Document Title:	**Evaluation of High Density SNP Microarrays to Obtain Phenotypic and Ancestry Information from Skeletal Remains**
Author(s):	**Giovanna M. Vidoli, Ph.D., Amy Mundorff, Jonathan Davoren**
Document Number:	**254407**
Date Received:	**December 2019**
Award Number:	**2014-DN-BX-K035**

Final Summary Overview

Department of Justice, Office of Justice Programs National Institute of Justice
Grant #2014-DN-BX-K035

Evaluation of High Density SNP Microarrays to Obtain Phenotypic and Ancestry Information from Skeletal Remains

Submitted by: Giovanna M. Vidoli, PhD
Research Assistant Professor, Anthropology
865-974-1303; gvidoli@utk.edu

Prepared by: Giovanna Vidoli & Amy Mundorff
University of Tennessee
&
Jonathan Davoren
Bode Technology

Submission Date: 12 November 2019

The University of Tennessee
1 Circle Park Drive
Knoxville, TN 37996-0003

Project Start Date: 01 Jan 2015 – End Date: 3 June 2019

Table of Contents

Summary of Research Project

There are few pathways to positively identify unknown or fragmentary skeletal remains if antemortem records are unavailable to make anthropological, genetic, or odontological comparisons between the unknown case and a known individual. This creates a significant backlog of unidentified "cold" cases in medical examiner and coroners' offices. The Snapshot™ Forensic DNA Phenotyping System offered by Parabon Nanolabs, Inc., was developed for the Defense Threat Reduction Agency (DTRA) and with support from the Department of Defense (DTRA R&D Small Business Innovation Research Phase I and Phase II grant awarded for "SNAPSHOT: A System for Predicting Human Physical Traits from Sample DNA"), with the goal of not only inferring sex, biogeographical ancestry, hair and eye color, but also with the potential to give inferences as to skin coloring, freckling, facial shape, and regional ancestry. The ability to predict these characteristics from unidentified skeletal remains could revolutionize forensic casework and methods of human identification. However, while the Snapshot™ Forensic DNA Phenotyping System is a robust product and may have potential for success using low template and degraded skeletal remains, it has not yet been tested with human bone, which typically contain degraded DNA as well as varying levels of microbial DNA. Therefore, the aim of this grant was to characterize the performance of a pre-configured panel of ancestry and phenotypically informative SNPs on DNA extracted from a controlled sample of well-documented skeletal cases. In addition, we compared the SNP results with the individual's self-reported data and our anthropological analysis of ancestry, which is what is traditionally used at medical examiner's offices to develop a biological profile from unidentified skeletal remains. The goal of this evaluation is to arm the medico-legal community and, especially, forensic crime laboratories with an additional tool to aid in the positive identification of missing persons and unidentified skeletal remains.

Project Subjects

This study utilized human remains obtained through the Forensic Anthropology Center (FAC) Body Donation Program at the University of Tennessee, Knoxville. The FAC receives approximately 100 human donations each year for research. As part of the donation process, the donors complete a Biological Questionnaire where the individual indicates his or her self-designated race (White, Black, Hispanic, or Other), their natural hair color, and their eye color (Blue, Green, Gray, Brown, Hazel, and Other). The complete Biological Questionnaire can be downloaded and viewed at http://fac.utk.edu/pdf/Questionnaire.pdf. After death, the individual's body is received by the FAC, their personal details are anonymized with a donation number and biological samples including blood cards, fingernails, and hair samples are obtained. The body is placed outdoors at the Anthropology Research Facility (ARF) to decompose naturally. To control for taphonomic differences between buried and surface decomposition, only individuals who decomposed on the ground surface were considered for this study. As a result, these individuals will have decomposed in a similar environment and been exposed to similar taphonomic processes. Following decomposition, the skeletal remains are recovered, cleaned, and allowed to air dry prior to accession into the William Bass Donated Skeletal Collection. The cleaning process is typically limited to simply rinsing the bones with warm tap water and removing any remnant adhering tissue.

Project Design and Methods

This research had five primary objectives:

1. To establish the amount of DNA from bone necessary for the HumanOmniExpressExome BeadChip genotyping.
2. To derive SNP genotype predictions of sex, biogeographic ancestry, hair color, and eye color from bone samples of identified individuals using the HumanOmniExpressExome BeadChip.

3. To test the reproducibility of the genotype results from the HumanOmniExpressExome BeadChip data.
4. To assess the predictive value of the chip for inference of ancestry and phenotypic variables from DNA samples derived from bone tissue.
5. To compare the estimates of ancestry predicted from the HumanOmniExpressExome BeadChip against those derived from traditional forensic anthropological assessment.

This research was divided into three distinct stages. Stage I comprised evaluating the sensitivity and accuracy of the HumanOmniExpressExome BeadChip using different DNA concentrations of blood and bone samples, from three recently skeletonized individuals, for overall quality scores. Including DNA samples from blood along with samples from bone allowed a direct comparison between the different sample types' results necessary to establish the accuracy and the call rate of the SNPs at different amounts of intact and degraded DNA. Five different amounts of DNA from blood and bone were tested (250, 100, 50, 20, and 10 Ng) to investigate how the results vary at lower levels of DNA compared to the recommended 250 Ng of DNA. While manufacturer recommendations indicate 250 Ng of DNA from soft tissue is sufficient, it had yet to be determined whether that amount, or lesser amounts, might be sufficient for whole genome amplification. Therefore, the DNA samples at 100, 50, 20, and 10 Ng were also tested with an additional whole genome amplification step to increase the amount of DNA.

During Stage II, which ran concurrently with Stage I, the ancestry of 25 individuals was estimated using standard forensic anthropological methods. Ancestry refers to the physical and genetic reflection of the accumulation of deep population histories including migrations, environment, etc. Thus, groups of people who have a shared geographic origin and population history share some common genetic material and phenotypic traits that can be measured using

skeletal metric and non-metric techniques. The major broad ancestry, or ethnic, groups addressed here include: European, African, Hispanic, and Asian, with Native Americans considered a subset of the Asian ancestry group.

The results from Stage I and II were used to inform Stage III. During Stage III, the ideal amount of DNA, as determined in Stage I, was used to extract samples from an additional 22 skeletons with varying self-reported ethnicities. Eight samples were run twice to assess genotype reproducibility. Stage III was designed to assess the chip-based inferences of ancestry and phenotype compared to the donor's self-reported information to evaluate the predictive value of the chip. In addition, ancestry inferences derived from the molecular analyses were compared with those developed by anthropological assessment in Stage II. This comparison was used to appraise the strength, limitation, accuracy, and cost effectiveness of this new technology to current anthropological methods of ancestry estimation from skeletal remains.

Anthropological Analysis

Two experienced forensic anthropologists (Co-PIs Vidoli and Mundorff) independently assessed ancestry using standard anthropological methods. Thirty-one caliper based cranial measurements were taken following the standards outlined in *Data Collection Procedures for Skeletal Material* (Moore-Jansen *et al* 1994). The cranial measurements were entered in FORDISC 3.0 and the resulting statistics were analyzed for group classification. FORDISC (Jantz and Ousley, 2005) is the primary tool for metric ancestry assessment and the statistical basis for FORDISC is discriminant function analysis. In addition, morphometric (non-metric) traits on the skull were assessed and recorded for each individual. Morphological features of the skull (cranial non-metric traits or macromorphoscopic traits) are heritable and vary among and between human populations and are therefore useful in the assessment of ancestry. However, like metric traits, only broad

categories can be assessed (i.e., White, Black, Asian), and some methods are limited to discerning between Black and White only. Six non-metric traits, as outlined in Hefner and Ousley (2014), were scored and entered into either OSSA or HefneR (http://osteomics.com/hefneR/). The main difference between OSSA and HefneR is that the former provides only Black and White as categories while the later provides Black, White, Native American, and Asian. In addition, the statistics for each method differs.

DNA

Using varying DNA concentrations from blood and bone, the chip performance was evaluated with quality control checks and with genotyping performance results to determine the optimal amount for bone. Including DNA samples from blood along with samples from bone allowed a direct comparison between the different sample types' results necessary to establish the accuracy and the call rate of the SNPs at different amounts of intact and degraded DNA. Five different amounts of DNA from blood and bone were tested (250, 100, 50, 20, and 10 Ng) to investigate how the results vary at lower levels of DNA compared to the recommended 250 Ng of DNA. While manufacturer recommendations indicate 250 Ng of DNA from soft tissue is sufficient, it had yet to be determined whether that amount, or lesser amounts, might be sufficient for whole genome amplification. Therefore, the DNA samples at 100, 50, 20, and 10 Ng were also tested with an additional whole genome amplification step to increase the amount of DNA.

DNA was isolated from blood using the Qiagen® EZ1® DNA Investigator Kit on a Qiagen Biorobot EZ1. DNA from skeletal samples was isolated using a standard demineralization process followed by cleanup using the Qiagen QIAmp micro kit. DNA was quantified using the Life Technologies Quantifiler Trio system. As skeletal samples are typically variable for DNA yields, extractions were performed until at least 700 Ng total DNA was recovered.

SNP testing

The extracted DNA samples were assayed for SNP genotypes using the Illumina human Omni Express Exome system using the manufacturer recommended procedure and an Illumina iScan instrument. The raw data was analyzed using the Illumina Genome Studio software using the recommended settings.

Snapshot

The Snapshot solution requires the use of Illumina's HumanOmniExpressExome BeadChip, which examines SNPs from all three HapMap phases to capture the greatest amount of common SNP variation. This BeadChip includes over 273,000 functional exonic markers, and the Snapshot technology has shown success with this array with as little as ***50pg*** of genomic DNA (Greytak 2014a, 2014b). The Snapshot solution, developed by Parabon, uses machine learning model software that depicts predictive values along a continuum from the smallest observed prediction values for the trait to the largest (Greytak 2014a). Once SNPs are determined to be significant, they are combined into a predictive model for each trait using advanced machine learning methods. When a DNA sample is queried through this predictive model for each trait, they are converted to a percentile and then compared against the distribution of the observed values for each possible category of the trait. For example, the predictive model for an unknown individual's eye color is compared against the distribution for blue eyes and its consistency (0-100%) is measured according to where it falls. This is repeated for green, brown, etc., and the category with the highest consistency is reported as the predicted phenotype. Any category with consistency of less than 5% is considered an excluded phenotype. Because the values are not independent of one another, the consistency values across the categories will not equal 100%. For ancestry, the Snapshot solution offers estimations across seven populations: African, Middle Eastern, European, Central Asian, East Asian, Oceanian, and Native American (Greytak 2014b).

The Snapshot report will depict ancestry as the proportional membership in each ancestral population, such that the total will equal 100%. The data is displayed on both a global map with relative contribution of each population, and by percentage. The Snapshot DNA Phenotyping solution also has the potential to give regional ancestry (up to 28 regions within the seven major global ancestry groups), and the results are displayed on a plot detailing how the unknown individual clusters with subjects from well-established ancestral populations.

Data Analysis

The bone samples tested were approximately 6 years post-mortem and the blood collection occurred within a week death and was stored on blood cards for approximately 6 years. The bone samples had an average degradation index around 1.5 and the blood specimens G and E had a degradation index of 1 indicating little degradation had occurred. The DNA extracted from specimen Z had a degradation index of 3 indicating some degradation. All samples tested produced STR profiles and only the blood from specimen Z showed any indications of degradation.

SNP Results – Stage I

The SNP call rates for specimens G and E were both ~98.8% for blood at 250 Ng of input DNA while the bone gave ~86 and ~90%, respectively (Table A1). At the minimum DNA quantity tested, 10 Ng, the blood specimens G and E provided 71.9% and 89% call rates with 56,565 and 12,781 discordant genotypes, respectively. As the DNA was increased the call rate increased and the discordant genotypes fell (Table 1). With 250 Ng of DNA, from blood, specimens G and E both had 99.8 % call rates. For the DNA from bone specimen G had more than a 70% call rate for both 250 Ng and 100 Ng. The DNA from bone specimen E had more than 70% call rate only when 250 Ng was tested. When the DNA from bone was at less than 100 Ng for specimen G or

at less than 250 Ng for specimen E the call rates were around 50% or less and nearly half of the called genotypes were discordant with the results of the DNA from blood. The data quality, for specimen Z, was too low to make any conclusions on the results. The cause of the poor data quality from the blood in specimen Z needs to be investigated further but could be related to the degradation. In addition, due to the poor data quality, this individual was replaced in Stage III.

Table 1. SNP call rates and call differences across a range of input DNA.

	G - Blood		E - Blood		G - Bone		E - Bone	
Template (Ng)	Call Rate	Genotype Differences vs 250 Ng	Call Rate	Genotype Differences vs 250 Ng	Call Rate	Genotype Differences vs 250 Ng	Call Rate	Genotype Differences vs 250 Ng
250	99.8%	--	99.8%	--	86.1%	21,054	89.9%	13,608
100	83.7%	15,708	98.1%	640	72.4%	59,874	52.7%	229,133
50	81.4%	22,621	96.0%	1,097	50.2%	241,863	48.3%	260,943
20	75.0%	41,065	93.4%	3,071	63.5%	361,413	39.2%	210,140
10	71.9%	56,565	89.1%	12,781	32.3%	173,303	29.8%	163,351

The called SNPs for specimens E and G between 10 and 100 Ng of input DNA were at least 91.8% reproducible when compared to the blood sample at 250 Ng of input DNA. The called SNPs for specimen G at 100 and 250 Ng of bone DNA and specimen E at 250 Ng of input DNA were at least 91.4% reproducible when compared to the blood sample at 250 Ng of input DNA. As the input DNA was reduced the number of discordant allele calls increases especially for the bone specimens where up to approximately 57% of the allele calls were discordant.

Phenotyping – Stage I

The phenotyping and ancestry predictions for specimens G and E were consistent with the self-reported information for all blood samples. The DNA from bone was also consistent with the self-reported information but only when 100 and 250 Ng was tested for specimen G and 250 Ng was tested for specimen E. When the SNP call rates were below 70% the predictions were either not

significant or they differed from the expected results (Table A2). The poor predictions at lower levels of DNA from bone were expected due to the lower SNP call rates and half of the called SNP genotypes being incorrect.

SNP Results – Stage III

A total of 250 Ng of DNA was used for 29 samples and 100 Ng was used for 2 samples in Stage III. From the 31 samples tested 14 had call rates above the 70% while the remaining were lower down to a 48% call rate (Table 2). The 6 samples tested in duplicate showed discordance levels of 0.07 – 0.94% of the SNPs (Table 3). The discordant genotypes generally increased for samples with lower call ranges.

Table 2. SNP call rates for the 31 samples tested in Stage III.

Sample	Call Rate	Sample	Call Rate	Sample	Call Rate
E_2	95.7%	G_1	67.60%	M	61.8%
D_1	94.3%	X	67.00%	L	60.3%
R	94.1%	T	66.90%	U	58.5%
E_1	93.9%	J	71.8%	I	58.1%
D_2	93.5%	F	69.4%	S_1	57.6%
P_1	81%	O	66.5%	V	57.4%
P_2	80.1%	Y	65.9%	S_2	56.3%
H_1	74.7%	C	65.3%	B	55.7%
H_2	74.7%	Q	65.2%	W	53.4%
A	72.7%	K	61.8%	N	48.4%
G_2	70.1%				

Table 3. Comparison of SNP calls for duplicated samples

Sample pair	SNP call differences	SNPs where both failed	SNPs where 1 was called and 1 failed
G	0.5262%	20.6%	21.0%
H	0.3229%	17.2%	16.1%
P	0.2603%	12.9%	13.0%
S	0.9419%	28.0%	30.2%
D	0.0711%	3.7%	4.8%
E	0.0769%	3.0%	4.5%

Phenotyping – Stage III

Self-reported race was compared with the SNP and anthropological data (Table A4). Unfortunately, due to low SNP call rates (Table 2), phenotypic predictions were only available for 14 samples. The 14 samples with phenotyping results were from specimens A, D, E, G, H, P, R, T, and X as 5 were processed in duplicate. For 11 of the individuals there was agreement between the self-reported ancestry, the SNP predicted ancestry, and the ancestry predicted by

Craniometrics. The disagreement among the self-reported, SNP, and craniometrics in three samples (samples B, L, and O), reflects the difficulty of assessing the ancestry of individuals who self-identify as Hispanic (sample B), who have mixed ancestry (sample L), or also of how FORDISC is applied and interpreted (sample O). In addition, these samples had low call rates (<70%). On the other hand, Sample X self-identified as Hispanic, had SNP predictions that were from the Americas, and FORDISC categorized this individual Guatemalan. That is not to say the individual was from Guatemala but rather that his facial and cranial metrics were consistent with someone from Central America. The agreement among the 3 data sets demonstrates that even with more difficult samples, ancestry can be correctly predicted.

Findings

- The SNP chip testing results had large significant drop in accuracy when the call rates are below 70%. Below a 70% call rate the accuracy becomes so low that the phenotype and ancestry predictions become uninformative.
- DNA from blood samples was gave accurate phenotype and ancestry predictions at 10 Ng of input DNA however bones required at least 250 Ng of DNA to have a call rate over 70%.
- Phenotypic information was determined for DNA from approximately 50% of the bone samples and 67% of the blood samples at a 6-year post mortem interval.
- Anthropological assessments were mostly consistent with self-reported ancestry. However, the non-metric assessments, especially using OSSA, which has limited categories, did not capture the phenotypic variation in human groups. While the sample size was smaller than expected, there was overall concordance between the genetic and anthropological ancestry predictions.

Implications for Criminal Justice Policy and Practice in the United States

The greatest potential gain from this research is the demonstration of a new probative tool to assist with identifying human skeletal remains for the forensic and criminal justice community. This research will enhance forensic science practice for criminal justice purposes. Applying the Illumina's HumanOmniExpressExome BeadChip in a medico-legal context will provide the criminal justice system and forensic science practitioners with an additional tool to aid in the identification of skeletal remains, which typically lack enough probative identifying information to match with a missing person's report. In lieu of other investigative leads, additional genetic information extracted from bone samples will improve identification efforts of skeletal remains. In particular, the data derived from the molecular analysis of bone is critically valuable for instances of partial or fragmentary remains (fleshed and skeletal) and juvenile remains, increasing the identification potential of the most challenging forensic cases. As a result, the number of unidentified remains in medical examiner and coroner's offices across the country, the time to positively identify unidentified skeletal remains, and, consequently, the financial expenditures, would all decrease. The scientific output generated by the forensic application of this chip will also help inform law enforcement on questions of social identity frequently given by families for missing person's reports. As such, improved policies regarding what informative (phenotypic) data is gathered from family members of missing individuals will ensure that the genetic data generated from a skeleton's DNA can more easily be matched to a missing person's file because they both contain in-common categorical choices. In addition, SNP genotyping on skeletal remains will impact criminal justice practice regarding the examination, DNA sampling, and recording of skeletal remains, particularly juvenile skeletal remains. There is currently no method to determine sex, ancestry, or certain physical characteristics in juvenile skeletal remains and one of the greatest

potential contribution of the application of the Illumina's HumanOmniExpressExome BeadChip, will be to provide otherwise inaccessible information on decomposed or skeletonized remains of minors and children. Finally, the ability to gain phenotypic information for fragmentary remains would facilitate identification in mass fatality incidents with high levels of fragmentation. This research also demonstrated limitations to this technology. Phenotypic predictions were not possible for a number of samples because the necessary SNP call rates could not be obtained.

Dissemination of Research Findings

Results from this study will be presented at the American Academy of Forensic Sciences Annual Conference in February 2020, and a manuscript is in preparation for submission to Forensic Science International: Genetics. Law enforcement's interest in using SNP chips has grown rapidly in the past year due to their ability to find relatedness beyond 3rd cousins. As such, the results of this project may affect practice and policy governing the handling and examination of unidentified skeletal remains by medico-legal agencies across the United States. Therefore, we also intend to submit an article focusing on skeletal identification management considerations to Forensic Science, Policy, and Management. Finally, best practice procedures in corroborating genetic with anthropological data to broaden the sphere of information for skeletal remains will be made available for training of medico-legal or law enforcement personnel involved in missing person's cases and identification of unidentified human remains.

Appendices

Citations

Greytak, Ellen McRae (2014a) *Using Genome-wide SNPs for DNA Phenotyping and Kinship Inference.* Poster presented at the 25th International Symposium on Human Identification, Phoenix, AZ, September 2014

Greytak, Ellen McRae (2014b) *DNA Phenotyping with Parabon® Snapshot™ Predicting Physical Appearance from Crime Scene DNA.* Poster presented at the International Association of Chiefs of Police 121st Annual Conference and Exposition, Orlando, FL, October 21, 2014

Hefner JT, Ousley SD. (2014) Statistical classification methods for estimating ancestry using morphoscopic traits. *Journal of Forensic Sciences* 59(4):883-90.

Jantz, RL and Ousley SD (2005) FORDISC 3: Computerized Forensic Discriminant Functions. Version 3.0. The University of Tennessee, Knoxville**.**

Moore-Jansen, P.M., S.D. Ousley, and R.J. Jantz.(1994) *Data Collection Procedures for Forensic Skeletal Material*. Report of Investigations No.48. Department of Anthropology, University of Tennessee, Knoxville.

Tables

Table A1. The Phenotyping and ancestry predictions across a range of input DNA from blood and bone during Stage I.

			Large Autosomal		Small Autosomal		Y - Male					
	Well	Sample Name	CT	Quantity (ng/µl)	CT	Quantity (ng/µl)	CT	Quantity (ng/ul)	IPC Ct	Male Quantity/ Human Quantity	Degradation Index	Total DNA (Ng)
Bone	A4	E1	23.6	**1.46**	25.4	**2.02**	24.1	**1.90**	27.9	94.00%	1.38	72.9
	B4	E2	22.5	**2.97**	24.1	**5.17**	23.4	**3.03**	28.2	58.62%	1.74	148.3
	C4	E3	21.3	**6.73**	22.8	**12.59**	21.6	**10.30**	28.9	81.80%	1.87	336.6
	H4	E4	23.3	**1.79**	25.0	**2.63**	23.6	**2.66**	28.0	100%	1.47	89.5
	C5	E5	25.2	**0.49**	27.1	**0.63**	25.8	**0.63**	27.9	100%	1.28	24.6
	F5	E6	25.7	**0.36**	27.3	**0.55**	26.1	**0.50**	27.8	91.15%	1.54	17.8
	F3	G1	24.7	**0.67**	26.4	**0.99**	25.0	**1.05**	27.6	100%	1.48	33.5
	G3	G2	23.6	**1.45**	25.5	**1.89**	24.1	**1.87**	27.9	99.27%	1.30	72.3
	H3	G3	22.4	**3.20**	24.2	**4.62**	22.8	**4.54**	27.8	98.34%	1.44	160.2
	G4	G4	24.3	**0.90**	26.1	**1.24**	24.9	**1.17**	27.7	94.38%	1.37	45.1
	B5	G5	26.0	**0.29**	27.7	**0.41**	26.5	**0.40**	27.5	98.18%	1.41	14.4
	E5	G6	26.8	**0.17**	28.4	**0.25**	27.0	**0.28**	27.3	100%	1.53	8.3
	Z2	Z1	25.0	**0.56**	26.7	**0.84**	25.4	**0.79**	27.8	94.29%	1.50	27.9
	E4	Z2	24.0	**1.10**	25.6	**1.74**	24.4	**1.54**	27.7	88.48%	1.58	55.1
	F4	Z3	23.3	**1.80**	24.8	**3.06**	23.6	**2.65**	27.8	86.70%	1.70	90.1
	A5	Z4	24.4	**0.85**	26.1	**1.26**	24.8	**1.20**	27.8	95.09%	1.48	42.5
	D5	Z5	27.1	**0.13**	29.0	**0.16**	27.7	**0.17**	27.4	100%	1.24	6.5
	G5	Z6	27.2	**0.12**	29.1	**0.15**	27.9	**0.16**	27.4	100%	1.20	6.2
Blood	B6	E_01	18.2	**56.04**	21.0	**45.39**	19.5	**40.89**	29.0	90.09%	1.00	2802.2
	C6	E_02	17.8	**73.49**	20.5	**61.64**	19.2	**51.76**	28.9	83.97%	1.00	3674.7
	H5	G_01	23.3	**1.75**	26.2	**1.18**	24.6	**1.35**	28.1	100%	1.00	87.5
	A6	G_02	23.0	**2.23**	25.3	**2.21**	24.0	**2.03**	28.1	91.88%	1.00	111.3
	D6	Z_01	23.1	**2.01**	23.8	**6.04**	22.4	**6.00**	28.4	99.38%	3.01	100.4
	E6	Z_02	23.0	**2.18**	23.5	**7.47**	22.2	**7.06**	28.0	94.51%	3.43	109.0

Table A2. Phenotyping and ancestry predictions for a range of template amounts for samples E and G from blood and bone.

	Sample Template	Skin color	confidence	Eye Color	confidence	Hair Color	confidence	Freckles	confidence	Sex	Ancestry	
Blood DNA	G	Fair / Very Fair	80.4%	Green/Blue	72%	Blond / Brown	70.9%	Few / Some	70.8%	M	NW Europe	91.1%
	250 Ng	Not: Dark Olive / Dark	94.3%	Not: Brown / Black	99.2%	Not Black	94.7%				NE Europe	7.4%
	G	Fair / Very Fair	96.4%	Green/Hazel	92.4%	Blond / Red	72.4%	Few / Some	75%	M	NW Europe	95.2%
	100 Ng	Not: light Olive / Dark Olive / Dark	96.4%	Not: Blue / Brown / Black	92.4%	Not Black	94.7%				NE Europe	4.8%
	G	Fair / Very Fair	94.6%	Green/Hazel	75.7%	Blond / Red	86.3%	Few / Some	79.2%	M	NW Europe	90.2%
	50 Ng	Not: light Olive / Dark Olive / Dark	94.6%	Not: Brown / Black	99.3%	Not Black	97.9%				NE Europe	6.9%
	G	Fair / Very Fair	98.2%	Green/Blue	76.2%	Brown / Red	74.4%	Few / Some	75%	M	NW Europe	96.4%
	25 Ng	Not: light Olive / Dark Olive / Dark	98.2%	Not: Brown / Black	99.3%	Not Black	90.5%				NE Europe	3.6%
	G	Fair / Very Fair	91.1%	Green/Hazel	87.4%	Blond / Red	89.4%	Few / Some	79.2%	M	Central East Europe	100%
	10 Ng	Not: light Olive / Dark Olive / Dark	91.1%	Not: Brown / Black	98.8%	Not Black	99.5%					
Bone DNA	G	Fair / Very Fair	78.6%	Green/Blue	76.4%	Blond / Red	81.7%	Few / Some	45.8%	M	NW Europe	82.2%
	250 Ng	Not: Dark Olive / Dark	94.3%	Not: Brown / Black	99.3%	Not Black	96.8%	Not Zero	91.7%		Caucasus	10.4%
											NE Europe	7.4%
	G	Fair / Light Olive	73.9%	Green/Blue	72.4%	Blond / Red	78.4%	Few / Some	70.8%	M	NW Europe	78.8%
	100 Ng	Not: Dark Olive / Dark	91.4%	Not: Brown / Black	98.9%	Not Black	96.8%				East Europe	8.2%
											Caucasus	7.6%
	G	Dark Olive / Light Olive	85.7%	Hazel / Green	93.1%	Brown / Blond	75.5%	Some / Many	1.9%	M	Europe	51%
	50 Ng	Not: Fair or Very Fair	99.7%	Not: Brown / Blue / Black	93.1%	Not Black	91.1%	Not Zero	99.99%		Africa	27%
											Cent/South Asia	9%
											East Asia	8%
	G	Dark Olive / Light Olive	92.9%	Hazel / Green	91.9%	Blond / Red	86.2%	Some / Many	19.8%	M	Europe	45%
	25 Ng	Not: Dark or Very Fair or Fair	92.9%	Not: Brown / Blue / Black	91.9%	Not Black	96.8%	Not Zero	99.99%		Africa	29%
											Cent/South Asia	11%
											East Asia	9%
	G	Dark Olive / Light Olive	85.7%	Hazel / Green	77.5%	Blond / Brown	70.2%	Some / Many	32.1%	M	Europe	49%
	10 Ng	Not: Very Fair / Fair	99.1%	Not Black	99.99%	Not Black	93.2%	Not Zero	99.99%		Africa	28%
											East Asia	9%
											Central Asia	8%

	Sample Template	Skin color	confidence	Eye Color	confidence	Hair Color	confidence	Freckles	confidence	Sex	Ancestry	
Blood DNA	E	Very Fair / Fair	94.6%	Hazel / Brown	92.8%	Brown / Black	98%	Few / Some	12.5%	M	NW Europe	84.4%
	250 Ng	Not: Light Olive / Dark Olive / Dark	94.6%	Not: Green / Blue / Black	92.8%	Not: Red / Blond	98%	Not Zero	99.3%		Central W Europe	11.8%
	E	Very Fair / Fair	94.6%	Hazel / Brown	97%	Brown / Black	98%	Few / Some	17.5%	M	NW Europe	94.7%
	100 Ng	Not: Light Olive / Dark Olive / Dark	94.6%	Not: Green / Blue / Black	97%	Not: Red / Blond	98%	Not Zero	99%		SW Europe	5.1%
	E	Very Fair / Fair	96.4%	Hazel / Brown	90.8%	Brown / Black	98.7%	Few / Some	16.7%	M	NW Europe	92%
	50 Ng	Not: Light Olive / Dark Olive / Dark	96.4%	Not: Green / Blue / Black	90.8%	Not: Red / Blond	98.7%	Not Zero	99.0%		SW Europe	4%
											CW Europe	4%
	E	Very Fair / Fair	94.6%	Hazel / Brown	83.5%	Brown / Black	85.5%	Few / Some	29.2%	M	NW Europe	100%
	20 Ng	Not: Light Olive / Dark Olive / Dark	94.6%	Not Black	99.99%	Not Red	94.7%	Not Zero	93.1%			
	E	Very Fair / Fair	99.99%	Green / Hazel	93.2%	Brown / Black	88.5%	Few / Some	50.0%	M	NW Europe	100%
	10 Ng	Not: Light Olive / Dark Olive / Dark	99.99%	Not: Blue/Brown/Black	93.2%	Not Red	90.7%					
Bone DNA	E	Very Fair / Fair	96.4%	Hazel / Brown	89.5%	Brown / Black	93.4%	Few / Some	16.7%	M	NW Europe	86.3%
	250 Ng	Not: Light Olive / Dark Olive / Dark	96.4%	Not: Blue / Black	97.9%	Not: Red / Blond	93.4%	Not Zero	98.3%		SW Europe	10.6%
		Light Olive / Dark Olive	97.6%	Hazel / Green	91.1%	Brown / Red	77.3%	Some / Many	61.3%	M	Europe	55%
	E	Not: Very Fair / Fair / Dark	97.6%	Not: Brown / Blue / Black	91.1%	Not: Black	90.0%	Not Zero	99.99%		Africa	26%
	100 Ng										East Asia	8%
											Central Asia	6%
		Light Olive / Dark Olive	95.8%	Hazel / Brown	74.1%	Brown / Red	99.5%	Some / Many	73.1%	M	Europe	49%
	E	Not: Very Fair / Fair / Dark	95.8%	Not: Blue / Black	90.2%	Not: Black / Brown	99.5%	Not Zero	99.99%		Africa	29%
	50 Ng										East Asia	9%
											Central Asia	7%
		Dark Olive / Light Olive	92.9%	Hazel / Green	79.2%	Blond / Brown	71.5%	Few / Some	0.0%	M	Europe	49%
	E	Not: Dark / Very Fair / Fair	92.9%	Not: Black	99.99%	Not Black	93.2%	Not Zero	99.99%		Africa	28%
	20 Ng										Cent/South Asia	9%
											East Asia	8%
		Light Olive / Dark Olive	96.5%	Hazel / Brown	78.3%	Brown / Red	92.7%	Some / Many	0.5%	M	Europe	50%
	E	Not: Fair / Very Fair / Dark	96.5%	Not: Black	99.99%	Not: Black / Brown	92.7%	Not Zero	99.3%		Africa	28%
	10 Ng										East Asia	8%
											Central Asia	7%

Table A3. Phenotyping and ancestry predictions from 250 Ng of DNA extracted from bone.

	Sample	Skin color	confidence	Eye Color	confidence	Hair Color	confidence	Freckles	confidence	Sex	Ancestry	
Bone DNA	G_1	Fair / Very Fair	88.6%	Green / Blue	82.5%	Brown / Blond	89%	Few / Some	4.2%	M	North European	80.4%
		Not: Brown / Dark Brown	99.99%	Not: Brown / Black	99.4%	Not Reddish	95.2%	Not Zero	99.7%		West Africa	6.78%
											West - Central Asia	5.78%
	G_2	Fair / Very Fair	90.4%	Blue / Green	88.2%	Blond	98.2%	Few / Some	33.3%	M	North European	73.63%
		Not: Light Brown / Brown / Dark Brown	90.4%	Not: Brown / Black	99.6%	Not: Brown / Black	98.2%	Not Zero	96.2%		West Africa	10.25%
											Central Asia	5.84%
	H_1	Brown / Dark Brown	95.8%	Brown / Black	50.1%	Reddish	98.8%	Zero / Few	71.9%	M	West Africa	56.3%
		Not: Lt Brown / Fair / Very Fair	95.8%	Not Blue / Green	99.99%	Black	99.99%				South Africa	20.5%
						Not Blond / Brown	99.99%				North European	9.3%
	H_2	Brown / Dark Brown	95.8%	Brown / Black	50.1%	Reddish	95.4%	Few / Many	52.6%	M	West Africa	57.6%
		Not: Lt Brown / Fair / Very Fair	95.8%	Not Blue / Green	99.99%	Black	99.99%				South Africa	18.8%
						Not Blond / Brown	99.99%				North European	9.6%
	P_1	Fair / Very Fair	88.6%	Hazel / Green	95.2%	Brown / Black	91.2%	Few / Zero	78.9%	F	North European	94.2%
		Not: Brown / Dark Brown	99.3%	Not: Brown / Blue / Black	95.2%	Not Blond	91.2%					
	P_2	Fair / Very Fair	90.4%	Hazel / Green	94.5%	Brown / Black	88.9%	Few / Some	71.9%	F	North European	92.2%
		Not: Brown / Dark Brown	90.4%	Not: Brown / Blue / Black	94.5%							
	D_1	Fair / Very Fair	82.0%	Hazel / Green	95.9%	Reddish	95.9%	Few / Some	78.8%	M	North Europe	81.2%
		Not: Brown / Dark Brown	99.3%	Not: Blue / Brown / Black	95.9%	Black / Brown	99.8%				Southeast Europe	16.8%
						Not Blond	99.8%					
	D_2	Fair / Very Fair	82.0%	Hazel / Green	96%	Reddish	96.1%	Few / Some	78.8%	M	North Europe	81.9%
		Not: Brown / Dark Brown	99.3%	Not: Blue / Brown / Black	96%	Brown / Black	99.8%				Southeast Europe	16.7%
						Not Blond	99.8%					
	E_1	Fair / Very Fair	91%	Hazel / Brown	91%	Brown / Black	99.8%	Some / Few	8.3%	M	North European	96.8%
		Not: Light Brown / Brown / Dark Brown	91%	Not: Green / Blue / Black	91%	Not Blond	99.8%	Not Zero	99.3%			
	E_2	Fair / Very Fair	90.4%	Hazel / Brown	90.8%	Brown / Black	99.8%	Some / Few	8.3%	M	North European	96.2%
		Not: Light Brown / Brown / Dark Brown	90.4%	Not: Green / Blue / Black	90.8%	Not Blond	99.8%	Not Zero	99.3%			
	A	Fair / Very Fair	89.9%	Blue / Green	87.2%	Reddish	99.8%	Few / Many	5.3%	F	North Europe	57.1%
		Not: Brown / Dark Brown	99.99%	Not: Brown / Black	99.6%	Brown / Blond	96.3%	Not Zero	99.7%		Middle East	14.7%
						Not Black	96.3%				North East Europe	14.6%
	R	Brown / light Brown	93.5%	Brown / Black	55%	Black	99.9%	Zero / Few	96.5%	M	West Africa	51.99%
		Not: Bark Brown / Fair / Very Fair	93.5%	Not Blue / Green	99.99%	Not Brown / Blond	99.9%	Not: Some / Many	96.5%		North Europe	34.4%
											East Africa	9.1%
	T	Fair / Very Fair	80.8%	Blue / Green	89.5%	Blond / Brown	96.3%	Some / Few	8.3%	M	North Europe	79.6%
		Not: Brown / Dark Brown	99.30%	Not: Brown / Black	99.4%	Not Black	96.3%	Not Zero	99.7%		Western Middle East	11.4%
	X	Light Brown / Brown	97.2%	Brown / Black	50.1%	Reddish	97.9%	Few / Some	50.0%	M	Central America	52.7%
		Not: Fair / Dark Brown / Very Fair	97.2%	Not: Blue / Green	99.99%	Black	99.99%				South America	22.4%
						Not: Blond / Brown	99.99%				Brazil	6.3%

Table A4. Self-reported, SNP, and anthropological ancestry results

Sample	Self reported	SNP	Craniometrics	Non-metric
A	White	Europe	White	White
B	Hispanic	Europe	Black	
C	White	No data	White	White
D	White	Europe	White	American Indian
E	White	Europe		
F	Hispanic	No data	White	White
G	White	Europe	White	Other
H	Black	Africa	Black	Black
I	White	No data	White	White
J	Black	No data	Black	Black
K	Black	Africa	Black	Black
L	White/Asian	mixed	White	White
M	White	No data	Hispanic	White
N	White	No data	White	White
O	White	Europe	American Indian	White
P	White	Europe	White	White
Q	White	Europe	White	White
R	Black	Africa	Black	Black
S	White	No data	White	White
T	White	Europe	White	White
U	White	No data	White	White
V	Black	No data	White	White
W	White/American Indian	No data	White	American Indian
X	Hispanic	Americas	Guatemalan	Black
Y	White	Europe	White	White