rs2733832 rs2871875 rs3002288 rs3782974 rs3794604 rs4253231 rs4278697 rs4778137 rs4778138 rs4778177 rs4778185 rs4778190 rs4778220 rs4778232 rs4778241 rs4810147 rs6058017 rs6785780 rs683 rs7170852 rs7170989 rs7173419 rs7174027 rs7175046 rs7176632 rs7176759 rs7179994 rs7183877 rs728404 rs7495174 rs7643410 rs7975232 rs8024968 rs916977 rs9476886 rs9584233 rs977588 rs977589
List C	rs10001971 rs10007810 rs1003719 rs10108270 rs1015362 rs10209564 rs10235789 rs10236187 rs1040045 rs1040404 rs1042602 rs10496971 rs10510228 rs10511828 rs10512572 rs10513300 rs1074265 rs10839880 rs10954737 rs1105879 rs1110400 rs11164669 rs11227699 rs1126809 rs1129038 rs11547464 rs11631797 rs11652805 rs12130799 rs12203592 rs12439433 rs12452184 rs12544346 rs12592730 rs12593929 rs12629908 rs12657828 rs12821256 rs1289399 rs12896399 rs12906280 rs12913823 rs12913832 rs1296819 rs1325127 rs1325502 rs13267109 rs13400937 rs1357582 rs1369093 rs1393350 rs1407434 rs1408799 rs1408801 rs1426654 rs143384 rs1448485 rs1471939 rs1500127 rs1503767 rs1510521 rs1513056 rs1513181 rs1533995 rs1540771 rs1569175 rs1597196 rs1635168 rs 1667394 rs16891982 rs16950979 rs16950987 rs1760921 rs17793678 rs1800401 rs1800407 rs1800414 rs1805005 rs1805006 rs1805007 rs1805008 rs1805009 rs1837606 rs1871428 rs1879488 rs192655 rs1950993 rs199501 rs2001907 rs200354 rs2030763 rs2033111 rs2069398 rs2070586 rs2070959 rs2073730 rs2073821 rs2125345 rs214678 rs2228479 rs2238289 rs2240202 rs2240203 rs2252893 rs2269793 rs2277054 rs2278202 rs2306040 rs2330442 rs2346050 rs2357442 rs2397060 rs2416791 rs2424905 rs2424928 rs2504853 rs2532060 rs2594935 rs260690 rs2627037 rs26722 rs2702414 rs2709922 rs2724626 rs2733832 rs2835370 rs2835621 rs2835630 rs2899826 rs2946788 rs2966849 rs2986742 rs3118378 rs316598 rs316873 rs32314 rs35264875 rs35414 rs37369 rs3737576 rs3739070 rs3745099 rs3768056 rs3784230 rs3793451 rs3793791 rs3794604 rs3822601 rs3829241 rs385194 rs3935591 rs3940272 rs3943253 rs4458655 rs4463276 rs4530349 rs4666200 rs4670767 rs4673339 rs471360 rs4738909 rs4746136 rs4778138 rs4778232 rs4778241 rs4781011 rs4798812 rs4800105 rs4821004 rs4880436 rs4891825 rs4900109 rs4908343 rs4911414 rs4911442 rs4918842 rs4925108 rs4951629 rs4955316 rs4984913 rs507217 rs5768007 rs6058017 rs6104567 rs6422347 rs642742 rs6451722 rs6464211 rs647325 rs6493315 rs6541030 rs6548616 rs6556352 rs6759018 rs683 rs7029814 rs705308 rs7170852 rs7174027 rs7179994 rs7183877 rs7219915 rs7238445 rs7277820 rs728405 rs731257 rs734873 rs7421394 rs7495174 rs7554936 rs7657799 rs772262 rs7745461 rs7803075 rs7844723 rs798443 rs7997709 rs8021730 rs8024968 rs8028689 rs8035124 rs8041209 rs8113143 rs818386 rs870347 rs874299 rs881728 rs885479 rs892839 rs916977 rs9291090 rs9319336 rs946918 rs948028 rs9522149 rs9530435 rs9782955 rs9809104 rs9845457 rs9894429 rs989869

List A, List B, and List C of the SNPs are mentioned in Table S7.

expressiveness of discrete self-reported values.

Previous research found a set of genetic variants associated with eye color. For example, Mushailov et al. (44) identified five SNPs and Walsh et al. (45) identified 21 SNPs significantly associated with eye color. We identified 65 SNPs in the literature that produced fair predictions (see List A; they include all of the SNPs in List B minus the five SNPs of Mushailov et al. (44) and overlapping SNPs in List C); 98 SNPs that produced good results (see List B); and 241 SNPs that produced good predictions (see List C) (Table S8).

Mushailov et al. (44) found that there exist significant interactions between variants associated with eye color. For the combination of covariates which produced the best prediction, we included 1,000 genomic PCs, five eye color associated SNPs (rs12896399, rs6119471, rs16891982, rs12913832, and rs12203592) (44), nine specific interactions among these five SNPs as reported by Mushailov et al. (44), and excluded age and sex.

**Skin Color Prediction from Genome.** Skin pigmentation varies with latitude, suggesting that skin color variation is likely driven by natural selection in response to UV radiation levels (47). While the principal genes influencing eye and hair color are now largely identified, our understanding of the genetics of skin color variation is still far from complete (42), especially since fair skin color of European and East Asian populations seem to have arose independently (48–50).

In GWAS and other analyses a number of distinct genes were implicated in skin color variation, including: the MC1R, its inhibitor ASIP, OCA2, HERC2, ALC45A2, SLC24A5, and IRF4 (51–56). A number of skin color prediction models were built using different subsets of SNPs, including: a six-SNP model (57), a seven-SNP model (58) and a ten-SNP model (59). However, all of the predictive models used discrete qualitative phenotypes (skin color binned as “light”, “medium” and “dark” or some variation thereof) and the number of individuals and their ancestry variation was limited. In addition, the applicability of some of the models was limited to homozygous genotypes, whereas heterozygous genotypes were not considered at all (57, 58). Here, we sought to determine genetic features predictive for skin color across ancestries.

For the skin color prediction model, we included age and sex (both predicted from the genome), the first 1,000 PCs, which capture the ancestry information, and seven previously identified SNPs (rs12913832, rs1545397, rs16891982, rs1426654, rs885479, rs6119471, rs12203592) (58) as covariates. Unlike the model by Spichenok et al. (58), seven SNPs used in the skin color prediction model are encoded as minor allele counts instead of homozygous allele representation, as in preliminary experiments the latter representation yielded a 0.02 reduction in  $R_{CV}^2$ ; these SNPs along with their annotation are listed in Table S9.

We mainly compared two prediction approaches: ridge regression and extreme gradient boosting. Table S10 shows the results from extreme gradient boosting (XGBoost), which outperformed ridge regression. The number of estimators (`n_estimators`), maximum depth of a tree (`max_depth`), subsample proportion of instances chosen to grow a tree (`subsample`) and step size shrinkage to prevent overfitting (`eta`) were tuned using CV; the best performance was obtained when parameters were set to `n_estimators=1000`, `max_depth=2`, `subsample=0.9`, and `eta=0.01`.

While reported SNPs themselves are highly predictive of skin color, their independent contribution on top of a baseline of 1,000 genomic PCs is still marginal even in our best performing model ( $\approx 1$  to 3%) and most skin color variation was captured by the first three genomic PCs. True versus predicted skin color for 1,022 participants are given in Fig. S17.

**Height/BMI/Weight Prediction from the Genome.** We included age, sex, the first 1,000 genomic PCs, and associated SNPs from other studies in our height prediction model. We used 696 SNPs that previously had been associated to height on a large-scale GWAS meta-analysis (60) to build a height prediction model (we excluded one SNP rs2735469 among 697 previously identified SNPs in Wood et al. (60) since it did not pass our MAF threshold of 0.1% in our data set). For the BMI prediction model, we included 96 previously identified as BMI-associated SNPs (61) (we excluded one SNP rs12016871 among the reported SNPs by Locke et al. (61) because its MAF < 0.1%). For the weight prediction model, we used both the height-associated 696 SNPs and BMI-associated 96 SNPs. We used self-reported age and predicted sex from the genome as covariates.

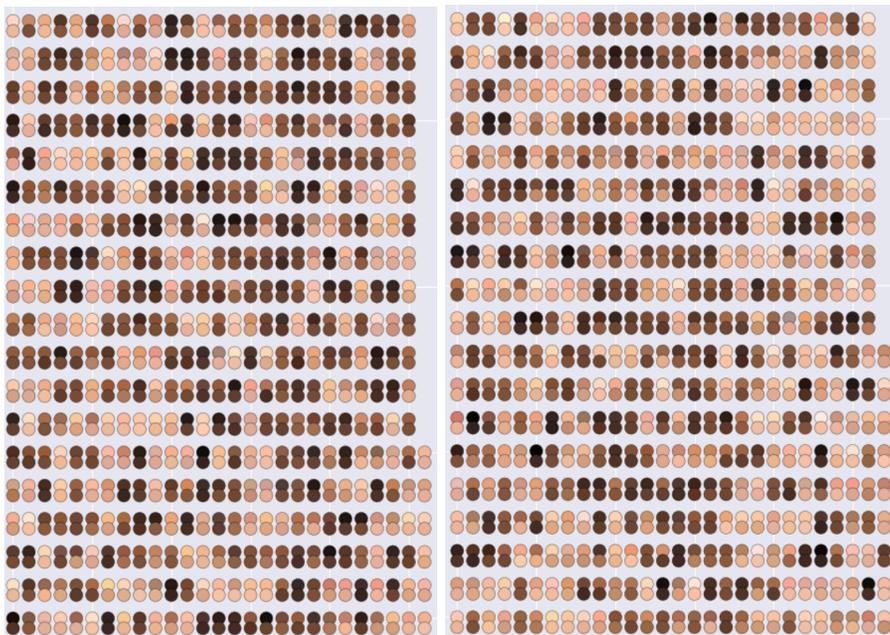
The underlying effect size of each of the selected SNPs for height/BMI/weight was expected to be small (62) and it would be difficult to accurately estimate these effect sizes on our cohort. Thus, instead of estimating the effect size of 696 SNPs + 96 SNPs on individuals from our database, we used the previously estimated effect sizes from a large scale meta-analysis of 253,288 individuals of the GIANT consortium for height SNPs (60) and 339,224 individuals for BMI SNPs (61). Then, for height and BMI predictions, one aggregated feature was created for height and BMI which is the sum of 696 SNPs and 96 SNPs weighted by their effect sizes, respectively. The prediction performances are summarized in Tables S11 and S12.

Regarding height, Tables S11 and S12 and Fig. S18 show the  $R_{CV}^2$  and mean absolute error (*MAE*) between the observed and predicted heights by our model with different features. The prediction model including only age as a feature (Fig. S18A) has an *MAE* of 7.4cm and  $R_{CV}^2$  of 0.02 (males: *MAE* = 5.5cm,  $R_{CV}^2$  = -0.01, females: *MAE* = 5.3cm,  $R_{CV}^2$  = 0.01). The prediction model with age and sex (Fig. S18B) has an *MAE* of 5.4cm and  $R_{CV}^2$  of 0.44. The prediction model with age, sex and the first 1,000 genomic PCs (Fig. S18C) has an *MAE* of 5.1cm and  $R_{CV}^2$  of 0.50. When we replaced genomic PCs by 696 height-associated SNPs in the previous model (Fig. S18D), we achieved the best predictive model, having an *MAE* of 4.9cm and  $R_{CV}^2$  of 0.53. Adding both, SNPs and genomic PCs into the same model yielded an identical performance (Fig. S18E), indicating that the height-associated SNPs sufficiently tag relevant ancestry variation and genome-wide SNP variation that is represented by the genomic PCs. Corresponding predictions stratified by both sexes (leaving out the sex covariate) yield an *MAE* and an  $R_{CV}^2$  of 5.3cm and 0.07 in males, and 4.7cm and 0.20 in females, respectively (Fig. S19).

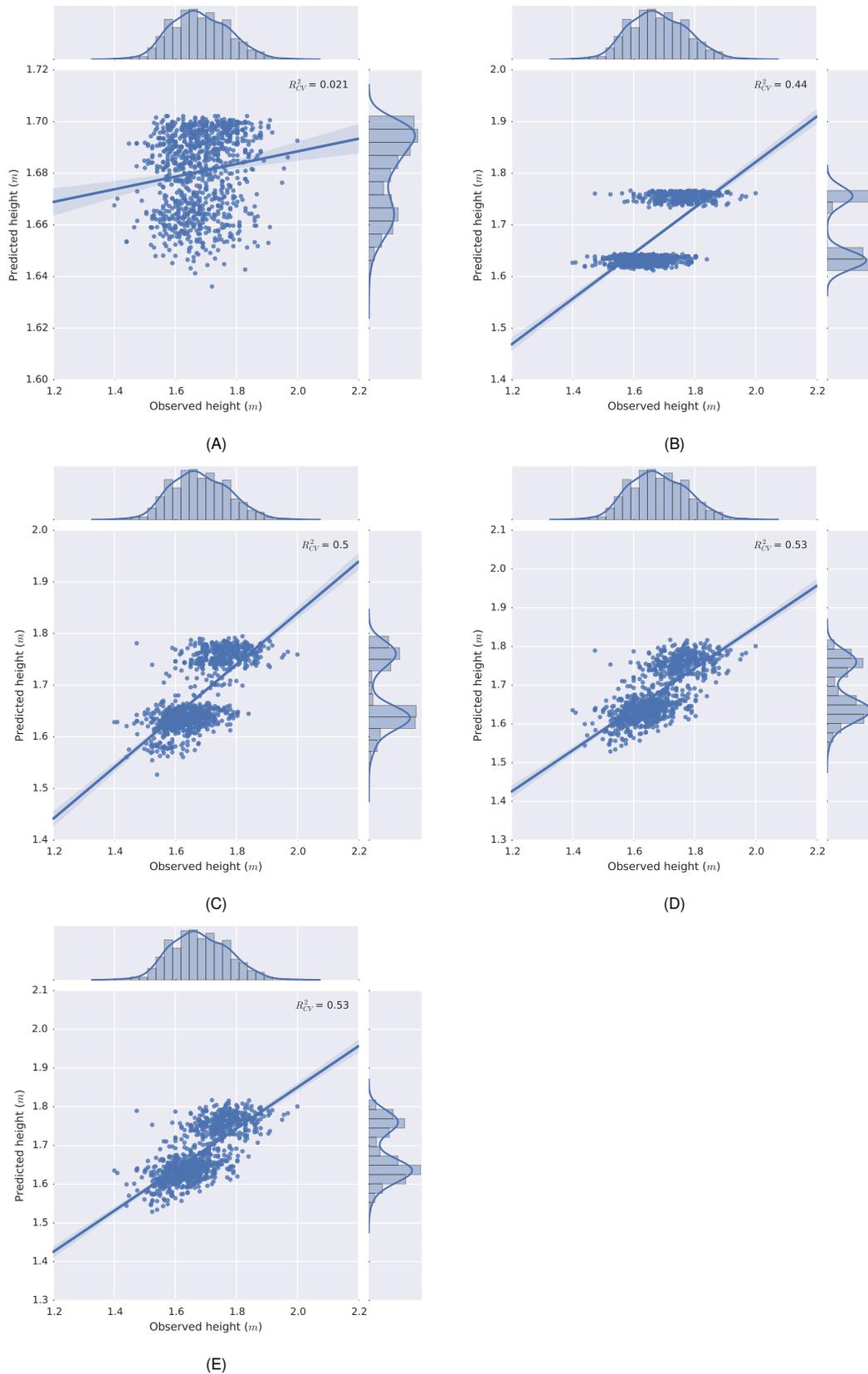
Regarding BMI, Tables S11 and S12, and Fig. S20 show the  $R_{CV}^2$  and *MAE* between the observed and predicted BMI by our model with different features. When the BMI predictive model includes only age as a feature (Fig. S20A), the *MAE* is 5.60kg/m<sup>2</sup> and  $R_{CV}^2$  of 0.07. The prediction model with age and sex (Fig. S20B) has an *MAE* of 5.53kg/m<sup>2</sup> and  $R_{CV}^2$

**Table S9. List of variants used for skin color prediction with annotations.**

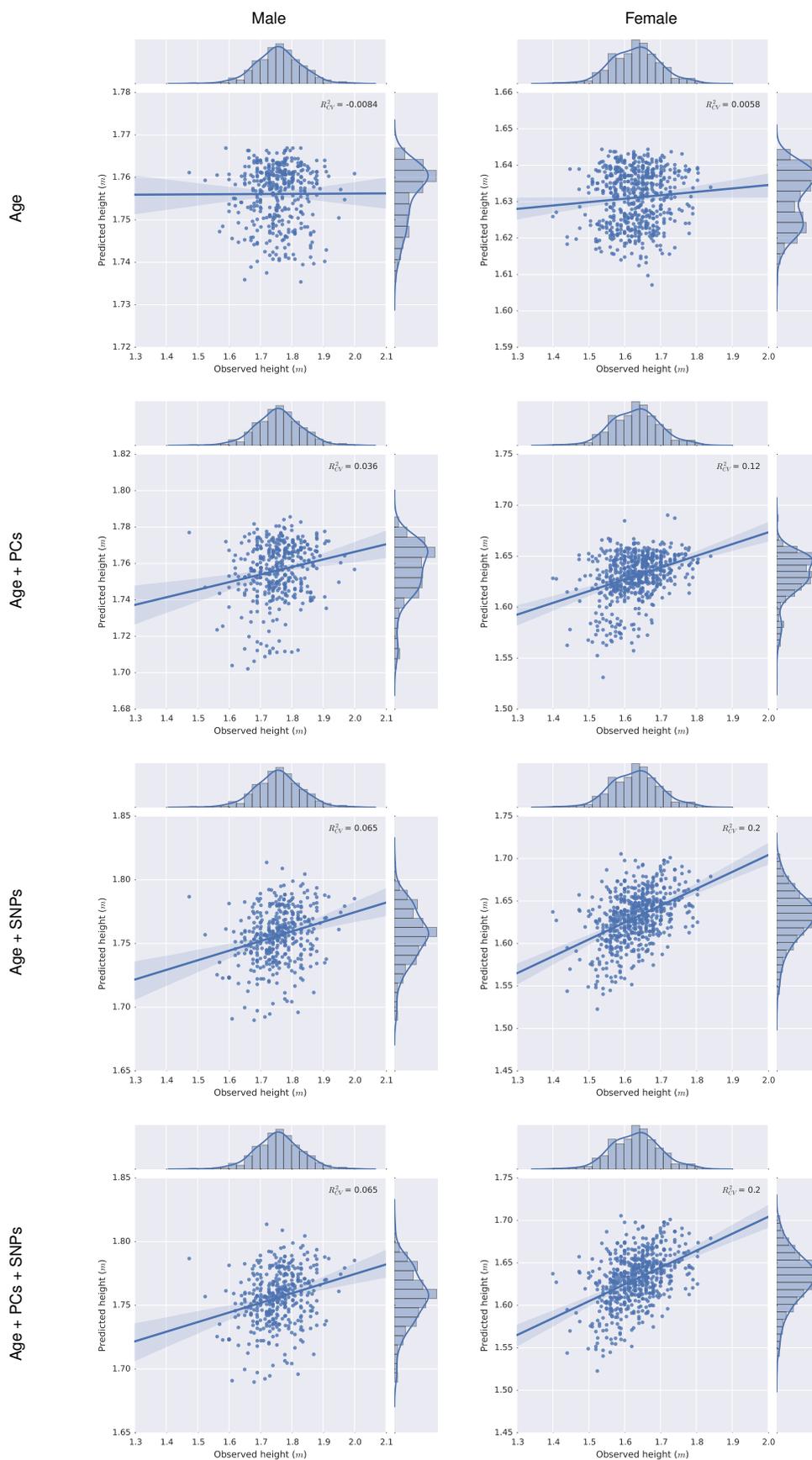
Gene	SNP ID	Variation Type	Annotation
HERC2	rs12913832	Predicted transcription factor binding site for OCA2	Blue/brown eye color (in European and East Asian population); Reduced melanin content in cultured human melanocytes.
OCA2	rs1545397	Intron	East Asian population.
SLC45A2	rs16891982	Missense	European population; Blue/brown eye color (in European population); Reduced melanin content in cultured human melanocytes.
SLC24A5	rs1426654	Missense	European population; Reduced melanin content in cultured human melanocytes.
MC1R	rs885479	Missense	East Asian population.
ASIP	rs6119471	Near 5'-end, predicted transcription factor binding site	African population.
IRF4	rs12203592	Intron	Blue/brown eye color (in European population).



**Fig. S17.** Observed and predicted skin colors for 1,022 individuals. The top circle shows observed, the bottom circle shows predicted skin color from an Extreme Boosted Tree model with 1,000 PCs, predicted age, predicted sex and seven SNPs as covariates.



**Fig. S18.** Scatter plots between predicted height and observed height with different features in ten-fold CV. (A) Age; (B) Age + Sex; (C) Age + Sex + 1,000 genomic PCs; (D) Age + Sex + height SNPs (696 height associated SNPs); (E) Age + Sex + 1,000 genomic PCs + height SNPs (696 height associated SNPs).



**Fig. S19.** Height prediction with different covariates stratified by sex. Covariates are age, age + 1,000 genomic PCs, age + SNPs, and age + 1,000 genomic PCs + SNPs from top to bottom rows for male (left column) and female subjects (right column).

**Table S10. Regression results by extreme gradient boosting on skin color using different covariate sets.**

Covariate Set	R	G	B
Age	0.13	0.15	0.14
Sex	-0.01	-0.01	-0.01
PCs	0.71	0.76	0.78
Age + Sex + PCs	0.76	0.79	0.80
Age + Sex + PCs + SNPs	0.77	0.79	0.81
SNPs	0.68	0.71	0.75

Values are  $R_{CV}^2$ . Phenotypes are represented in RGB space. Covariates are different combinations of age, sex, top 1,000 genomic PCs and seven SNPs listed in Table S9.

of 0.09. The prediction model with age, sex and first 1,000 genomic PCs (Fig. S20C) has an  $MAE$  of  $5.30kg/m^2$  and  $R_{CV}^2$  of 0.17. When we also add 96 BMI associated SNPs to the above model, we achieved a model of equivalent quality in terms of  $MAE$ . Prediction models with age, sex, 96 BMI-associated SNPs, without (Fig. S20D) or with (Fig. S20E) genomic PCs, also achieved an  $MAE$  of  $5.30kg/m^2$  and  $R_{CV}^2$  of 0.17, indicating that on our data SNPs do not add relevant independent genomic variation to a model containing PCs or vice versa. Corresponding predictions stratified by both sexes (leaving out the sex covariate) yield  $MAE$  and  $R_{CV}^2$  of  $3.94kg/m^2$  and 0.11 in males, and  $6.15kg/m^2$  and 0.15 in females, respectively (Fig. S21).

Regarding weight, Tables S11 and S12, and Fig. S22 show the  $R_{CV}^2$  and  $MAE$  between the observed and predicted weight by our model with different features. The prediction model with only age as a feature (Fig. S22A) has an  $MAE$  of  $16.57kg$  and  $R_{CV}^2$  of 0.04. The prediction model with age and sex (Fig. S22B) has an  $MAE$  of  $16.40kg$  and  $R_{CV}^2$  of 0.05. The prediction model with age, sex, and the first 1,000 genomic PCs (Fig. S22C) has an  $MAE$  of  $15.54kg$  and  $R_{CV}^2$  of 0.15. Replacing genomic PCs with 696 height-associated SNPs and 96 BMI-associated SNPs (Fig. S22D) did not improve prediction accuracy and even slightly reduced prediction quality, with an  $MAE$  of  $15.59kg$  and an  $R_{CV}^2$  of 0.14. Re-adding genomic PCs left the results unchanged (Fig. S22E). These results reflect the fact that SNP effect sizes had been estimated on different phenotypes. Corresponding predictions stratified by both sexes (leaving out the sex covariate) yield  $MAE$  and  $R_{CV}^2$  of  $13.53kg$  and 0.10 in males, and  $16.77kg$  and 0.15 in females, respectively (Fig. S23).

**Sex Prediction from the Genome.** To predict sex from the genome, we first estimated the copy number for chromosome X ( $CCN\_chrX$ ) and Y ( $CCN\_chrY$ ) (see Materials and Methods). Males are expected to have one copy of chromosome X and one copy of chromosome Y and females are expected to have two copies of chromosome X. Fig. S24 shows the distributions for  $CCN\_chrX$  vs  $CCN\_chrY$  computed for all the individuals in our dataset. Sex chromosome copy numbers are predictive of sex, as can be predicted in Fig. S24. We performed rule-based sex prediction as follows: individuals with  $CCN\_chrY \leq 0.25$  were predicted as female, regardless of the value of  $CCN\_chrX$ . Individuals with  $CCN\_chrY > 0.25$  were predicted as male. Among male individuals in our dataset, we identified a putative case with XXY aneuploidy, also known as Klinefelter’s syndrome (63). This case was identified using

the following rule:  $1.5 < CCN\_chrX \leq 2.5$ . Note, that a single case is not sufficient to perform a rigorous statistical assessment of the proposed rule. As sex chromosome aneuploidy is expected to appear in the general population, the individual has been included in all further analysis. If necessary, these rules could be extended to address other cases of sex chromosome aneuploidy.

When predicting self-reported gender from sex, our chromosome copy number ( $CCN$ )-based rules achieved an accuracy of 99.6%. Four inconsistencies and two missing annotations were observed in 1,061 individuals. For the four errors, three female individuals were predicted as male and one male individual was predicted as female. A closer look at these cases indicated that for all of them the self-reported gender did in fact not reflect their sex. The individual with Klinefelter’s syndrome, karyotype 47, XXY, was annotated and predicted as male, as expected. Our sex prediction from  $CCN$  of the genome is highly accurate and could be used to identify gender mismatches.

**Predicting Age, Sex and Ancestry from the Face and Voice.**

To quantify how well face and voice capture information about age, sex and five regions of ancestry, we predicted these traits from observed face shape, face color, landmark distances, and voice  $i$ -vectors using ridge regression. As input features for prediction from face shape and color we used 1,000 of the corresponding PCs. As input features for prediction from voice, we used all 100 available  $i$ -vectors and voice pitch. Similarly, we used all landmark distances for prediction. This approach is helpful for extracting demographic information from face and voice where such information is useful but not otherwise accessible. In addition, it leads to higher select and performance compared to directly matching observed to predicted values for face and voice.

**Predicting Male Pattern Baldness from the Genome.**

We trained a model to predict Male Pattern Baldness (androgenic alopecia), which has an estimated heritability of 80% (64). We did not collect a baldness phenotype during the sample collection phase, so we assigned a phenotype by inspecting the facial images. For each image, we assigned a degree of baldness on a scale of one through seven using the Norwood-Hamilton scale for androgenic alopecia, with one representing no hair loss. Our data did not have many bald samples (see distribution of the phenotype in Fig. S25). Further, were not able to distinguish true baldness from cosmetic effects like head shaving based on the face images. We trained a model on this phenotype using SNPs associated with androgenic alopecia in GWAS (65–67). The most predictive covariate was the age of a person, achieving an  $R_{CV}^2$  of 0.29 in male individuals. We did not observe an improvement over age from using genetic information in the form of reported SNPs or genomic PCs. However, when using age predicted from the genome ( $R_{CV}^2$  of 0.05 in males), we did see an improvement from reported SNPs ( $R_{CV}^2$  of 0.09 in males) and PCs ( $R_{CV}^2$  of 0.08 in males) over predicted age alone. See full results in Table S13. Such low values of  $R_{CV}^2$  are expected to only yield marginal improvements in re-identification performance, as can be seen from Figs. S33 and S34.

**Hair Color Prediction from the Genome.** We were not able to generate usable models for hair color from this subject

**Table S11.  $R^2_{CV}$  for height, BMI, weight with different covariates in ridge regression.**

Covariate Set	Height ( $R^2_{CV}$ )			BMI ( $R^2_{CV}$ )			Weight ( $R^2_{CV}$ )		
	male	female	all	male	female	all	male	female	all
Age	-0.01	0.01	0.02	0.08	0.05	0.07	0.05	0.04	0.04
Age (+ Sex)	-0.01	0.01	0.44	0.08	0.05	0.09	0.05	0.04	0.05
Age + 1,000 PCs (+ Sex)	0.04	0.12	0.50	0.11	0.16	0.17	0.09	0.16	0.15
Age + 696 SNP Height (+ Sex)	0.07	0.20	0.53	–	–	–	0.10	0.15	0.15
Age + 96 SNP BMI (+ Sex)	–	–	–	0.11	0.15	0.17	0.09	0.15	0.14
Age + 1,000 PCs + 696 SNP Height (+ Sex)	0.07	0.20	0.53	–	–	–	0.10	0.15	0.15
Age + 1,000 PCs + 96 SNP BMI (+ Sex)	–	–	–	0.11	0.15	0.17	0.09	0.15	0.15
Age + 1,000 PCs + 696 SNP Height + 96 SNP BMI (+ Sex)	–	–	–	–	–	–	0.10	0.15	0.14

Sex has been used only in the analysis of “all” samples.

**Table S12. Mean absolute error for height, BMI, weight with different covariates in ridge regression.**

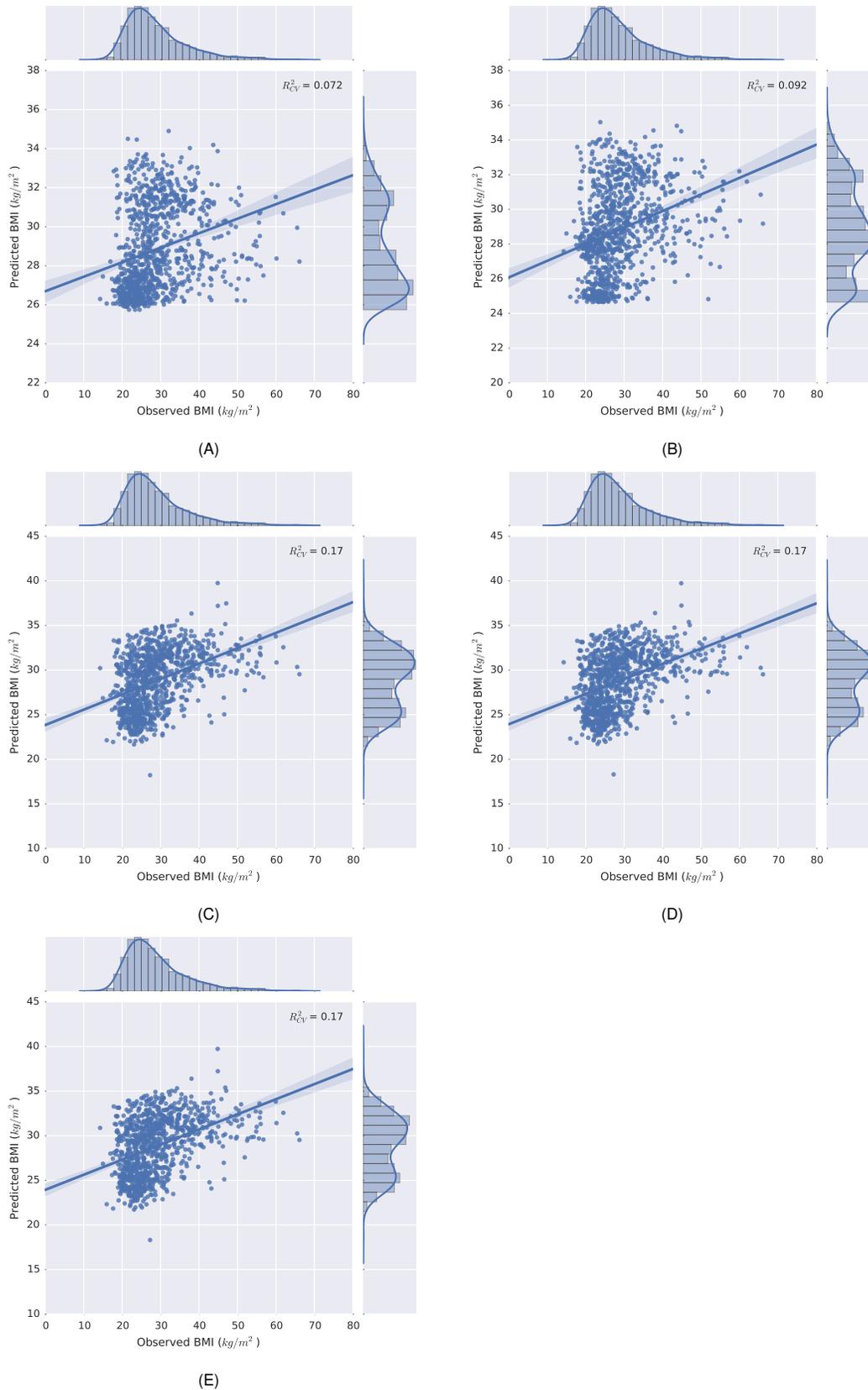
Covariate Set	Height ( $MAE$ , cm)			BMI ( $MAE$ , kg/m <sup>2</sup> )			Weight ( $MAE$ , kg)		
	male	female	all	male	female	all	male	female	all
Age	5.5	5.3	7.4	3.99	6.57	5.60	13.82	18.15	16.57
Age (+ Sex)	5.5	5.3	5.4	3.99	6.57	5.53	13.82	18.15	16.40
Age + 1,000 PCs (+ Sex)	5.4	4.9	5.1	3.92	6.13	5.30	13.57	16.68	15.54
Age + 696 SNP Height (+ Sex)	5.3	4.7	4.9	–	–	–	13.50	16.70	15.54
Age + 96 SNP BMI (+ Sex)	–	–	–	3.94	6.15	5.30	13.57	16.74	15.59
Age + 1,000 PCs + 696 SNP Height (+ Sex)	5.3	4.7	4.9	–	–	–	13.50	16.70	15.54
Age + 1,000 PCs + 96 SNP BMI (+ Sex)	–	–	–	3.94	6.15	5.30	13.57	16.74	15.54
Age + 1,000 PCs + 696 SNP Height + 96 SNP BMI (+ Sex)	–	–	–	–	–	–	13.53	16.77	15.59

Sex has been used only in the analysis of “all” samples.

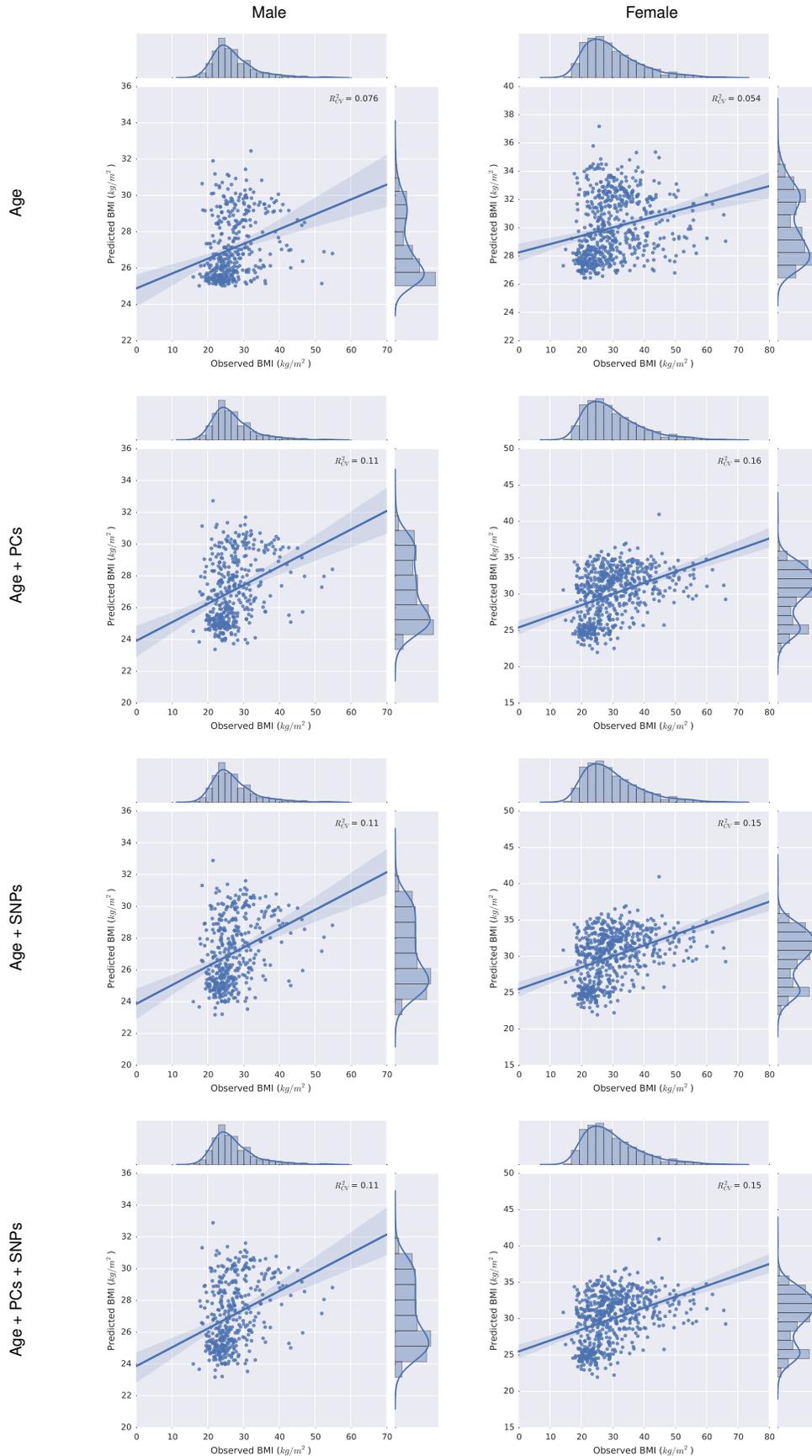
**Table S13. Regression results by ridge regression for male pattern baldness in male individuals.**

Covariate Set	Self-Reported Age		Predicted Age	
	$R^2_{CV}$	$MAE$	$R^2_{CV}$	$MAE$
SNPs	0.06	0.54	0.06	0.54
1,000 PCs	0.07	0.53	0.07	0.53
Age	0.29	0.47	0.05	0.57
Age + 1,000 PCs	0.28	0.47	0.08	0.53
Age + SNPs	0.29	0.47	0.09	0.53
Age + 1,000 PCs + SNPs	0.29	0.47	0.09	0.53

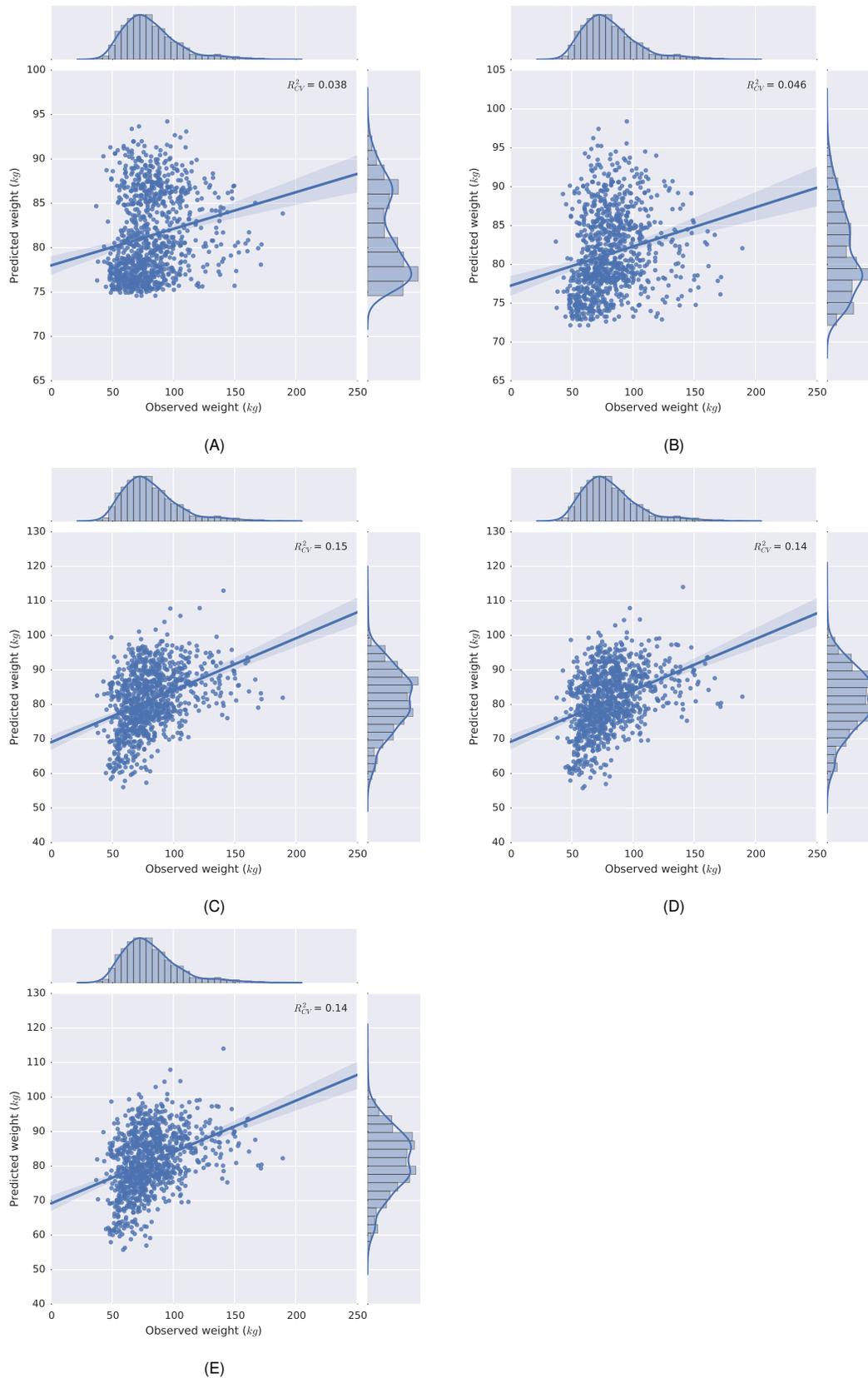
$R^2_{CV}$  and mean absolute errors ( $MAE$ ) for models using different covariate sets. PCs refer to genomic PCs, predicted age is age predicted from the genome.



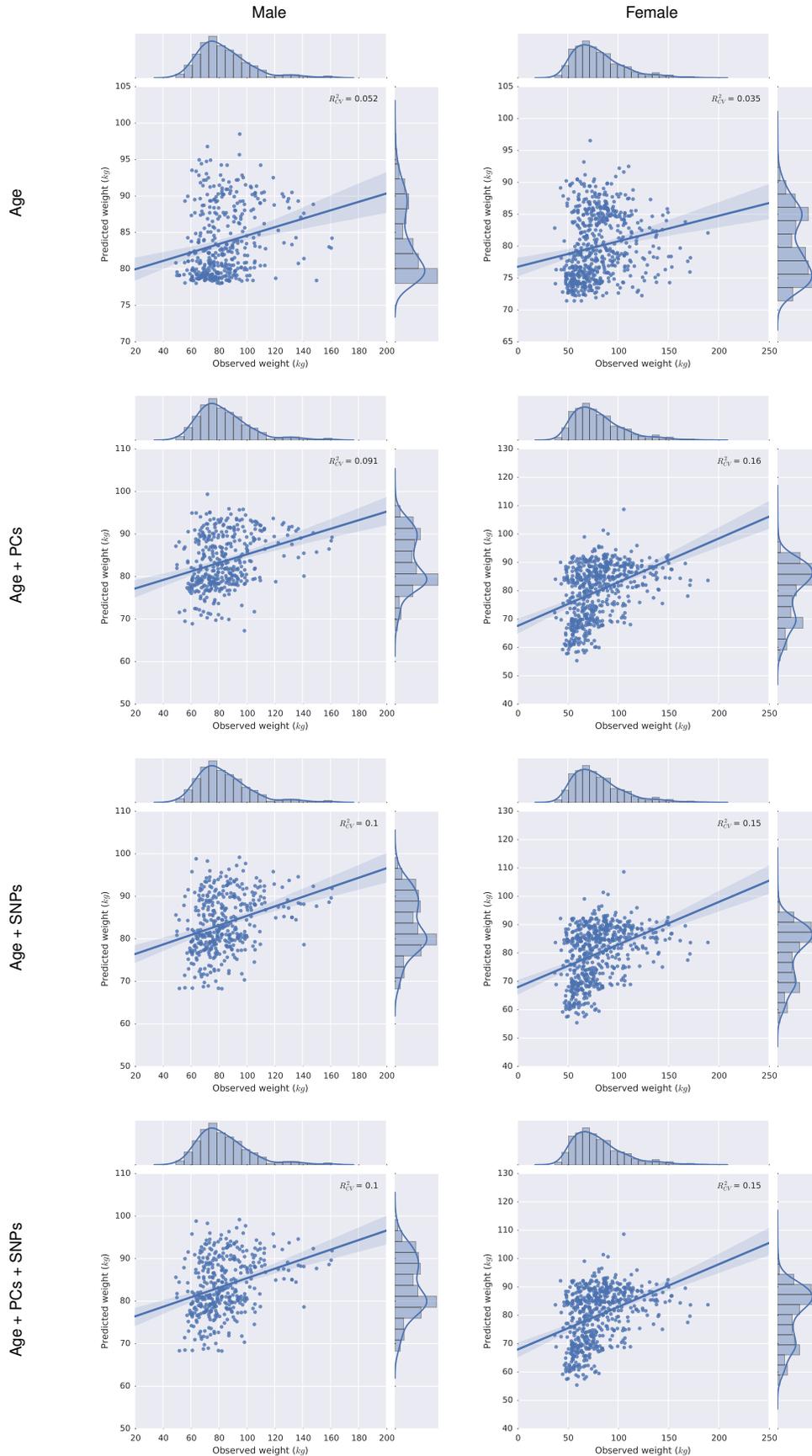
**Fig. S20.** Scatter plots between predicted BMI and observed BMI with different features in ten-fold CV with 4,082 individuals. (A) Age; (B) Age + Sex; (C) Age + Sex + 1,000 genomic PCs; (D) Age + Sex + height SNPs (96 height associated SNPs); (E) Age + Sex + 1,000 genomic PCs + 96 BMI SNPs.



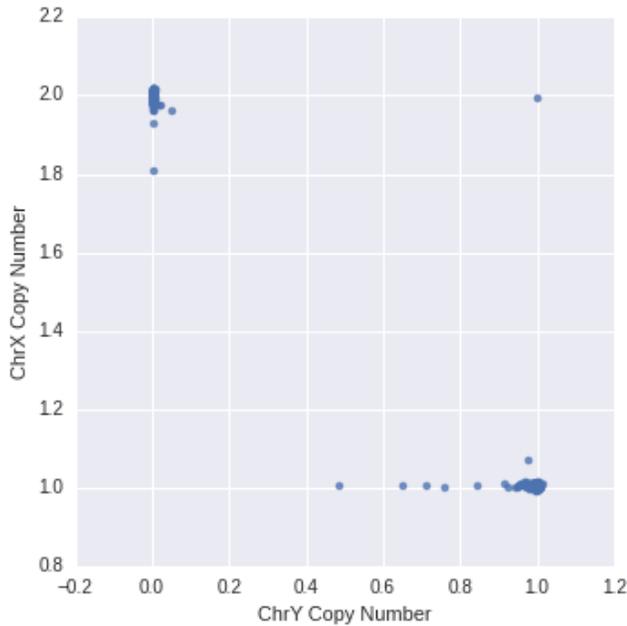
**Fig. S21.** BMI prediction with different covariates stratified by sex. Covariates are age, age + 1,000 genomic PCs, age + SNPs, and age + 1,000 genomic PCs + SNPs from top to bottom rows for male (left column) and female subjects (right column).



**Fig. S22.** Scatter plots between predicted weight and observed weight with different features. (A) Age; (B) Age + Sex; (C) Age + Sex + 1,000 genomic PCs; (D) Age + Sex + height SNPs (696 height associated SNPs) + 96 BMI SNPs; (E) Age + Sex + 1,000 genomic PCs + height SNPs (696 height associated SNPs) + 96 BMI SNPs.



**Fig. S23.** Weight prediction with different covariates stratified by sex. Covariates are age, age + 1,000 genomic PCs, age + SNPs, and age + 1,000 genomic PCs + SNPs from top to bottom rows for male (left column) and female subjects (right column).



**Fig. S24.** Distributions for chromosome X vs chromosome Y copy number estimates. We obtained the final rules for predicting sex, as detailed in the supplementary text, by inspecting this plot.

population. The population was not selected for hair color, and after eliminating individuals with over 50% African ancestry and subjects who did not have a usable hair color value due to their age, use of artificial hair color, or baldness we were left with 198 subjects, including 19 red heads and 20 blonds, which we found insufficient to reliably train predictive models of hair color.

**Metric Learning for Individual Identification.** To measure similarity between a de-identified genome  $g \in \mathcal{G}$  and a set of phenotypic measurements derived from an image and basic demographic information  $p \in \mathcal{P}$  for an individual whose identity is known, we propose the following two-step approach. First, we find a mapping of phenotypes,  $\psi_{\mathcal{P}} : \mathcal{P} \rightarrow \mathcal{E}_{\mathcal{P}}$ , and a mapping of genomes,  $\phi_{\mathcal{P}} : \mathcal{G} \rightarrow \mathcal{E}_{\mathcal{P}}$ , into a common  $D$ -dimensional embedding-space  $\mathcal{E}_{\mathcal{P}} \in \mathbb{R}^D$ . Finding such mappings consists of a combination of PC analysis and predictive modeling. The second step is to learn an optimal similarity  $\delta_{\mathcal{P}} : \mathcal{E}_{\mathcal{P}} \times \mathcal{E}_{\mathcal{P}} \rightarrow \mathbb{R}$  that allows numeric comparison of mapped phenotypes  $\psi_{\mathcal{P}}(p)$  and genomes  $\phi_{\mathcal{P}}(g)$ .

**Learning a Similarity Function.** For any given phenotypic measurement, we have defined suitable embeddings. Phenotypes that are a single number, such as height, weight, or age, are simply represented by their phenotype value. For phenotypes that comprise high-dimensional objects, such as images, or voice samples, we have defined embeddings to capture a maximum amount of information relevant for matching. For example, facial images provide information on the shape and the color of the face. Additionally, a facial image may also provide information about sex, ancestry, and the age of the person in the image. Consequently, we embedded images into a set of PC dimensions that capture shape and color information, and

additional dimensions for sex, ancestry, and age.

Given  $\psi_{\mathcal{P}} : \mathcal{P} \rightarrow \mathcal{E}_{\mathcal{P}}$ , the function for embedding a phenotype, and  $\phi_{\mathcal{P}} : \mathcal{G} \rightarrow \mathcal{E}_{\mathcal{P}}$ , the function for embedding a genome, we learn an optimal similarity  $\delta_{\mathcal{P}}$ , that takes embedded phenotype  $\psi_{\mathcal{P}}(p)$  and genotype  $\phi_{\mathcal{P}}(g)$  and outputs a similarity. A list of all embeddings is given in Table S14. As not all dimensions of the embedding space  $\mathcal{E}_{\mathcal{P}}$  can be expected to yield equal amounts of information for judging similarity between phenotypes and genomes, we learned optimally weighted similarity functions  $\delta_{\mathcal{P}}$  to improve re-identification.

$$\delta_{\mathcal{P}}(\psi_{\mathcal{P}}(p_n), \phi_{\mathcal{P}}(g)) = \sum_{d=1}^D w_d |\psi_{\mathcal{P}}(p_n)_d - \phi_{\mathcal{P}}(g)_d|, \quad [2]$$

where the weights  $w_d$  reflect the importance of  $d$ -th dimension of  $\mathcal{E}_{\mathcal{P}}$ .

To obtain on optimal  $\delta_{\mathcal{P}}$ , we define the probability that a genome  $g$  belongs to the same person as the  $n$ -th phenotype profile  $p \in \mathcal{S}$  out of a lineup  $\mathcal{S} = [p_1, \dots, p_N]$  of size  $N$ , to be by expression 3 for  $\eta(g, p_n, \mathcal{S})$ .

$$\eta(g, p_n, \mathcal{S}) = \frac{\exp(\delta_{\mathcal{P}}(\psi_{\mathcal{P}}(p_n), \phi_{\mathcal{P}}(g)))}{\sum_{p_{n'} \in \mathcal{S}} \exp(\delta_{\mathcal{P}}(\psi_{\mathcal{P}}(p_{n'}), \phi_{\mathcal{P}}(g)))}. \quad [3]$$

Plugging in the expression for  $\delta_{\mathcal{P}}$  given in expression 2, we obtain

$$\eta(g, p_n, \mathcal{S}) = \frac{\exp(\sum_{d=1}^D w_d |\psi_{\mathcal{P}}(p_n)_d - \phi_{\mathcal{P}}(g)_d|)}{\sum_{p_{n'} \in \mathcal{S}} \exp(\sum_{d=1}^D w_d |\psi_{\mathcal{P}}(p_{n'})_d - \phi_{\mathcal{P}}(g)_d|)}.$$

For a training data set  $\mathcal{D}$  consisting of combinations of lineups  $\mathcal{S}$  and corresponding genomes  $g$ , we can maximize the resulting log-likelihood  $\mathcal{L}$  with respect to the weights  $w_d$  using the YASMET (31) software\*.

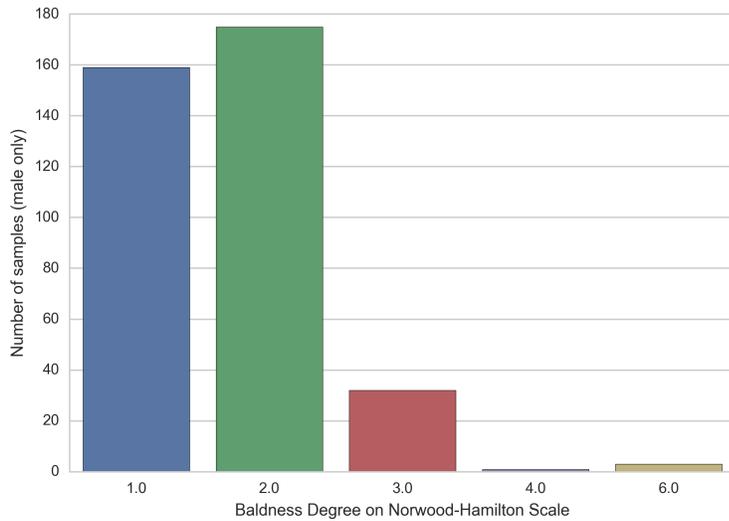
$$\mathcal{L} = \sum_{(g, \mathcal{S}) \in \mathcal{D}} \left( \sum_{p_n \in \mathcal{S}} \ln \eta(g, p_n, \mathcal{S}) \right).$$

In Fig. S28, we compare  $m_{10}$  and  $s_{10}$  using YASMET and the cosine distance for different combinations of phenotypes. Fig. S26 shows the corresponding receiver operator characteristic (ROC) curves, for all individuals and stratified by gender. For 25 out of 26 settings, YASMET showed better performance than cosine (one-sided binomial  $P = 4.0 \times 10^{-7}$ ), demonstrating that our proposed metric learning approach properly adjusted the weights to achieve high identification performance.

In Fig. S29 we show  $m_{10}$  and  $s_{10}$  using YASMET for individuals stratified by (A) male and (B) female sex. Fig. S32 shows  $m_{10}$  and  $s_{10}$  using YASMET for an analysis that included 277 SNPs reported by Qiao et al. (68). We did not observe an improvement in re-identification performance.

**Select Performance Simulation.** We simulated independent Gaussian distributed traits  $y_n$  for 1,000 individuals as the sum of a Gaussian distributed predictor  $p_n$  and an unpredictable Gaussian noise component  $\epsilon_n$ :

\* <http://www.fjoch.com/yasmet.html>

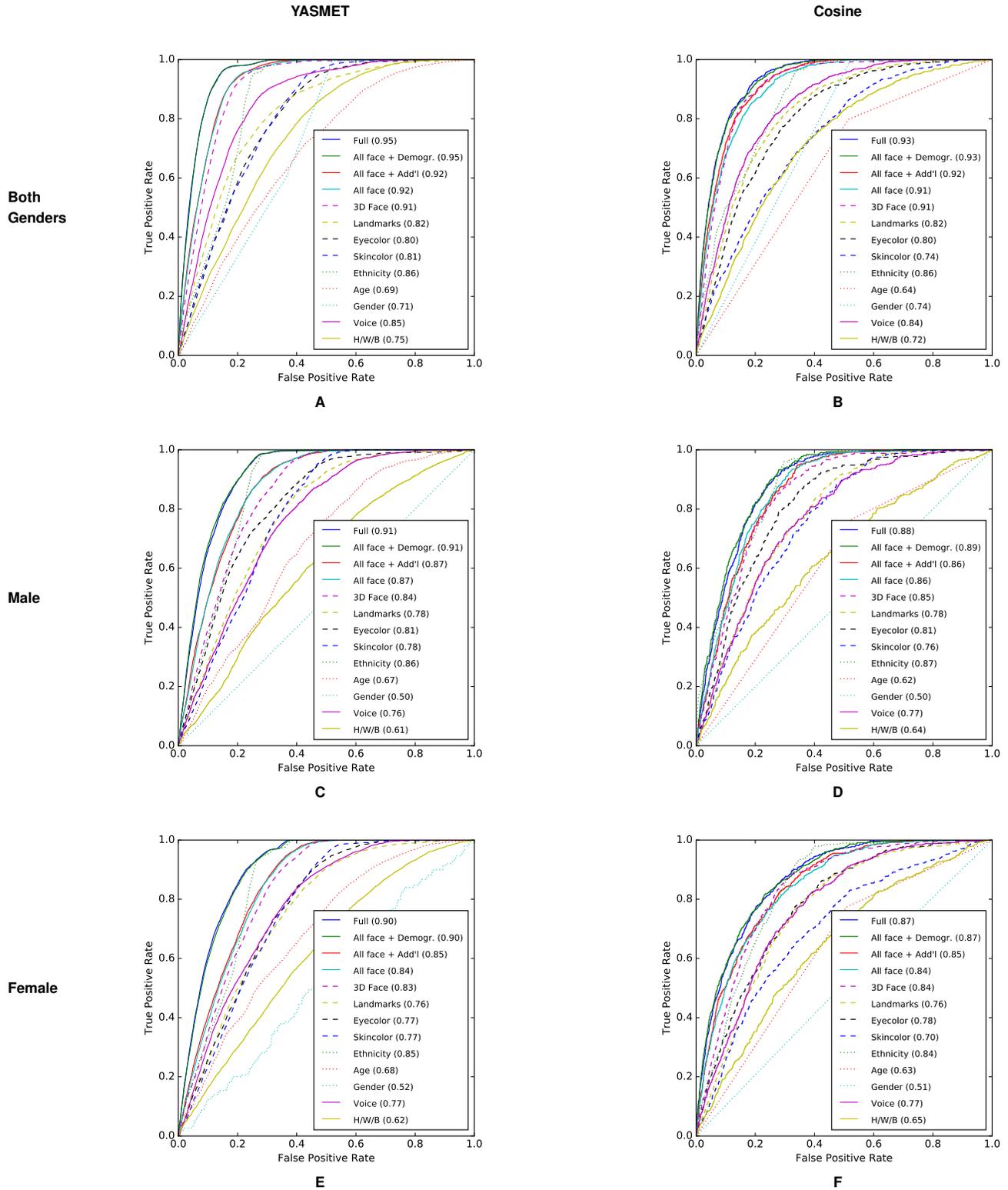


**Fig. S25.** Histogram of male pattern baldness on the Norwood-Hamilton Scale in male individuals.

**Table S14. List of models used for Identification.**

	Label	Target	Phenotype Source	Genome Source
Demogr.	Gender	Gender	= self-reported Gender	= Sex ( $\sim$ XY)
	Age	Age	= self-reported Age	$\sim$ Telomeres + XY
	Ethnicity	Ethnicity	= self-reported Ethnicity	$\sim$ Genome PCs
Add'l	Height/Weight/BMI	Height	= measured Height	$\sim$ SNPs + Genome PCs + Sex ( $\sim$ XY) + Age ( $\sim$ Telomeres, XY)
		Weight	= measured Weight	$\sim$ SNPs + Genome PCs + Sex ( $\sim$ XY CCN) + Age ( $\sim$ Telomeres, XY)
		BMI	= Weight/Height <sup>2</sup>	$\sim$ SNPs + Genome PCs + Sex ( $\sim$ XY) + Age ( $\sim$ Telomeres + XY)
	Voice	Voice <i>i</i> -vectors	= Voice <i>i</i> -vectors	$\sim$ SNPs + Genome PCs + Sex ( $\sim$ XY) + Age ( $\sim$ Telomeres + XY)
		Gender	$\sim$ Voice <i>i</i> -vectors	= Sex ( $\sim$ XY)
		Age	$\sim$ Voice <i>i</i> -vectors	$\sim$ Telomeres + XY
	Ancestry	$\sim$ Voice <i>i</i> -vectors	= Ancestry	
All Face	Eyecolor	Eyecolor	= measured Eyecolor	$\sim$ SNPs + Genome PCs
	Skincolor	Skincolor	= measured Skincolor	$\sim$ SNPs + Genome PCs
	Landmarks	Landmarks	= manual annotation	$\sim$ SNPs + Genome PCs + Sex ( $\sim$ XY) + Age ( $\sim$ Telomeres + XY)
		Gender	$\sim$ Landmarks	= Sex ( $\sim$ XY)
		Age	$\sim$ Landmarks	$\sim$ Telomeres + XY
	3D Face	Ancestry	$\sim$ Landmarks	= Ancestry
		Shape PCs	= Shape PCs	$\sim$ Genome PCs + Sex ( $\sim$ XY CCN) + Age ( $\sim$ Telomeres + XY)
		Color PCs	= Color PCs	$\sim$ Genome PCs + Sex ( $\sim$ XY) + Age ( $\sim$ Telomeres + XY)
		Gender	$\sim$ Shape PCs + Color PCs	= Sex ( $\sim$ XY)
	Age	$\sim$ Shape PCs + Color PCs	$\sim$ Telomeres + XY	
	Ancestry	$\sim$ Shape PCs + Color PCs	= Ancestry	
Together	All Face	3D Face $\cup$ Landmarks $\cup$ Eyecolor $\cup$ Skincolor		
	All Face + Add'l	3D Face $\cup$ Landmarks $\cup$ Eyecolor $\cup$ Skincolor $\cup$ Voice $\cup$ Height/Weight/BMI		
	All Face + Demogr.	3D Face $\cup$ Landmarks $\cup$ Eyecolor $\cup$ Skincolor $\cup$ Ethnicity $\cup$ Age $\cup$ Gender		
	Full	3D Face $\cup$ Landmarks $\cup$ Eyecolor $\cup$ Skincolor $\cup$ Voice $\cup$ Height/Weight/BMI $\cup$ Ethnicity $\cup$ Age $\cup$ Gender		

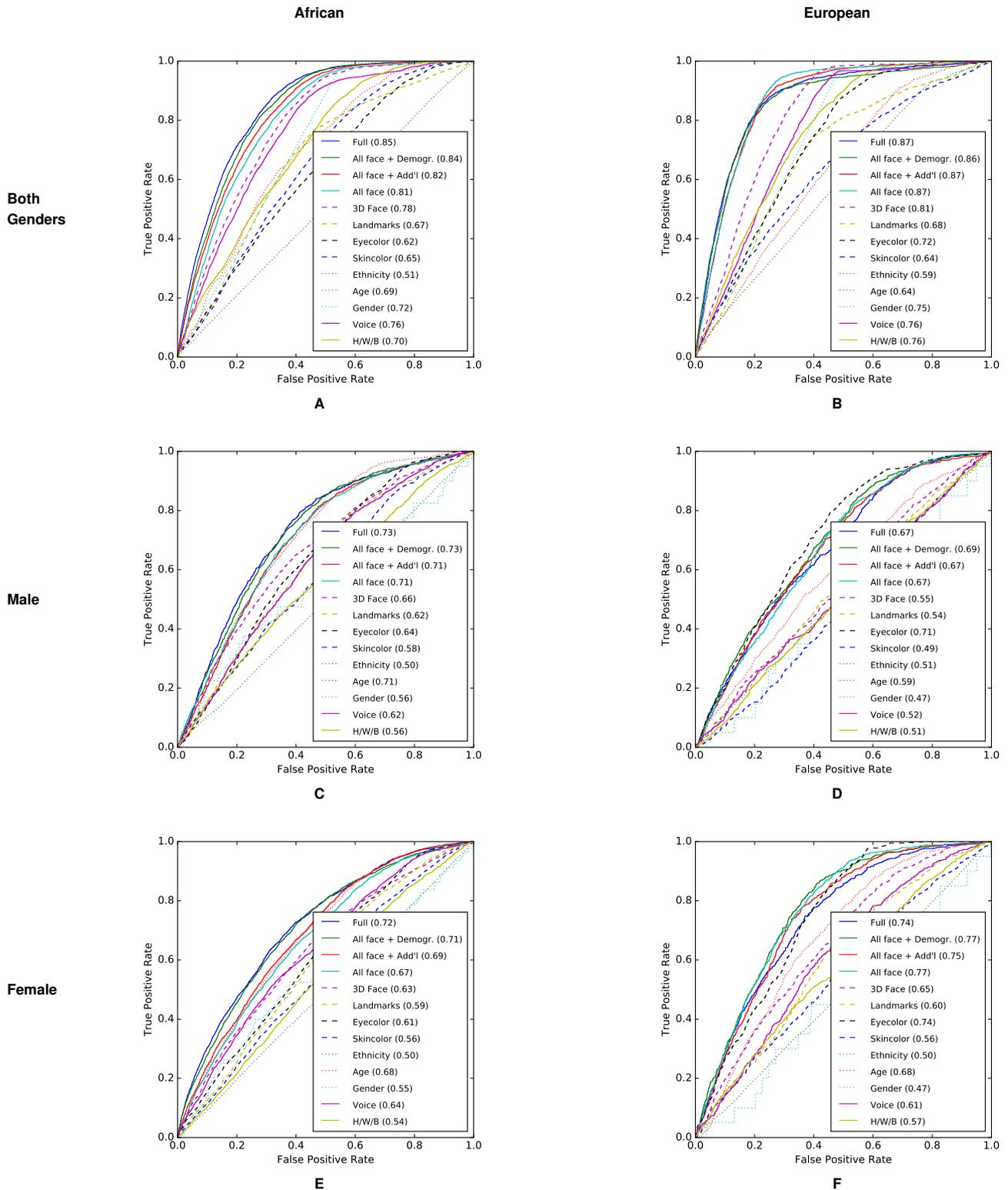
For each embedding, “Label”, all “Target” variables that are compared are listed. Each “Target” variable is predicted from a “Phenotype Source” and a “Genome Source”. “=” refers to use of observed values and “ $\sim$ ” refers to predictive models learned from data. For model covariates that are inferred from a model involving other variables, we show the model in parentheses. The embeddings under “Together” are unions of other embeddings. “Gender” is self reported gender. “Sex” is the genetic sex. “Ethnicity” is self-reported ethnicity. “Genome PCs” refers to 1,000 PCs inferred from common variation. “Ancestry” refers to five-region genomic ancestry proportions (AFR, EUR, EAS, AMR, CSA). “Telomeres” refers to estimated telomere length. “XY” refers to estimated CCNs for the sex chromosomes. “Voice *i*-vectors” refer to the voice *i*-vector embedding of voice samples. “Landmarks” refers to the XYZ values of 36 landmark locations (see Table S1). “Eyecolor” refers to RGB values of eyecolor estimated from the images. “Skincolor” refers to RGB values of facecolor estimated from the images. “Shape PCs” refers to the 3D image embedding of face shape. “Color PCs” refers to the 3D image embedding of face texture.



**Fig. S26.** ROC curves comparing YASMET and cosine in identification of individuals with different sets of features on individuals from all ethnicities. Features are learned between observed and predicted variables for (A) YASMET on both genders, (B) cosine on both genders, (C) YASMET on male individuals (D) cosine on male individuals, (E) YASMET on female individuals, and (F) cosine on female individuals.

$$\begin{aligned}
 y_n &= p_n + \epsilon_n. \\
 p_n &\sim N(0, R^2). \\
 \epsilon_n &\sim N(0, 1 - R^2).
 \end{aligned}$$

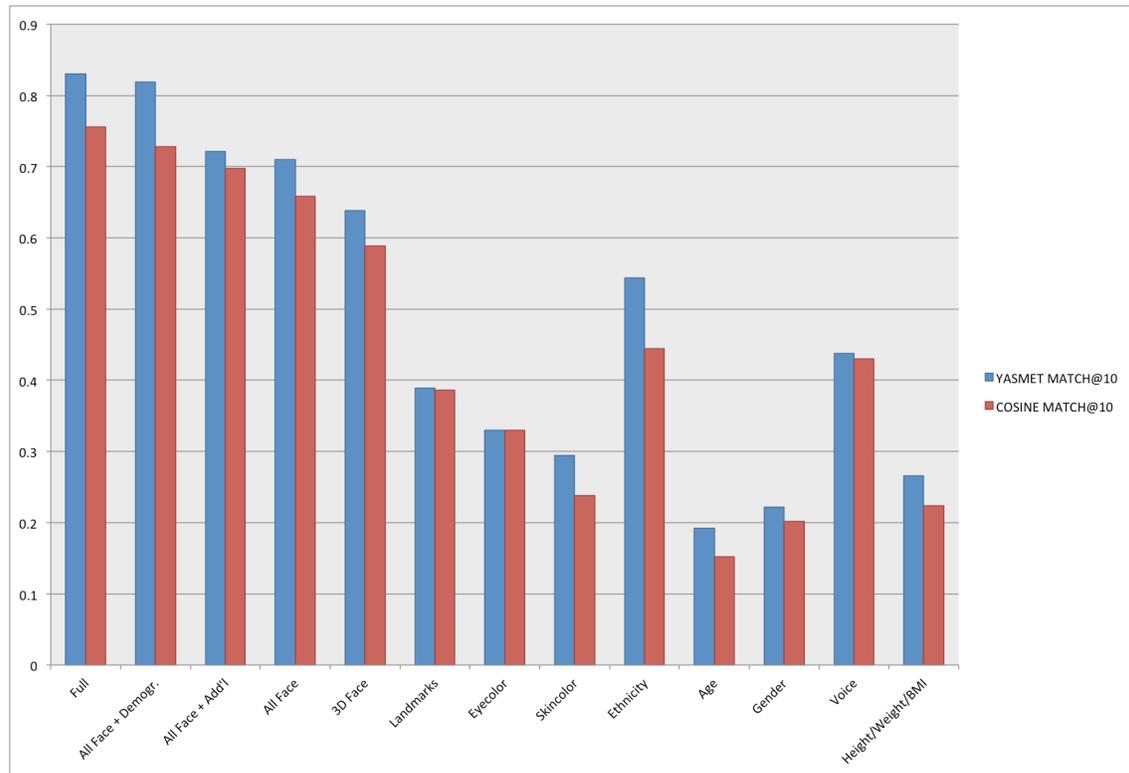
This way we achieve an expected variance explained of  $R^2$  for each trait. Fig. S33 shows how  $s_{10}$  changes for a single trait that can be predicted at a given  $R^2$  between 0 and 1.



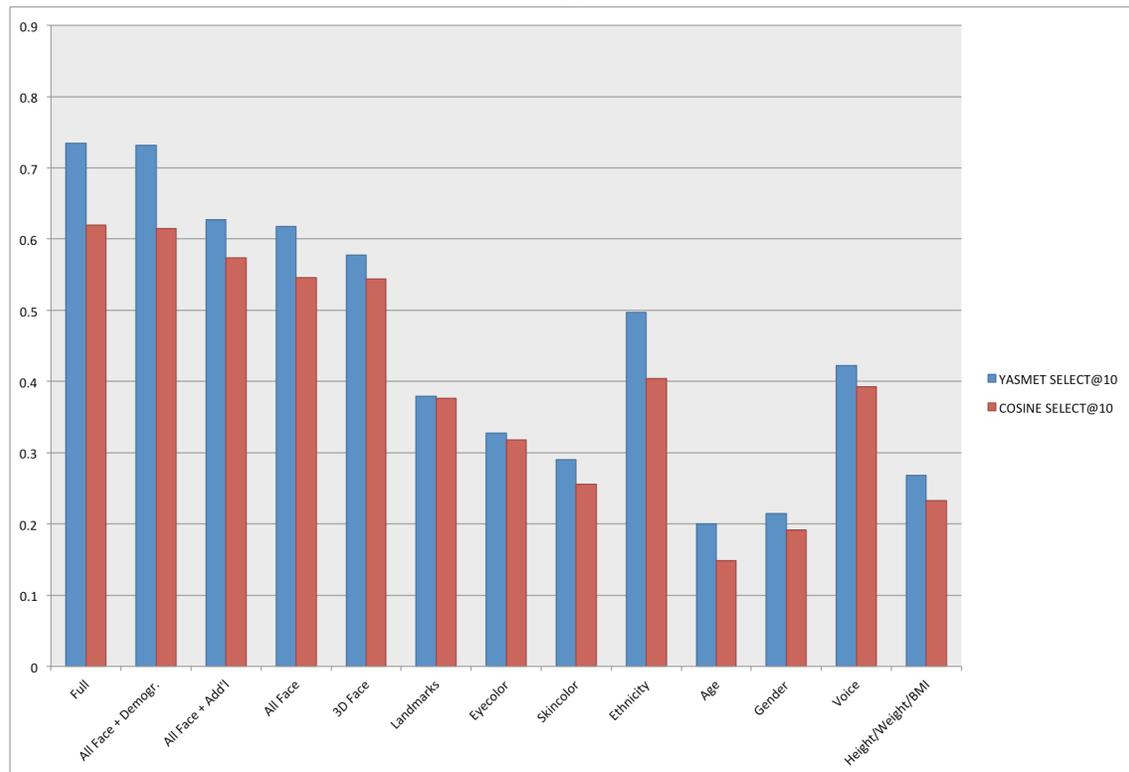
**Fig. S27.** ROC curves comparing identification between African American individuals and European individuals with different sets of features using YASMET. Features are learned between observed and predicted variables for YASMET on (A) African Americans of both genders, (B) Europeans of both genders, (C) male African Americans (D) on male Europeans, (E) female African Americans, and (F) female Europeans.

The figure indicates that for a single Gaussian distributed trait 55%  $R^2$  corresponds to roughly 1 bit of information and achieves an  $s_{10}$  of 0.2 (similar to perfect prediction of a single

binary trait like gender). Fig. S34 shows  $s_{10}$  as a function of the number of traits that each can be predicted at a given expected  $R^2$  between 1% and 90%.



A



B

**Fig. S28.** Select and match comparison between the similarity learned using YASMET and the cosine similarity on different combinations of phenotypes. (A)  $m_{10}$  and (B)  $s_{10}$ . In the  $x$ -axis, "Demogr." represents the combined ancestry, age, and sex, "Add'l" represents the combined voice and height/weight/BMI, "All Face" represents the combined 3D face, landmarks, eye color, and skin color, and "Full" represents the combined sets of phenotypes including "Demogr.", "Add'l", and "All Face".

		Male					Female				
		2	5	10	20	50	2	5	10	20	50
Full	Match	0.95	0.83	0.72	0.55	0.37	0.95	0.83	0.7	0.57	0.41
	Select	0.93	0.78	0.65	0.51	0.34	0.93	0.79	0.65	0.51	0.36
All Face + Demogr.	Match	0.95	0.84	0.71	0.55	0.36	0.95	0.83	0.7	0.59	0.42
	Select	0.93	0.79	0.65	0.51	0.34	0.92	0.79	0.66	0.53	0.36
All Face + Add'l	Match	0.92	0.73	0.51	0.33	0.16	0.91	0.7	0.53	0.37	0.19
	Select	0.88	0.67	0.49	0.32	0.18	0.88	0.65	0.48	0.33	0.17
All Face	Match	0.92	0.73	0.52	0.33	0.16	0.9	0.7	0.52	0.35	0.2
	Select	0.89	0.67	0.49	0.32	0.16	0.87	0.64	0.48	0.33	0.18
3D Face	Match	0.9	0.66	0.44	0.26	0.11	0.89	0.64	0.46	0.3	0.15
	Select	0.87	0.61	0.42	0.26	0.13	0.85	0.59	0.43	0.28	0.15
Landmarks	Match	0.82	0.52	0.3	0.18	0.069	0.82	0.5	0.29	0.16	0.058
	Select	0.78	0.48	0.29	0.16	0.079	0.77	0.46	0.29	0.15	0.073
Eyecolor	Match	0.85	0.56	0.35	0.2	0.069	0.81	0.52	0.32	0.17	0.071
	Select	0.82	0.54	0.34	0.2	0.084	0.78	0.49	0.31	0.19	0.08
Skincolor	Match	0.79	0.46	0.24	0.12	0.049	0.77	0.42	0.24	0.13	0.048
	Select	0.77	0.43	0.23	0.12	0.055	0.76	0.42	0.23	0.12	0.056
Ethnicity	Match	0.88	0.67	0.5	0.39	0.27	0.87	0.67	0.53	0.41	0.29
	Select	0.88	0.65	0.48	0.35	0.22	0.87	0.67	0.52	0.39	0.26
Age	Match	0.7	0.33	0.19	0.095	0.048	0.71	0.36	0.19	0.1	0.043
	Select	0.67	0.34	0.19	0.093	0.038	0.68	0.36	0.2	0.11	0.048
Gender	Match	0.51	0.2	0.098	0.046	0.02	0.51	0.16	0.069	0.081	0.013
	Select	0.5	0.2	0.1	0.05	0.02	0.51	0.15	0.087	0.05	0.02
Voice	Match	0.8	0.48	0.27	0.15	0.057	0.81	0.52	0.31	0.17	0.068
	Select	0.75	0.44	0.27	0.15	0.077	0.78	0.49	0.31	0.18	0.077
Height/Weight/BMI	Match	0.67	0.32	0.17	0.089	0.038	0.67	0.33	0.18	0.11	0.046
	Select	0.63	0.31	0.17	0.097	0.044	0.64	0.33	0.18	0.1	0.05
Random	Match	0.5	0.2	0.1	0.05	0.02	0.5	0.2	0.1	0.05	0.02
	Select	0.5	0.2	0.1	0.05	0.02	0.5	0.2	0.1	0.05	0.02

A

B

Fig. S29. Select and match results stratified by sex. Results are based on the YASMET similarity. Shown are different numbers of pool sizes from two to 50 within (A) male and (B) female subjects.

		African					European				
		2	5	10	20	50	2	5	10	20	50
Full	Match	0.9	0.66	0.47	0.31	0.17	0.9	0.75	0.57	0.4	0.2
	Select	0.86	0.61	0.44	0.29	0.16	0.88	0.67	0.5	0.34	0.19
All Face + Demogr.	Match	0.89	0.66	0.46	0.3	0.15	0.91	0.73	0.57	0.38	0.2
	Select	0.85	0.61	0.42	0.27	0.14	0.88	0.68	0.51	0.36	0.19
All Face + Add'l	Match	0.88	0.64	0.45	0.28	0.14	0.9	0.68	0.48	0.28	0.14
	Select	0.84	0.59	0.41	0.26	0.15	0.87	0.65	0.45	0.28	0.15
All Face	Match	0.85	0.59	0.4	0.24	0.11	0.89	0.69	0.49	0.31	0.12
	Select	0.81	0.55	0.38	0.23	0.12	0.88	0.65	0.46	0.3	0.13
3D Face	Match	0.83	0.53	0.33	0.19	0.076	0.85	0.55	0.34	0.2	0.081
	Select	0.79	0.5	0.32	0.19	0.087	0.82	0.55	0.35	0.21	0.1
Landmarks	Match	0.73	0.38	0.22	0.1	0.043	0.72	0.38	0.2	0.1	0.039
	Select	0.68	0.36	0.21	0.1	0.037	0.69	0.36	0.2	0.1	0.03
Eyecolor	Match	0.68	0.34	0.17	0.099	0.051	0.76	0.4	0.2	0.12	0.041
	Select	0.64	0.33	0.18	0.1	0.055	0.74	0.38	0.21	0.11	0.05
Skincolor	Match	0.66	0.31	0.16	0.085	0.037	0.71	0.34	0.2	0.11	0.038
	Select	0.65	0.3	0.15	0.081	0.031	0.64	0.34	0.2	0.13	0.056
Ethnicity	Match	0.53	0.2	0.11	0.057	0.025	0.61	0.31	0.19	0.1	0.04
	Select	0.51	0.21	0.11	0.058	0.025	0.57	0.28	0.17	0.099	0.051
Age	Match	0.72	0.36	0.19	0.097	0.041	0.67	0.32	0.19	0.098	0.039
	Select	0.69	0.36	0.2	0.1	0.042	0.64	0.33	0.2	0.094	0.041
Gender	Match	0.71	0.38	0.2	0.097	0.044	0.75	0.4	0.2	0.11	0.049
	Select	0.72	0.37	0.2	0.1	0.04	0.74	0.39	0.2	0.1	0.04
Voice	Match	0.79	0.47	0.29	0.17	0.07	0.78	0.44	0.23	0.13	0.053
	Select	0.76	0.46	0.29	0.17	0.076	0.77	0.44	0.25	0.13	0.053
Height/Weight/BMI	Match	0.74	0.4	0.22	0.11	0.051	0.8	0.45	0.25	0.14	0.052
	Select	0.71	0.37	0.23	0.14	0.077	0.77	0.43	0.24	0.14	0.057
Random	Match	0.5	0.2	0.1	0.05	0.02	0.5	0.2	0.1	0.05	0.02
	Select	0.5	0.2	0.1	0.05	0.02	0.5	0.2	0.1	0.05	0.02

Fig. S30. Select and match results stratified by Ethnicity. Results are based on the YASMET similarity. Shown are different numbers of pool sizes from two to 50 within (A) African and (B) European subjects.

		African					European					
		2	5	10	20	50	2	5	10	20	50	
Male	Full	Match	0.82	0.48	0.29	0.18	nan	0.72	0.4	0.23	0.11	nan
		Select	0.74	0.47	0.28	0.16	nan	0.67	0.37	0.24	0.13	nan
	All Face + Demogr.	Match	0.8	0.47	0.29	0.15	nan	0.74	0.41	0.2	0.12	nan
		Select	0.75	0.42	0.27	0.17	nan	0.7	0.38	0.24	0.13	nan
	All Face + Add'l	Match	0.74	0.47	0.28	0.17	nan	0.69	0.35	0.2	0.088	nan
		Select	0.7	0.44	0.27	0.15	nan	0.68	0.36	0.2	0.13	nan
	All Face	Match	0.77	0.47	0.28	0.16	nan	0.7	0.38	0.19	0.1	nan
		Select	0.71	0.42	0.28	0.15	nan	0.68	0.36	0.2	0.11	nan
	3D Face	Match	0.76	0.39	0.22	0.15	nan	0.58	0.26	0.14	0.082	nan
		Select	0.67	0.37	0.23	0.14	nan	0.56	0.25	0.13	0.073	nan
	Landmarks	Match	0.69	0.33	0.17	0.093	nan	0.61	0.24	0.13	0.053	nan
		Select	0.64	0.34	0.18	0.086	nan	0.57	0.25	0.12	0.07	nan
	Eyecolor	Match	0.68	0.34	0.16	0.09	nan	0.73	0.32	0.16	0.085	nan
		Select	0.63	0.34	0.17	0.082	nan	0.7	0.34	0.19	0.11	nan
	Skincolor	Match	0.62	0.32	0.17	0.098	nan	0.54	0.22	0.1	0.053	nan
		Select	0.6	0.3	0.18	0.096	nan	0.51	0.19	0.087	0.032	nan
	Ethnicity	Match	0.5	0.21	0.12	0.054	nan	0.55	0.2	0.086	0.05	nan
		Select	0.5	0.2	0.1	0.052	nan	0.5	0.2	0.1	0.05	nan
	Age	Match	0.71	0.37	0.21	0.095	nan	0.61	0.29	0.14	0.085	nan
		Select	0.67	0.37	0.22	0.098	nan	0.6	0.28	0.16	0.085	nan
Gender	Match	0.4	0.28	0.12	0.17	nan	0.42	0.26	0.17	0	nan	
	Select	0.38	0.28	0.17	0.1	nan	0.36	0.29	0.2	0	nan	
Voice	Match	0.67	0.34	0.17	0.072	nan	0.56	0.23	0.11	0.047	nan	
	Select	0.61	0.32	0.16	0.082	nan	0.53	0.22	0.12	0.065	nan	
Height/Weight/BMI	Match	0.59	0.28	0.17	0.089	nan	0.56	0.21	0.098	0.052	nan	
	Select	0.57	0.28	0.17	0.099	nan	0.52	0.19	0.092	0.05	nan	
Random	Match	0.5	0.2	0.1	0.05	0.02	0.5	0.2	0.1	0.05	0.02	
	Select	0.5	0.2	0.1	0.05	0.02	0.5	0.2	0.1	0.05	0.02	

		A					B					
		2	5	10	20	50	2	5	10	20	50	
Female	Full	Match	0.8	0.5	0.32	0.21	0.11	0.8	0.52	0.31	0.17	nan
		Select	0.72	0.46	0.31	0.2	0.099	0.78	0.49	0.3	0.19	nan
	All Face + Demogr.	Match	0.79	0.51	0.3	0.19	0.11	0.82	0.51	0.32	0.17	nan
		Select	0.73	0.46	0.29	0.18	0.097	0.79	0.49	0.32	0.2	nan
	All Face + Add'l	Match	0.77	0.46	0.29	0.18	0.085	0.84	0.52	0.29	0.17	nan
		Select	0.71	0.43	0.27	0.18	0.087	0.79	0.5	0.3	0.16	nan
	All Face	Match	0.74	0.4	0.25	0.15	0.069	0.85	0.51	0.29	0.12	nan
		Select	0.69	0.39	0.24	0.14	0.077	0.79	0.48	0.29	0.15	nan
	3D Face	Match	0.71	0.36	0.21	0.11	0.054	0.71	0.34	0.19	0.097	nan
		Select	0.65	0.34	0.21	0.11	0.045	0.66	0.34	0.19	0.11	nan
	Landmarks	Match	0.64	0.32	0.15	0.075	0.029	0.6	0.31	0.15	0.09	nan
		Select	0.6	0.29	0.16	0.077	0.027	0.61	0.28	0.17	0.077	nan
	Eyecolor	Match	0.68	0.32	0.18	0.11	0.054	0.77	0.39	0.21	0.097	nan
		Select	0.63	0.32	0.18	0.1	0.052	0.77	0.39	0.24	0.13	nan
	Skincolor	Match	0.56	0.24	0.14	0.064	0.027	0.52	0.29	0.14	0.092	nan
		Select	0.55	0.23	0.12	0.044	0.019	0.53	0.27	0.18	0.09	nan
	Ethnicity	Match	0.47	0.2	0.11	0.053	0.023	0.54	0.2	0.1	0.065	nan
		Select	0.5	0.2	0.099	0.048	0.021	0.5	0.2	0.1	0.05	nan
	Age	Match	0.69	0.36	0.2	0.11	0.058	0.69	0.33	0.16	0.075	nan
		Select	0.67	0.37	0.21	0.12	0.048	0.66	0.33	0.18	0.075	nan
Gender	Match	0.5	0.25	0.14	0.11	0.02	0.36	0.32	0.1	0	nan	
	Select	0.52	0.23	0.14	0.075	0	0.32	0.32	0.2	0	nan	
Voice	Match	0.68	0.35	0.21	0.11	0.057	0.67	0.34	0.19	0.12	nan	
	Select	0.64	0.34	0.19	0.11	0.047	0.65	0.31	0.15	0.078	nan	
Height/Weight/BMI	Match	0.57	0.26	0.12	0.064	0.031	0.61	0.25	0.15	0.087	nan	
	Select	0.54	0.24	0.12	0.05	0.022	0.58	0.24	0.13	0.078	nan	
Random	Match	0.5	0.2	0.1	0.05	0.02	0.5	0.2	0.1	0.05	0.02	
	Select	0.5	0.2	0.1	0.05	0.02	0.5	0.2	0.1	0.05	0.02	

**Fig. S31.** Select and match results stratified by sex and ethnicity. Results are based on the YASMET similarity. Shown are different numbers of pool sizes from two to 50 within (A) African American males, (B) European males (C) African American females, (D) European females. "nan" values indicate that the size of the test sample sets were too small to run the lineup size.

		2	5	10	20	50
Full	Match	0.97	0.92	0.83	0.7	0.53
	Select	0.93	0.83	0.74	0.62	0.45
All Face + Demogr.	Match	0.98	0.91	0.82	0.7	0.53
	Select	0.93	0.83	0.73	0.61	0.45
All Face + Add'l	Match	0.96	0.85	0.72	0.55	0.33
	Select	0.91	0.76	0.63	0.48	0.28
All Face	Match	0.95	0.84	0.71	0.53	0.32
	Select	0.9	0.76	0.62	0.46	0.29
3D Face	Match	0.94	0.8	0.64	0.46	0.25
	Select	0.89	0.74	0.58	0.42	0.24
Landmarks	Match	0.87	0.61	0.39	0.24	0.1
	Select	0.81	0.55	0.38	0.23	0.11
Eyecolor	Match	0.85	0.55	0.33	0.19	0.075
	Select	0.8	0.52	0.33	0.19	0.085
Skincolor	Match	0.81	0.5	0.29	0.16	0.065
	Select	0.79	0.47	0.29	0.16	0.07
Ethnicity	Match	0.9	0.71	0.54	0.41	0.27
	Select	0.87	0.66	0.5	0.36	0.25
Age	Match	0.69	0.36	0.19	0.099	0.043
	Select	0.66	0.35	0.2	0.11	0.043
Gender	Match	0.74	0.4	0.22	0.12	0.051
	Select	0.73	0.39	0.21	0.11	0.049
Voice	Match	0.88	0.66	0.44	0.26	0.11
	Select	0.84	0.61	0.42	0.26	0.12
Height/Weight/BMI	Match	0.77	0.46	0.27	0.14	0.061
	Select	0.74	0.43	0.26	0.15	0.065
Random	Match	0.5	0.2	0.1	0.05	0.02
	Select	0.5	0.2	0.1	0.05	0.02

Fig. S32. Select and match results for models with 277 SNPs reported by Qiao et al. (68) for different numbers of pool sizes from two to 50 on all subjects.

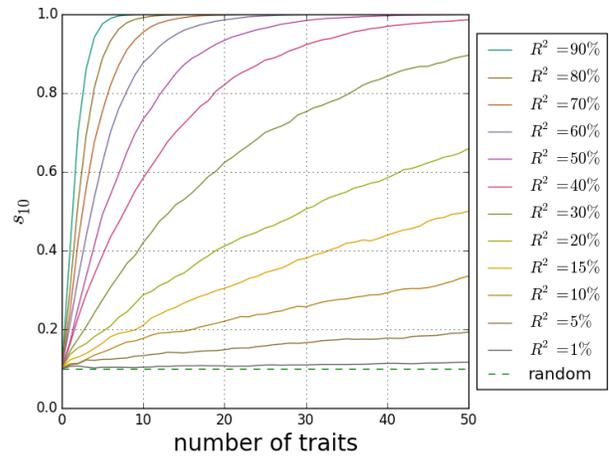


Fig. S34. Simulation of  $s_{10}$  as a function of number of traits. The plot shows how  $s_{10}$  performance changes as a function of the number of traits for different expected  $R^2$ . Random predictions (green dashed line) would achieve an  $s_{10}$  performance of 10% irrespective of the number of traits.

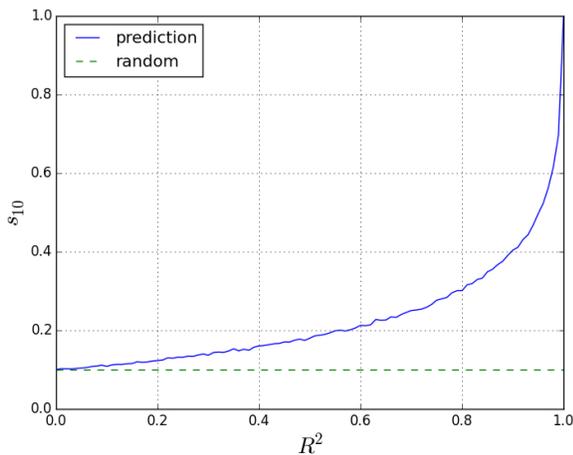


Fig. S33. Simulation of  $s_{10}$  as a function of  $R^2$  for a single trait. The plot shows simulation results for a single independently Gaussian distributed trait as a function of expected  $R^2$  (blue solid line). A random prediction (green dashed line) would achieve an  $s_{10}$  performance of 10%.

**Landmark Genome-wide Association Analysis.** We have performed a GWAS of all pairwise distances (630 in total) for 36 facial landmarks in Table S1 on 1,036 individuals. We expect to have family relatedness and cryptic relatedness due to the particular characteristics of the study. Consequently, we attempted to remove related individuals from the landmark GWAS to help us to alleviate the problem of inflated association statistics due to related individuals in the GWAS. To remove the related individuals, we excluded related individuals iteratively so that the degree of relatedness between remaining individuals was preserved a certain degree. We computed the relatedness using the KING (69) algorithm on 11 million autosomal variants. After excluding individuals with a relatedness larger than 0.177 (corresponding to filtering out 1st degree relatedness) and individuals with missing or outliers in particular landmark distance measurements, we obtained data sets with 824-873 (870 average) individuals across different landmark distances. The exact number of individuals differed between traits because the traits had different numbers of missing and outlier values. To determine outlier values in a given landmark distance, we attempt to use the prediction landmark distances and distribution of the measured landmark distances. For a given landmark distance, we built ridge regression models using other landmark distances as features. The landmark distances used as features are the distances between the remaining 34 landmarks and each one of the two landmark points from the landmark distance, resulting in a total of 68 landmark distance features. A trait value was considered to be an outlier if the measured landmark distance was more than three standard deviations away from the out of sample predicted value in a 10 fold CV, or four standard deviations away from the distribution of measured values.

We filtered SNPs and indels to have a maximum missing call rate of 1%. We also included a set of 38 candidate SNPs from Adhikari et al. (40), for which we relaxed this constraint to a maximum of 10% missingness. We filtered variants whose frequency of the most frequent dosage value (0, 1, or 2) is less than 2% among the non-missing variants. Note that this is similar to a minor allele frequency filter of 1%, except that it also filters out presumably erroneous variants that are heterozygous in a majority of the individuals. After filtering, 11,441,529 autosomal variants remain. Finally, we performed LD pruning within 500kb windows, filtering out variants with a correlation of larger than 0.64, resulting in a set of 2,147,867 variants that we used for GWAS testing. In addition to standard GWAS testing, we separately performed association tests in male and female samples and combined the corresponding test statistics in a random effects meta-analysis.

We included sex and five genomic PCs as covariates to capture population structure and sex effects within remaining individuals. Dataset S1 shows SNPs that had  $P$ -values smaller than  $5 \times 10^{-8}$  in at least one of the two tests. 166 associations were significant in the combined association test, 351 associations were significant in the meta-analysis, with an overlap of 51 associations between the two approaches. Since we tested 630 phenotypes, we also computed estimates of the false discovery rate (FDR) for the combined GWAS. The lowest FDR was estimated as 0.28. Thus we could not find significantly associated SNPs when considering an FDR threshold of 5%.

Table S15 shows the list of candidate SNPs from a previ-

ously published GWAS on human facial variation (40) that achieved a  $P$ -value smaller than  $5 \times 10^{-4}$  on our data. Note that the replication study with 501 individuals within Adhikari et al. (40) also used  $6 \times 10^{-3}$  as the significance threshold when comparing the original results with replication study results. We used a less stringent reporting threshold on this set of SNPs as we have a lower number of individuals (870 on average) compared to the original study of Adhikari et al. (40) (6,000 individuals). A total of six SNPs was replicated, which are associated with six facial phenotypes (Brow ridge protrusion, Upper lip thickness, Columella inclination, Nose protrusion, Nose tip angle, Nose wing breadth) as shown in Table S15.

### Equivalence between a model using large numbers of principal components and a ridge regression on common variants.

In the following, we will show that a ridge regression on the matrix of all common SNPs is equivalent to a ridge regression on the top  $N_{tr}$  (or more) PCs of the training data, where  $N_{tr}$  is the size of the training data set. In all of our experiments the size of the training data set is less than 1,000 (due to CV). This is sufficient to show that the common SNP PC models trained in our paper are mathematically equivalent to a ridge regression on the matrix of SNPs that had been used to compute PCs. We further note that such a ridge regression fit with regularization parameter  $\gamma$  is equivalent to BLUP prediction using a realized relationship matrix determined from the same set of SNPs and ratio of environmental variance over genetic variance  $\sigma_e^2/\sigma_g^2$  equal to  $\gamma$ , a model known as G-BLUP (70).

Let  $X \in R^{N_{tr} \times S}$  be the matrix of common variants in the training data set, with the training data set size  $N_{tr}$  less than the number of SNPs  $S$ . Let the vector of predictions for the test data set given the matrix of SNPs for the test data set  $X_{tests}$  be  $\hat{y}_{test}$ .

$$\hat{y}_{test} = X_{test} \beta_{Ridge}.$$

Plugging in the ridge estimate for  $\beta_{Ridge}$ , we get

$$\hat{y}_{test} = X_{test} \overbrace{(X^T X + \gamma I_S)^{-1} X^T}^{\beta_{Ridge}} y,$$

where  $\gamma$  is the ridge penalty and  $I_S$  is the  $S \times S$  identity matrix.

Further, let the singular value decomposition of  $X$  be  $X = U \Lambda V^T$ , where  $U \in R^{N_{tr} \times N_{tr}}$  is the orthogonal matrix of left singular vectors,  $V \in R^{S \times S}$  is the orthogonal matrix of right singular vectors and the  $N_{tr} \times S$  dimensional diagonal matrix  $\Lambda$  holds the  $N_{tr}$  singular values of  $X$ .

$$\begin{aligned} \hat{y}_{test} &= X_{test} (V \Lambda^T U^T U \Lambda V^T + \gamma I_S)^{-1} V \Lambda^T U^T y \\ &= X_{test} (V \Lambda^T \Lambda V^T + \gamma V V^T)^{-1} V \Lambda^T U^T y, \\ &= X_{test} V (\Lambda^T \Lambda + \gamma I_S)^{-1} V^T V \Lambda^T U^T y, \\ &= X_{test} V (\Lambda^T \Lambda + \gamma I_S)^{-1} \Lambda^T U^T y, \end{aligned}$$

We observe that  $X_{test} V$  is the matrix of all  $S$  projected PCs for the training data. Further, even though  $(\Lambda^T \Lambda + \gamma I_S)^{-1}$  is a full rank  $S \times S$  diagonal matrix, it is being multiplied by the  $S \times N_{tr}$  matrix  $\Lambda^T$ , which can be written

**Table S15. List of replicated SNPs from previous human facial variation GWAS (40).**

rsid	chr	Position	Associated Trait	<i>P</i> -Value	Reported association (40)
rs2235371	1	209790735	SL_PG	$6.7 \times 10^{-4}$	Brow ridge protrusion ( $1.8 \times 10^{-3}$ ★) Upper lip thickness ( $6.6 \times 10^{-3}$ ★)
rs2045323	4	153910747	ALL_LI	$3.5 \times 10^{-4}$	Columella inclination ( $3 \times 10^{-9}$ )
			ALL_SL	$9.5 \times 10^{-4}$	Nose protrusion ( $1 \times 10^{-9}$ )
			ALL_ST	$2.6 \times 10^{-4}$	Nose tip angle ( $2 \times 10^{-8}$ )
			ALR_LI	$1.0 \times 10^{-4}$	
			ALR_SL	$2.1 \times 10^{-4}$	
			ALR_ST	$4.3 \times 10^{-5}$	
			CPHR_STO	$4.2 \times 10^{-4}$	
			CPHL_STO	$1.0 \times 10^{-1}$	
			SBALL_LI	$9.8 \times 10^{-4}$	
			SBALR_LI	$2.9 \times 10^{-4}$	
			SBALR_SL	$5.4 \times 10^{-4}$	
			SBALR_STO	$7.5 \times 10^{-4}$	
			SBALL_STO	$1.8 \times 10^{-1}$	
rs12651681	4	154328210	EBL_ENL	$5.9 \times 10^{-5}$	Columella inclination ( $2.4 \times 10^{-8}$ )
			EBL_IRL	$1.1 \times 10^{-4}$	
			EBR_IRR	$3.8 \times 10^{-3}$	
			EBL_PIL	$8.5 \times 10^{-4}$	
			EBL_PIL	$6.3 \times 10^{-3}$	
			EBR_ENR	$3.4 \times 10^{-4}$	
rs12644248	4	154314240	SBALL_PG	$3.4 \times 10^{-4}$	Columella inclination ( $6.6 \times 10^{-9}$ )
			SBALR_PG	$1.3 \times 10^{-3}$	
rs12543318	8	87856112	CPHR_CHL	$7.8 \times 10^{-4}$	Brow ridge protrusion ( $2.9 \times 10^{-2}$ ★)
			CPHL_CHR	$3.4 \times 10^{-1}$	Columella inclination ( $1.5 \times 10^{-2}$ ★)
rs927833	20	22060939	ALL_LI	$3.4 \times 10^{-4}$	Nose wing breadth ( $1 \times 10^{-9}$ )
			ALL_SL	$7.2 \times 10^{-4}$	
			ALR_LI	$2.1 \times 10^{-4}$	
			ALR_SL	$6.8 \times 10^{-4}$	

We have used a less stringent threshold of  $1.0 \times 10^{-3}$  as we have fewer individuals (870 on average) than this study (6,000 individuals). *P*-value with ★ represent Bonferroni-adjusted *P*-values reported by Adhikari et al. (40). A total six SNPs were replicated, which are associated with six facial phenotypes (brow ridge protrusion, Upper lip thickness, columella inclination, nose protrusion, nose tip angle, nose wing breath). Gray marks opposite (L/R) results that are above the reporting threshold of  $1.0 \times 10^{-3}$ .

as  $\begin{bmatrix} \Lambda_{N_{tr}}; & 0_{(S-N_{tr}) \times (S-N_{tr})} \end{bmatrix}^T$ , where  $\Lambda_{N_{tr}}$  is the symmetric  $N_{tr} \times N_{tr}$  diagonal matrix holding all non-zero singular values of  $X$  and  $0_{(S-N_{tr}) \times (S-N_{tr})}$  is an  $(S - N_{tr}) \times (S - N_{tr})$  matrix of all zeros. It follows that only the top  $N_{tr}$  diagonal entries of  $(\Lambda^T \Lambda + \gamma I_S)^{-1}$  are relevant for the matrix multiplication with  $\Lambda^T$  whereas the lower part becomes zero.

It follows that we can rewrite the expression for the prediction above using only the top  $N_{tr} \times N_{tr}$  part of  $\Lambda$ , denoted by  $\Lambda_{N_{tr}}$  and using the  $N_{tr} \times N_{tr}$  identity matrix  $I_{N_{tr}}$  and the vector of zeros  $0_{(S-N_{tr}) \times (S-N_{tr})}$ :

$$\hat{y}_{test} = X_{test} V \begin{bmatrix} (\Lambda_{N_{tr}}^T \Lambda_{N_{tr}} + \gamma I_{N_{tr}})^{-1} \Lambda_{N_{tr}}^T U^T y. \\ 0_{(S-N_{tr}) \times (S-N_{tr})} \end{bmatrix}.$$

Dropping zeros, the expression finally can be reduced to one using only the first  $N_{tr}$  columns of  $V$ , denoted as  $V_{N_{tr}} \in R^{S \times N_{tr}}$ .

$$\begin{aligned} \hat{y}_{test} &= X_{test} V_{N_{tr}} (\Lambda_{N_{tr}}^T \Lambda_{N_{tr}} + \gamma I_{N_{tr}})^{-1} \Lambda_{N_{tr}}^T U^T y \\ &= X_{test} V_{N_{tr}} (\Lambda_{N_{tr}}^T U^T U \Lambda_{N_{tr}} + \gamma I_{N_{tr}})^{-1} \Lambda_{N_{tr}}^T U^T y, \end{aligned}$$

where re-introducing  $U^T U$  yielded the desired result.  $U \Lambda_{N_{tr}}$  is the  $N_{tr} \times N_{tr}$  matrix holding the top  $N_{tr}$  principal components of the training SNP matrix and  $X_{test} V_{N_{tr}}$  is the  $N_{test} \times N_{tr}$  matrix of projections for the test SNPs on the training PCs. This completes our proof (q.e.d.).

**Assessing baseline re-identification accuracy.** In the following, we will theoretically analyze the performance of several baseline strategies for re-identification and performance. These strategies may not utilize any genomic (or other) information on the individual of interest, but may use any information that is available on the lineup (*i.e.*, phenotype data), and even on the study cohort as a whole. Additionally, any baselines may use summary information for the cohort, such as sample proportions (*e.g.*, breakdowns of ethnicity, age or sex).

For a given lineup of size  $N$ , which has  $N_1 \geq N_2 \geq \dots \geq N_C$  individuals from each of  $C$  discrete subgroups or classes (*e.g.*, all individuals that have the same ethnicity and/or sex), with  $\sum_{c=1}^C N_c = N$ , the probability of any person  $n$  in the lineup matching the (unknown) genomic sample is equal to  $1/N$ .

$$P(n) = \frac{1}{N} \quad \forall n \in [1, \dots, N].$$

Under this setup we can compare two baseline strategies, picking an individual at random, or picking an individual that belongs to a particular strata in the data, for example the largest/smallest strata. We will see that both of these strategies have an expected accuracy of  $1/N$ . Consequently, we use  $1/N$  as a baseline accuracy for each re-identification experiment.

**Strategy 1: Pick a random individual from the lineup.** This is the baseline strategy that we chose. It has an expected accuracy of  $1/N$ .

**Strategy 2: Always pick a random individual that originates from the majority/minority class in the lineup.** Another baseline strategy may always pick from a specific class, typically the class with either the most, or the least members. In the following, we will see that this strategy has an identical expected accuracy of  $1/N$ .

It follows that the probability that the genomic sample comes from one class  $c$  is proportional to the number of individuals that belong to that class.

$$P(k) = \frac{N_c}{N} \quad \forall c \in [1, \dots, C].$$

The complement, namely that the genomic sample comes from a different class follows as

$$P(\bar{c}) = 1 - P(k) \quad \forall c \in [1, \dots, C].$$

The probability that the genomic sample belongs to a single selected individual from that class, given that the genomic sample belongs to any individual from that class follows by conditioning

$$P(n|c) = \frac{P(n)}{P(c)} = \frac{1}{N_c}$$

On the other side, the probability that the genomic sample belongs to a single selected individual from that class, given that the genomic sample belongs to an individual from a different class, is equal to 0.

The marginal probability is then

$$\underbrace{P(n|c) P(c)}_{1/N} + \underbrace{0 \cdot P(\bar{k})}_0 = \frac{1}{N}.$$

We can see that this strategy has the same expected accuracy of  $1/N$ .

- Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *science* 319(5866):1100–1104.
- 1000 Genomes Project Consortium, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 19(9):1655–1664.
- Wu MC, et al. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics* 89(1):82–93.
- Zook JM, et al. (2016) Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Scientific data* 3.
- Ding Z, et al. (2014) Estimating telomere length from whole genome sequence data. *Nucleic acids research* p. gku181.
- Telenti A, et al. (2016) Deep sequencing of 10,000 human genomes. *Proceedings of the National Academy of Sciences* p. 201613365.
- Long T, et al. (2017) Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nature Genetics*.
- Benjamini Y, Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic acids research* p. gks001.
- Zubakov D, et al. (2010) Estimating human age from T-cell DNA rearrangements. *Current Biology* 20(22):R970–R971.
- Ou XI, et al. (2012) Predicting human age with bloodstains by sjTREC quantification. *PLoS one* 7(8):e42412.
- Rausch T, et al. (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28(18):i333–i339.
- Hochheiser H, et al. (2011) The FaceBase Consortium: a comprehensive program to facilitate craniofacial research. *Developmental biology* 355(2):175–182.
- Turk M, Pentland A (1991) Eigenfaces for recognition. *Journal of cognitive neuroscience* 3(1):71–86.
- Li M, Yuan B (2004) A novel statistical linear discriminant analysis for image matrix: two-dimensional fisherfaces in *Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on*. (IEEE), Vol. 2, pp. 1419–1422.
- Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: Closing the gap to human-level performance in face verification in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1701–1708.
- Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence* 31(2):210–227.

18. Kämäräinen JK, Hadid A, Pietikäinen M (2011) Local representation of facial features in *Handbook of Face Recognition*. (Springer), pp. 79–108.
19. Turk G, O'Brien JF (2002) Modelling with implicit surfaces that interpolate. *ACM Transactions on Graphics (TOG)* 21(4):855–873.
20. Pérez P, Gangnet M, Blake A (2003) Poisson image editing in *ACM Transactions on Graphics (TOG)*. (ACM), Vol. 22, pp. 313–318.
21. Horn BK (1987) Closed-form solution of absolute orientation using unit quaternions. *JOSA A* 4(4):629–642.
22. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence* 24(4):509–522.
23. Amberg B, Romdhani S, Vetter T (2007) Optimal step nonrigid ICP algorithms for surface registration in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. (IEEE), pp. 1–8.
24. Guo J, Mei X, Tang K (2013) Automatic landmark annotation and dense correspondence registration for 3D human facial images. *BMC bioinformatics* 14(1):1.
25. Georgiades AS, Belhumeur PN, Kriegman DJ (2001) From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE transactions on pattern analysis and machine intelligence* 23(6):643–660.
26. Yu T, Ahuja N (2005) Simultaneous estimation of texture map and pseudo illumination from multiple views. *Urbana* 51:61801.
27. LeCun Y, Bengio Y (1995) Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361(10):1995.
28. Sobel I, Feldman G (1968) A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project* in pp. 271–272.
29. Dehak N, Kenny PJ, Dehak R, Dumouchel P, Ouellet P (2011) Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing* 19(4):788–798.
30. Hasan MR, Jamil M, Rahman MGRMS (2004) Speaker identification using mel frequency cepstral coefficients. *variations* 1:4.
31. Och FJ, Ney H (2002) Discriminative training and maximum entropy models for statistical machine translation in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. (Association for Computational Linguistics), pp. 295–302.
32. Gallil Z (1986) Efficient algorithms for finding maximum matching in graphs. *ACM Computing Surveys (CSUR)* 18(1):23–38.
33. Lee SH, Goddard ME, Visscher PM, van der Werf JH (2010) Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. *Genetics Selection Evolution* 42(1):1.
34. Craig DW, et al. (2011) Assessing and managing risk when sharing aggregate genetic variant data. *Nature Reviews Genetics* 12(10):730–736.
35. Savoye I, Loos R, Carels C, Derom C, Vlietinck R (1998) A genetic study of anteroposterior and vertical facial proportions using model-fitting. *The Angle Orthodontist* 68(5):467–470.
36. Carson E (2006) Maximum likelihood estimation of human craniometric heritabilities. *American Journal of Physical Anthropology* 131(2):169–180.
37. Liu F, et al. (2012) A genome-wide association study identifies five loci influencing facial morphology in europeans. *PLoS Genet* 8(9):e1002932.
38. Claes P, et al. (2014) Modeling 3D facial shape from DNA. *PLoS Genet* 10(3):e1004224.
39. Claes P, Hill H, Shriver MD (2014) Toward DNA-based facial composites: preliminary results and validation. *Forensic Science International: Genetics* 13:208–216.
40. Adhikari K, et al. (2016) A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nature communications* 7.
41. Paternoster L, et al. (2012) Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *The American Journal of Human Genetics* 90(3):478–485.
42. Liu F, Wen B, Kayser M (2013) Colorful DNA polymorphisms in humans in *Seminars in cell & developmental biology*. (Elsevier), Vol. 24, pp. 562–575.
43. Eiberg H, Mohr J (1995) Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. *European journal of human genetics: EJHG* 4(4):237–241.
44. Mushailov V, Rodriguez SA, Budimilija ZM, Prinz M, Wurmbach E (2015) Assay development and validation of an 8-SNP multiplex test to predict eye and skin coloration. *Journal of forensic sciences* 60(4):990–1000.
45. Walsh S, et al. (2013) The HirisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics* 7(1):98–115.
46. Bito LZ, Matheny A, Cruickshanks KJ, Nondahl DM, Carino OB (1997) Eye color changes past early childhood: the louisville twin study. *Archives of ophthalmology* 115(5):659–663.
47. Sturm RA, Duffy DL (2012) Human pigmentation genes under environmental selection. *Genome biology* 13(9):1.
48. Edwards M, et al. (2010) Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. *PLoS Genet* 6(3):e1000867.
49. McEvoy B, Beleza S, Shriver MD (2006) The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Human molecular genetics* 15(suppl 2):R176–R181.
50. Tishkoff SA, Verrelli BC (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annual review of genomics and human genetics* 4(1):293–340.
51. Sulem P, et al. (2007) Genetic determinants of hair, eye and skin pigmentation in europeans. *Nature genetics* 39(12):1443–1452.
52. Branicki W, Brudnik U, Kupiec T, Wolańska-Nowak P, Wojas-Pelc A (2007) Determination of phenotype associated SNPs in the MC1R gene. *Journal of forensic sciences* 52(2):349–354.
53. Rees JL (2004) The genetics of sun sensitivity in humans. *The American Journal of Human Genetics* 75(5):739–751.
54. Bonilla C, et al. (2005) The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. *Human genetics* 116(5):402–406.
55. Oetting WS, King RA (1999) Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism. *Human mutation* 13(2):99–115.
56. Stokowski RP, et al. (2007) A genomewide association study of skin pigmentation in a South Asian population. *The American Journal of Human Genetics* 81(6):1119–1132.
57. Hart KL, et al. (2013) Improved eye-and skin-color prediction based on 8 SNPs. *Croatian medical journal* 54(3):248–256.
58. Spichenok O, et al. (2011) Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Science International: Genetics* 5(5):472–478.
59. Maroñas O, et al. (2014) Development of a forensic skin colour predictive test. *Forensic Science International: Genetics* 13:34–44.
60. Wood AR, et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics* 46(11):1173–1186.
61. Locke AE, et al. (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518(7538):197–206.
62. Yang J, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* 42(7):565–569.
63. Klinefelter Jr HF, Reifstein Jr EC, ALBRIGHT JR F (1942) Syndrome characterized by gynecomastia, aspermatogenesis without A-Leydigism, and increased excretion of follicle-stimulating hormone 1. *The Journal of Clinical Endocrinology & Metabolism* 2(11):615–627.
64. Nyholt DR, Gillespie NA, Heath AC, Martin NG (2003) Genetic basis of male pattern baldness. *Journal of Investigative Dermatology* 121(6):1561–1564.
65. Richards JB, et al. (2008) Male-pattern baldness susceptibility locus at 20p11. *Nature genetics* 40(11):1282–1284.
66. Li R, et al. (2012) Six novel susceptibility loci for early-onset androgenetic alopecia and their unexpected association with common diseases. *PLoS Genet* 8(5):e1002746.
67. Lai CH, et al. (2013) Androgenic alopecia is associated with less dietary soy, higher blood vanadium and rs1160312 1 polymorphism in Taiwanese communities. *PLoS one* 8(12):e79789.
68. Qiao L, et al. (2016) Detecting genome-wide variants of eurasian facial shape differentiation: DNA based face prediction tested in forensic scenario. *bioRxiv* p. 062950.
69. Manichaikul A, et al. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* 26(22):2867–2873.
70. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829.