The Snapshot DNA Phenotyping pipeline was applied to each subject to predict global and regional ancestry using admixture analysis. The two lower-coverage subjects were both missing the SNP rs12913832, which is the primary SNP associated with eye color, so eye color and hair color were predicted assuming AA and AG genotypes, which are the most common genotypes in the Middle Eastern population. There were also very few skin color SNPs available for these two subjects, resulting in low-confidence predictions. The graphs below show the relative consistency of each possible phenotype category.

Three-dimensional face morphology was predicted by predicting the values of face PCs, which were then transformed into 3D objects. Each subject was compared to the previous subject and heat maps were calculated to show the differences among the face predictions. These differences were then emphasized to create caricatured faces, which were combined with the pigmentation predictions to create composites of the individuals’ likely appearance at age 25 by a forensic artist.

The low-coverage whole genome sequencing alignment file for subject HG00119 from the 1000 Genomes Project was downloaded, which has ~5.3X coverage. This data was randomly subsampled to 2.0X, 0.5X, and 0.05X. Low-coverage imputation was run on each subsample using a reference panel with subject HG00119 removed. Accuracy was determined by comparing the imputed genotypes to the genotypes in the 1000 Genomes phase 3 call set. As shown below, even at very low coverages, the vast majority of SNP genotypes can be accurately recovered by low-coverage imputation.

The low-coverage imputation pipeline using genotype likelihoods was implemented. This technique uses hundreds or thousands of linked SNPs to statistically infer the most likely genotype for each target SNP, resulting in much higher call rates. Instead, a low-coverage imputation pipeline was used, where the called SNPs are too sparse for this technique to operate successfully.

The low-coverage whole genome sequencing alignment file for subject HG00119 from the 1000 Genomes Project was downloaded, which has ~5.3X coverage. This data was randomly subsampled to 2.0X, 0.5X, and 0.05X. Low-coverage imputation was run on each subsample using a reference panel with subject HG00119 removed. Accuracy was determined by comparing the imputed genotypes to the genotypes in the 1000 Genomes phase 3 call set. As shown below, even at very low coverages, the vast majority of SNP genotypes can be accurately recovered by low-coverage imputation.

The two lower-coverage subjects were both missing the SNP rs12913832, which is the primary SNP associated with eye color, so eye color and hair color were predicted assuming AA and AG genotypes, which are the most common genotypes in the Middle Eastern population. There were also very few skin color SNPs available for these two subjects, resulting in low-confidence predictions. The graphs below show the relative consistency of each possible phenotype category.

Three-dimensional face morphology was predicted by predicting the values of face PCs, which were then transformed into 3D objects. Each subject was compared to the previous subject and heat maps were calculated to show the differences among the face predictions. These differences were then emphasized to create caricatured faces, which were combined with the pigmentation predictions to create composites of the individuals’ likely appearance at age 25 by a forensic artist.

The low-coverage whole genome sequencing alignment file for subject HG00119 from the 1000 Genomes Project was downloaded, which has ~5.3X coverage. This data was randomly subsampled to 2.0X, 0.5X, and 0.05X. Low-coverage imputation was run on each subsample using a reference panel with subject HG00119 removed. Accuracy was determined by comparing the imputed genotypes to the genotypes in the 1000 Genomes phase 3 call set. As shown below, even at very low coverages, the vast majority of SNP genotypes can be accurately recovered by low-coverage imputation.