ARTICLE

The Contribution of Neanderthals to Phenotypic Variation in Modern Humans

Michael Dannemann[1] and Janet Kelso[1,*]

Assessing the genetic contribution of Neanderthals to non-disease phenotypes in modern humans has been difficult because of the absence of large cohorts for which common phenotype information is available. Using baseline phenotypes collected for 112,000 individuals by the UK Biobank, we can now elaborate on previous findings that identified associations between signatures of positive selection on Neanderthal DNA and various modern human traits but not any specific phenotypic consequences. Here, we show that Neanderthal DNA affects skin tone and hair color, height, sleeping patterns, mood, and smoking status in present-day Europeans. Interestingly, multiple Neanderthal alleles at different loci contribute to skin and hair color in present-day Europeans, and these Neanderthal alleles contribute to both lighter and darker skin tones and hair color, suggesting that Neanderthals themselves were most likely variable in these traits.

Introduction

Interbreeding between Neanderthals and early modern humans has been shown to have contributed about 2% Neanderthal DNA to the genomes of present-day non-Africans. This Neanderthal DNA has apparently had both positive and negative effects. Together with the rapid decrease in Neanderthal ancestry after introgression, the depletion of Neanderthal DNA around functional genomic elements in present-day human genomes suggests that a large fraction of Neanderthal alleles are deleterious in modern humans.[1–4] However, recent studies have also identified a number of introgressed Neanderthal alleles that have increased in frequency in modern humans and that might contribute to genetic adaptation to new environments. Adaptive variants in genes related to immunity, skin and hair pigmentation, and metabolism have been identified.[4–11]

The majority of Neanderthal alleles in the genomes of people today are, however, not strongly adaptive and are therefore present at low frequencies (<2%) in present-day populations. To date, the number of individuals for whom genotype and phenotype information is available has been limited, making it difficult to study archaic alleles that are at such low frequencies or to link them to phenotypic variation. A recent study used the electronic medical records and genotypes of 28,000 individuals to address the contribution of these less frequent Neanderthal alleles to clinical traits in modern humans. It showed that a large number of Neanderthal variants at different loci influence risk of a number of disease traits, including depression, skin lesions, and blood-clotting disorders, and that Neanderthals contributed both risk and protective alleles for these traits.[12] However, evaluating the broader contribution of Neanderthals to common phenotypic variation in modern humans, or inferring Neanderthal phenotypes, has not been possible largely because of the limited number of studies that collect genotype data together with common phenotype information.

In addition to collecting genotype data via a custom genotyping array, the UK Biobank has collected baseline phenotypes, including traits related to physical appearance, diet, sun exposure, and behavior, as well as disease, for more than 500,000 people.[13] The pilot dataset including genotypes and phenotypes for more than 150,000 of the individuals was recently made available for study. Using these data, we studied the contribution of Neanderthals to common human phenotypic variation in 112,338 individuals from the UK Biobank to determine the set of traits to which Neanderthals have contributed and to evaluate the relative contribution of archaic and non-archaic alleles to common phenotypic variation in modern humans.

Material and Methods

Datasets from the UK Biobank

We obtained genotype and phenotype data from the pilot phase of the UK Biobank project.[13] Genotyping was performed with two arrays (UK BiLEVE and UK Biobank Axiom) that share 95% of markers, resulting in a merged dataset with genotype information for 152,729 individuals across 822,111 genomic sites.

Filtering Genotype Data

UK Biobank quality control (QC) included tests for batch, array, plate, and sex effects, as well as departures from Hardy-Weinberg equilibrium and discordance across control replicates. We used information provided by the UK Biobank to remove a total of 40,391 individuals; of these, 480 were related according to a kinship inference analysis, 17,308 had significantly decreased heterozygosity levels, and 32,443 had substantial non-European ancestry according to self-reported information and a principle-component analysis of the SNP data. Extensive documentation of the QC for these data is available on the UK Biobank's website.

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany
*Correspondence: kelso@eva.mpg.de
https://doi.org/10.1016/j.ajhg.2017.09.010.
© 2017 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Annotating Non-archaic and Archaic-like SNPs

A total of 825,927 polymorphic sites were genotyped. We took a two-step approach to annotate SNPs on the basis of whether they carried an allele of putative archaic origin. First, we identified potentially introgressed alleles by selecting SNPs that had one fixed allele in Yoruba individuals, an African population with little to no inferred Neanderthal DNA (1000 Genomes Project[14] phase 3), and a different allele in a heterozygous or homozygous state in the genome of the Altai Neanderthal[15] and that segregated in any of the UK Biobank individuals (we refer to these variants as archaic-like SNPs [aSNPs]). We then expanded this by requiring that the identified aSNPs overlap confidently inferred tracts of Neanderthal introgression in modern humans[4] that have a Neanderthal posterior probability greater than 0.9 and a length of at least 0.02 cM. In the construction of this introgression map, a number of criteria were used to ensure that the identified haplotypes were highly likely to be of introgressed origin: (1) alleles were required to be shared between non-Africans and Neanderthals but not be present in sub-Saharan Africans, (2) haplotype lengths had to be consistent with admixture ~50,000 years ago, and (3) haplotypes had to have a lower divergence to a Neanderthal reference genome than to African genomes.

We then collapsed sets of SNPs that were in high linkage disequilibrium (LD) into one representative tag SNP. To do so, we used PLINK[16] (parameters: –ld-window-r2 0.8 –ld-window 99999) and computed LD between all SNPs among the 152,729 individuals by combining sets of SNPs with $r^2 > 0.8$ into clusters. For clusters with at least one aSNP, we selected a random aSNP as the tag SNP. In clusters without aSNPs, we chose a random tag SNP. Non-archaic SNPs and aSNPs with no other SNPs in high LD were defined to be their own tag SNP. We identified a total of 534,341 tag SNPs, of which 6,671 were of putative archaic origin and 527,670 were of non-archaic origin.

To ensure a robust correlation between genotypes and phenotypes, we required each tag SNP to have a reasonable representation of both alleles. We therefore kept all tag SNPs where at least 100 individuals were heterozygous and at least 20 were homozygous for the minor allele, resulting in 6,210 archaic-like tag SNPs and 439,749 non-archaic tag SNPs.

Phenotype Data

Baseline phenotype data were available for different subsets of individuals (Table S1). Of these phenotypes, we used the 136 (including diet, cognitive functions, physical measurements, and self-reported medical conditions) for which data were available for at least 80,000 individuals (Table S1). We excluded phenotypes with complex measurements (e.g., electrocardiography). Phenotypes were represented either in categorical form (72 phenotypes) or as continuous variables (64 phenotypes) (Table S1).

Correlation of Genotype and Phenotype Data

Linear or logistic regression is typically used in association testing to account for potentially confounding covariates such as sex, age, and ancestry; however, applying this standard approach to the UK Biobank is challenging because some of the phenotypes are represented in categorical form for two or more categories, whereas other phenotypes are continuous. Linear regression or generalized linear models are widely used for continuous variables and require knowledge of the distribution of data to be modeled. This distribution is likely to differ between phenotypes, and its assessment is not always trivial. Logistic regression is typically applied to binary phenotypes, such as disease phenotypes. However, many of the categorical phenotypes in the UK Biobank have more than two categories and therefore cannot be transformed into binary data. Another option is to use a multinomial logistic regression, which would require testing each of the categories independently and would vastly increase the complexity of the analysis. We therefore used the chi-square test (for categorical data) and Spearman's correlation (for continuous data) because these statistics make fewer assumptions and are directly applicable to the two classes of phenotypes (categorical and continuous) in the UK Biobank. We excluded categorical data categories for which fewer than 1,000 individuals were available. However, neither test accounts for covariates such as ancestry, age, and sex. There is a strong correlation between ancestry and the presence of Neanderthal alleles. We therefore carefully selected individuals with very little variation in ancestry. There is no *a priori* reason to assume any correlation between Neanderthal ancestry and factors such as age and sex (and no previous study has shown such a correlation). We explicitly tested the impact of these factors on our results by (1) comparing results of linear models with and without covariates and (2) showing that these results were consistent with the results we obtained with a chi-square test (Table S2). To do so, we selected all 21 binary phenotypes and computed an association with all aSNPs by using (1) a chi-square test, (2) a logistic regression without any other covariates, (3) a logistic regression with age and sex as covariates, and (4) a logistic regression with age and sex as covariates and all interactions between age, sex, and genotype.

We found that the correlation between association p values with archaic alleles was between rho = 0.99999 and rho = 1 (Spearman's correlation) for the comparisons of (2) and (3) and of (2) and (4), suggesting that including age and sex has only a marginal impact on the estimation of the association p value.

To estimate the similarity between the results of a logistic regression without covariates (2) and those of a chi-square test (1), we also correlated association p values between the binary phenotypes and archaic alleles for (1) and (2) and found that they ranged between rho = 0.65 and rho = 0.67 (Spearman's correlation; all $p << 1.0 \times 10^{-16}$), suggesting that both tests show highly similar results.

Additionally, we correlated genotypes for all aSNPs used in this study with age and sex and found that there was no significant correlation between these two factors and the aSNP genotypes (false-discovery rate [FDR] < 0.05, min $FDR_{sex} = 0.33$, min $FDR_{age} = 0.28$).

These results suggest that age and sex have very little impact on our calculation of the phenotype association for binary phenotypes, and we infer that non-binary phenotypes are also not likely to be affected by these factors. Applying more sophisticated methods to the analysis of specific phenotypes could increase power to detect additional associations.

For both tests, we considered associations that reached $p < 1.0 \times 10^{-8}$ as significant. This addresses the multiple-testing problem encountered when the associations between 136 phenotypes and approximately 6,000 aSNPs are evaluated (family-wise error rate $= 1.0 \times 10^{-8} \times 6{,}000 \times 136 = 0.01$).

Phenotypic Impact of Archaic and Non-archaic Alleles

For all tag aSNPs, we computed an association p value between genotype and phenotype for each phenotype. We then clustered

tag aSNPs into archaic allele-frequency bins of size 1% and selected frequency-matched non-archaic tag SNPs by matching the number of non-archaic alleles from each frequency bin to the number of archaic alleles. For each phenotype, we created 1,000 random frequency-matched non-archaic sets and computed for each tag SNP an association p value for the phenotype.

To determine whether the archaic p value distributions were shifted to lower or higher significant p values than the non-archaic distributions, we determined the distances between the sets of archaic and non-archaic distributions. More specifically, for each phenotype, we computed empirical p values for the component aSNPs with associations $p < 1.0 \times 10^{-4}$ and compared their cumulative density distribution with the 1,000 non-archaic cumulative density p value distributions (Table S3). We selected the aSNP at which the distance between the archaic distribution and the non-archaic distribution was largest. We corrected all p values for each phenotype for multiple testing by using the Benjamini-Hochberg approach.

Candidate-Gene Analysis and Molecular Mechanism

Given that archaic alleles are typically present on longer haplotypes that we cannot determine directly from the UK Biobank array data, we used the 1000 Genomes[14] (phase 3) individuals to identify aSNPs that were not directly genotyped in the UK Biobank. We computed LD between these by using PLINK (see Annotating Non-archaic and Archaic-like SNPs) and combined sets of aSNPs with $r^2 > 0.8$ between all pairs into a haplotype. We defined the borders of the inferred archaic-like haplotype to be the most distant two aSNPs (Table 1).

We then assigned all 13 candidate tag aSNPs with an association p value $< 1.0 \times 10^{-8}$ (Table 1) to archaic haplotypes inferred from 1000 Genomes.

To determine the targets of these significantly associated aSNPs, we identified overlapping protein-coding genes (Ensembl version GRCh37) or assigned the haplotype to the nearest gene if there was no direct overlap. For each archaic-like haplotype, we identified protein sequence and regulatory variants among the aSNPs in each haplotype and computed the predicted effect of the amino acid changes by using the VEP.[17] Two of the haplotypes with significantly associated aSNPs carried an archaic missense allele (Table 1). To determine whether significantly associated aSNPs might modify gene regulation, we used a previously published set of associations between archaic haplotypes and differential expression in 48 human tissues from the Genotype-Tissue Expression (GTEx) dataset.[18] Of the haplotypes with significantly associated aSNPs, eight were also associated with the expression change of a nearby gene (within 50 kb) in at least one tissue (Table 1).

Testing whether Inferred Archaic Haplotypes Exceed the Length Expected by Incomplete Lineage Sorting

We tested whether the lengths of archaic haplotypes exceeded the length of segments resulting from incomplete lineage sorting (ILS) by using a conservative age of the Altai Neanderthal according to a mutation rate of 1.0×10^{-9} per base pair per year and applying the approach presented by Huerta-Sánchez et al.[19] and the average recombination rates[20] at the inferred haplotype. We corrected the p values obtained from that approach for multiple testing by using the Benjamini-Hochberg method and added them to Table 1.

Haplotype Trees for Candidate Loci

For each of the 13 inferred archaic haplotypes with significant phenotype associations, we extracted the genomic sequences of all 1000 Genomes phase 3 individuals, as well as the genome sequences of the Altai Neanderthal, Denisovan, and chimpanzee (pantro4) (Table 1). We removed non-variable sites and sites where either of the archaic individuals was polymorphic. We then clustered the haplotypes of the combined set of modern and ancient humans together with the chimpanzee into core haplotypes by combining haplotypes that differed by fewer than ~1/1,000 bases. Rooted neighbor-joining trees based on the consensus sequences of the resulting core haplotypes and with chimpanzee as an outgroup were computed and are displayed in Figure S1.

Results

We analyzed 136 baseline phenotypes in 112,338 individuals of British ancestry from the UK Biobank pilot study. A total of 822,111 SNPs directly genotyped in this cohort were classified as either "archaic" or "non-archaic" on the basis of their inclusion in a previously published map of Neanderthal ancestry[4] and their similarity to the Altai Neanderthal genome[15] (Material and Methods). We note that LD between Neanderthal introgressed alleles tends to be higher than LD between non-introgressed alleles because of the timing of Neanderthal introgression. To ensure that the phenotype associations with archaic and non-archaic haplotypes were unbiased, we selected a random tag SNP for each set of SNPs in high LD ($r^2 > 0.8$) and labeled these as "archaic" if the LD set contained at least one ancient SNP and as "non-archaic" otherwise. To ensure sufficient power to detect the phenotypic contribution of each allele, we filtered all tag SNPs for a minimum minor allele frequency (Material and Methods), resulting in a final set of 6,210 archaic tag SNPs and 439,749 non-archaic tag SNPs. We then retained only variants on archaic haplotypes that exceeded the length expected by ILS (Material and Methods).

Phenotypes in the UK Biobank are represented either as categorical (72 phenotypes) or continuous (64 phenotypes) data (Table S1). Linear or logistic regression is typically used in association testing to account for potentially confounding covariates such as sex, age, and ancestry. To avoid testing each of the categories independently, which vastly increases the complexity of the analysis, we applied two different tests: for continuous data, we applied Spearman's correlation to test for an association between each tag SNP and the phenotypic measurement, whereas for categorical data, we used a chi-square test to test for associations between tag SNPs and phenotypes (Material and Methods) and considered only associations that reached $p < 1.0 \times 10^{-8}$ as significant. By comparing our results to those of linear models for subsets of the data, we found that covariates such as age and sex had very little impact on our calculations of phenotype association (Material and Methods and Table S2).

For 11 phenotypes, a total of 15 associations reached genome-wide significance ($p < 1.0 \times 10^{-8}$; Tables 1 and S4). Among these 15 associations were Neanderthal alleles that increase both sitting height and height attained at age

Table 1. Archaic Alleles with Genome-wide-Significant Phenotype Associations

Phenotype	Meta-phenotype	Tag aSNP	Association p Value	Neanderthal Allele Frequency	Data Type	Archaic Haplotype (hg19)	Overlapping Gene(s)	Missense Mutations	Associated eQTLs	FDR ILS Test
Hair color (natural before graying)	sun exposure	chr16: 89,947,203 (rs62052168)	3.7×10^{-202}	0.097	categorical	chr16: 89,813,988–90,008,296	*SPIRE2*, *TCF25*, *MC1R*, *TUBB3*, *FANCA*	–	*FANCA*: muscle (skeletal), lung, pancreas, esophagus (muscularis), adipose (subcutaneous), nerve (tibial), artery (tibial), whole blood	1.84×10^{-9}
									SPIRE2: muscle (skeletal), heart (atrial appendage), adipose (visceral; omentum); skin (not sun exposed; suprapubic), minor salivary gland, esophagus (muscularis), esophagus (mucosa), esophagus (gastresophageal junction), testis, skin (sun exposed; lower leg), adipose (subcutaneous), nerve (tibial), artery (tibial), heart (left ventricle), cells (transformed fibroblasts), artery (aorta), pituitary	
									TCF25: uterus, brain (putamen; basal ganglia)	
									TUBB3: vagina, esophagus (mucosa)	
									MC1R: breast (mammary tissue)	
									DBNDD1: breast (mammary tissue), skin (not sun exposed; suprapubic), skin (sun exposed; lower leg), whole blood	
									GAS8-AS1 (MIM: 605179): testis	
									DEF8: skin (sun exposed; lower leg)	
									GAS8 (MIM: 605178): brain (spinal cord; cervical c-1)	
Skin color	sun exposure	chr6: 45,553,288 (rs115127056)	4.21×10^{-30}	0.075	categorical	chr6: 45,533,261–45,680,205	*RUNX2*	–	*RUNX2*: brain (cerebellum), brain (hippocampus), brain (cerebellar hemisphere)	$<2.2 \times 10^{-22}$
Ease of skin tanning	sun exposure	chr9: 16,804,167 (rs10962612)	1.59×10^{-22}	0.77	categorical	chr9: 16,720,122–16,804,167	*BNC2*	–	*BNC2*: muscle (skeletal)	1.62×10^{-12}
Hair color (natural before graying)	sun exposure	chr14: 92,793,206 (rs77004437)	4.56×10^{-21}	0.089	categorical	chr14: 92,767,097–92,801,297	*SLC24A4*	–	*SLC24A4*: muscle (skeletal)	0.008
Skin color	sun exposure	chr9: 16,904,635 (rs62543578)	1.6×10^{-14}	0.19	categorical	chr9: 16,891,561–16,915,874	*BNC2**	–	–	0.001

(*Continued on next page*)

Table 1. ***Continued***

Phenotype	Meta-phenotype	Tag aSNP	Association p Value	Neanderthal Allele Frequency	Data Type	Archaic Haplotype (hg19)	Overlapping Gene(s)	Missense Mutations	Associated eQTLs	FDR ILS Test
Comparative height size at age 10 years	early life factors	chr19: 31,033,240 (rs56199929)	3.97×10^{-14}	0.16	categorical	chr19: 30,982,165–31,041,053	*ZNF536*	–	–	1.79×10^{-6}
Pulse rate (automated reading)	blood pressure	chr6: 121,947,984 (rs55913590)	6.48×10^{-14}	0.029	continuous	chr6: 121,910,814–122,062,861	*GJA1** (MIM: 121014)	–	–	3.8×10^{-4}
Morning or evening person (chronotype)	sleep	chr2: 239,316,043 (rs75804782)	3.57×10^{-10}	0.12	categorical	chr2: 239,316,043–239,470,654	*ASB1*	*ASB1* (chr2: 239,344,412)	*TRAF3IP1*: testis, liver	$<2.2 \times 10^{-22}$
Skin color	sun exposure	chr11: 89,996,325 (rs74918882)	5.54×10^{-10}	0.041	categorical	chr11: 89,996,325–90,041,511	*CHORDC1**	–	–	0.03
Impedance of leg (left)	impedance measures	chr15: 84,716,986 (rs12902672)	1.46×10^{-9}	0.27	continuous	chr15: 84,703,470–85,114,447	*ADAMTSL3* (MIM: 609199), *GOLGA6L4*	*ADAMTSL3* (chr15: 84,706,461)	*NMB* (MIM: 162340): muscle (skeletal), minor salivary gland, adrenal gland, pancreas, esophagus (muscularis), esophagus (mucosa), stomach, small intestine (terminal ileum), colon (transverse), testis, skin (sun exposed; lower leg), artery (tibial), cells (transformed fibroblasts), spleen, liver	1.17×10^{-5}
									WDR73 (MIM: 616144): heart (atrial appendage), brain (cortex), thyroid, esophagus (muscularis), nerve (tibial), ovary, brain (anterior cingulate cortex; BA24)	
									SLC28A1 (MIM: 606207): breast (mammary tissue)	
									ZNF592 (MIM: 613624): lung, pancreas, liver	
									GOLGA6L4: small intestine (terminal ileum)	
									SEC11A: brain (anterior cingulate cortex; BA24)	
									ALPK3 (MIM: 617608): brain (cerebellar hemisphere)	
									ADAMTSL3: brain (amygdala)	
Incidence of childhood sunburn	sun exposure	chr9: 16,804,167 (rs10962612)	1.49×10^{-9}	0.77	continuous	chr9: 16,720,122–16,804,167	*BNC2*	–	*BNC2*: muscle (skeletal)	1.62×10^{-12}

(*Continued on next page*)

Table 1. ***Continued***

Phenotype	Meta-phenotype	Tag aSNP	Association p Value	Neanderthal Allele Frequency	Data Type	Archaic Haplotype (hg19)	Overlapping Gene(s)	Missense Mutations	Associated eQTLs	FDR ILS Test
Sitting height	body-size measures	chr10: 70,019,371 (rs12571093)	1.52×10^{-9}	0.16	continuous	chr10: 70,009,572–70,059,496	*PBLD* (MIM: 612189)	–	*PBLD*: muscle (skeletal), brain (cortex), brain (caudate; basal ganglia), brain (putamen; basal ganglia)	0.002
									ATOH7 (MIM: 609875): artery (coronary), breast (mammary tissue), skin (not sun exposed; suprapubic), minor salivary gland, adrenal gland, pancreas, esophagus (gastresophageal junction), colon (transverse), adipose (subcutaneous), artery (tibial), brain (cerebellum), artery (aorta), spleen	
									MYPN (MIM: 608517): brain (putamen; basal ganglia)	
Hair color (natural before graying)	sun exposure	chr6: 503,851 (rs71550011)	2.91×10^{-9}	0.07	categorical	chr6: 503,851–544,833	*EXOC2*	–	*EXOC2*: cells (transformed fibroblasts)	0.004
Daytime dozing or sleeping (narcolepsy)	sleep	chr10: 94,711,457 (rs112294410)	4.09×10^{-9}	0.017	categorical	chr10: 94,574,048–94,756,023	*EXOC6*	–	–	$<2.2 \times 10^{-22}$
Impedance of leg (right)	impedance measures	chr15: 84,716,986 (rs12902672)	5.54×10^{-9}	0.27	continuous	chr15: 84,703,470–85,114,447	*ADAMTSL3*, *GOLGA6L4*	*ADAMTSL3* (chr15: 84,706,461)	*NMB*: muscle (skeletal), minor salivary gland, adrenal gland, pancreas, esophagus (muscularis), esophagus (mucosa), stomach, small intestine (terminal ileum), colon (transverse), testis, skin (sun exposed; lower leg), artery (tibial), cells (transformed fibroblasts), spleen, liver	1.17×10^{-5}
									WDR73: heart (atrial appendage), brain (cortex), thyroid, esophagus (muscularis), nerve (tibial), ovary, brain (anterior cingulate cortex; BA24)	
									SLC28A1: breast (mammary tissue)	
									ZNF592: lung, pancreas, liver	
									GOLGA6L4: small intestine (terminal ileum)	
									SEC11A: brain (anterior cingulate cortex; BA24)	
									ALPK3: brain (cerebellar hemisphere)	
									ADAMTSL3: brain (amygdala)	

This table shows archaic alleles with genome-wide-significant associations (column 4, $p < 1.0 \times 10^{-8}$) and their corresponding phenotype (column 1) and meta-phenotype (column 2). Only archaic alleles on confidently inferred archaic introgressed haplotypes are included. The archaic allele frequency in the UK Biobank cohort is given in column 5. Gene identifiers for overlapping or nearest genes (marked with an asterisk) are in column 8. Abbreviations are as follows: eQTL, expression quantitative trait locus; FDR, false-discovery rate; and ILS, incomplete lineage sorting.

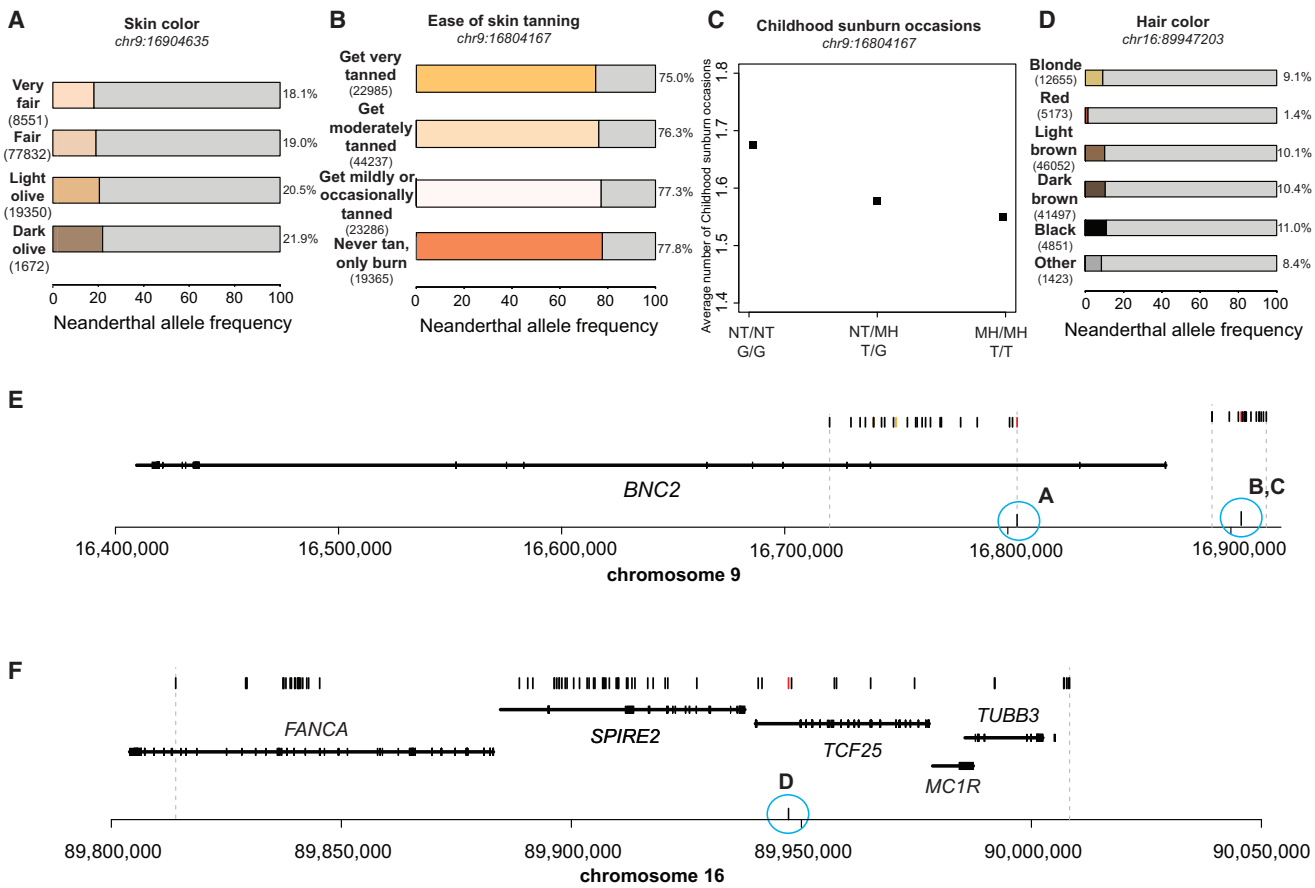


Figure 1. Archaic Haplotypes Associated with Skin and Hair Phenotypes

(A–D) Neanderthal allele frequency in percentage (x axis) and the number of individuals in the UK Biobank cohort for four aSNPs that show strong associations with skin and hair phenotypes (y axis): chr9: 16,904,635 (rs62543578) associated with skin color (A), chr9: 16,804,167 (rs10962612) associated with ease of skin tanning (B) and incidence of childhood sunburn (C) (illustrated are the average numbers of childhood sunburns for individuals with the three genotypes), and chr16: 89,947,203 (rs62052168) associated with hair color (D).

(E and F) The genomic locations of introgressed haplotypes for the aSNPs showing significant associations in (A)–(D). Gray vertical lines denote the extent of the inferred archaic haplotypes on chromosomes 9 (E) and 16 (F). At the top, we show all aSNPs that are within the inferred archaic haplotypes and are present in any 1000 Genomes individual. The associated tag SNPs directly genotyped by the UK Biobank are marked in red, and other aSNPs within the archaic haplotypes and genotyped in the UK Biobank are marked in orange. The associated tag aSNPs represented in (A)–(D) are marked on the x axis.

10 years, alleles that reduce measures of leg impedance (suggesting reduced body fat composition), and alleles that increase resting pulse rate (Table 1). Strikingly, more than half of the significantly associated alleles that we identified are related to skin and hair traits, consistent with previous evidence that genes associated with skin and hair biology are over-represented in introgressed archaic regions.[4,9,11] It was previously only possible to speculate about the precise effect of the introgressed alleles on skin and hair phenotypes on the basis of the genes that were in or near the introgressed haplotypes. We can now directly determine the effect of Neanderthal alleles on these traits in modern humans by correlating Neanderthal ancestry with phenotypes of individuals in the UK Biobank cohort.

The strongest association we found in this study was an archaic allele under-represented among red-haired individuals. This archaic allele is on an introgressed haplotype composed of 71 aSNPs and encompassing five genes: *FANCA* (MIM: 607139), *SPIRE2* (MIM: 609217), *TCF25* (MIM: 612326), *MC1R* (MIM: 155555), and *TUBB3* (MIM: 602661) (rs62052168, $p = 3.7 \times 10^{-202}$; Figure 1 and Table 1). *MC1R* is a key genetic determinant of pigmentation and hair color and is therefore a good candidate for this association. More than 20 variants in *MC1R* have been shown to alter hair color in humans.[21–28] None of the variants resulting in red hair in modern humans are present in either of the two high-coverage Neanderthal genomes that have been sequenced (Table S5). Therefore, Neanderthals appear not to carry any of the variants associated with red hair in modern humans. Further, a Neanderthal-specific variant (p.Arg307Gly) postulated to reduce the activity of MC1R and result in red hair was identified by PCR amplification of *MC1R* in two Neanderthals.[29] However, this putative Neanderthal-specific variant is also not present in the Neanderthals

genomes that have been sequenced to date, suggesting that if this variant was present in Neanderthals, it was rare. Using the high-coverage Neanderthal genomes, we identified only one additional Neanderthal-specific MC1R amino acid change for which the effect on hair color is unknown. However, it is polymorphic among Neanderthals, indicating that any phenotype that it confers was variable in Neanderthals (Table S5). Finally, because the introgressed haplotype we identified in this cohort is underrepresented among red-haired individuals, we conclude that if variants contributing to red hair were present in Neanderthals, they were probably not at high frequency.

We also identified strongly associated archaic alleles on two unlinked introgressed haplotypes near *BNC2* (MIM: 608669), a gene that has been previously associated with skin pigmentation in Europeans.[30] The first archaic haplotype (chr9: 16,720,122–16,804,167) is tagged by an archaic allele (rs10962612) that has a frequency of more than 66% in European populations (Table S6 and Figure 1) and is associated with increased incidence of childhood sunburn ($p = 1.5 \times 10^{-9}$) and poor tanning ($p = 1.6 \times 10^{-22}$) in the UK Biobank cohort (Table 1). A Neanderthal haplotype in this region was previously identified by Vernot and Akey,[11] and the association with sun sensitivity is consistent with the previous finding that Neanderthal alleles on this haplotype result in an increased risk of keratosis.[12] All of the Neanderthal-like SNPs overlapping *BNC2* on this haplotype have significant scores in a test for recent positive selection in Europeans[31] (singleton density score > 3), perhaps indicating their importance in recent local adaptation.

Interestingly, a second, less-frequent (19%) archaic haplotype near *BNC2* (chr9: 16,891,561–16,915,874; rs62543578; Table S6) shows strong associations with darker skin pigmentation in individuals with British ancestry in the UK Biobank cohort ($p = 1.6 \times 10^{-14}$; Figure 1 and Table 1). These results suggest that multiple alleles in and near *BNC2*, some of which are contributed by Neanderthals, have different effects on pigmentation in modern humans. Our analysis identified six additional associations ($p < 1.0 \times 10^{-8}$) contributing to variation in skin and hair biology at other introgressed loci (Table 1). Individuals with blonde hair show a higher frequency of the Neanderthal haplotype at chr6: 503,851–544,833 (overlapping *EXOC2* [MIM: 615329]), whereas individuals with darker hair color show higher Neanderthal ancestry at chr14: 92,767,097–92,801,297 (overlapping *SLC24A4* [MIM: 609840]). Two further archaic haplotypes on chromosomes 6 (chr6: 45,533,261–45,680,205, overlapping *RUNX2* [MIM: 600211]) and 11 (chr11: 89,996,325–90,041,511; nearest gene: *CHORDC1* [MIM: 604353]) are both significantly associated with lighter skin color (Table 1). The apparent variation in the phenotypic effects of Neanderthal alleles in this cohort demonstrates that it is difficult to confidently predict Neanderthal skin and hair color.

Additionally, it is not clear that phenotypic inference from single variants for which a function is known on the modern human genetic background provides sufficient evidence for extrapolating effects in Neanderthals, especially given the challenges with predicting complex phenotypes in present-day humans on the basis of genomic data.[32]

In addition to the introgressed haplotypes contributing to skin and hair traits, we also found two archaic haplotypes that contribute significantly to differences in sleep patterns (Table 1). One of the introgressed SNPs modifies the coding sequence of *ASB1* (MIM: 605758; rs3191996, p.Ser37Lys; Material and Methods). Archaic alleles near *ASB1* (tag aSNP: rs75804782; Figure 2 and Table 1) and *EXOC6* (MIM: 609672; tag aSNP rs71550011; Table 1) are associated with a preference for being an "evening person" and an increased tendency for daytime napping and narcolepsy, respectively. Humans show wide variation in diurnal preferences and can be divided into "chronotypes," which have been shown to have a genetic component.[33] Two previous studies of chronotypes identified strongly associated SNPs in the *ASB1* region.[34,35] Of the 540 SNPs with significant genome-wide associations in Hu et al.[34] ($p < 1.0 \times 10^{-8}$), ten overlapped the region identified near *ASB1*, and four of these were labeled as introgressed archaic variants. Lane et al. identified two *ASB1*-adjacent SNPs that showed significant associations with chronotype.[35] Neither of these are of archaic origin, but they are in high LD with aSNPs on the associated haplotype (maximum $r^2 = 0.73$, based on Europeans in 1000 Genomes phase 3), suggesting that these are not independent signals. Given the association scores calculated by Hu et al.,[34] the association is stronger for the set of aSNPs (p values ranging from 3.4×10^{-6} to 2.6×10^{-9}; rs75804782 has the second-most-significant association at $p = 4.4 \times 10^{-9}$) than for the non-archaic SNPs reported by Lane et al.[35] (rs3769118, $p = 1.9 \times 10^{-6}$; rs11895698, $p = 3.2 \times 10^{-6}$), suggesting that the association is likely to be driven by the introgressed archaic haplotype. Because the natural length of day-night cycles differs according to latitude and influences circadian rhythms, we tested for a correlation between the Neanderthal allele frequency at *ASB1* and latitude in worldwide non-African populations.[36] We found a significant correlation between the frequency of the Neanderthal allele near *ASB1* (rs75804782) and latitude (Spearman's rho = 0.21, p = 0.03). The fact that populations further from the equator have higher frequencies of the Neanderthal allele at *ASB1* than populations nearer the equator (Figure 2B) is consistent with the influence of daylight exposure on circadian rhythm,[37] although the functional link between these genes and chronotype traits is unclear.

Given the large number of associations with skin and hair traits, it is tempting to speculate that Neanderthals might have had an outsized contribution to these phenotypes. However, the number of significant associations that can be identified for a trait is dependent on how polygenic the traits are and how they are measured. Power to measure the contribution of an allele depends also on

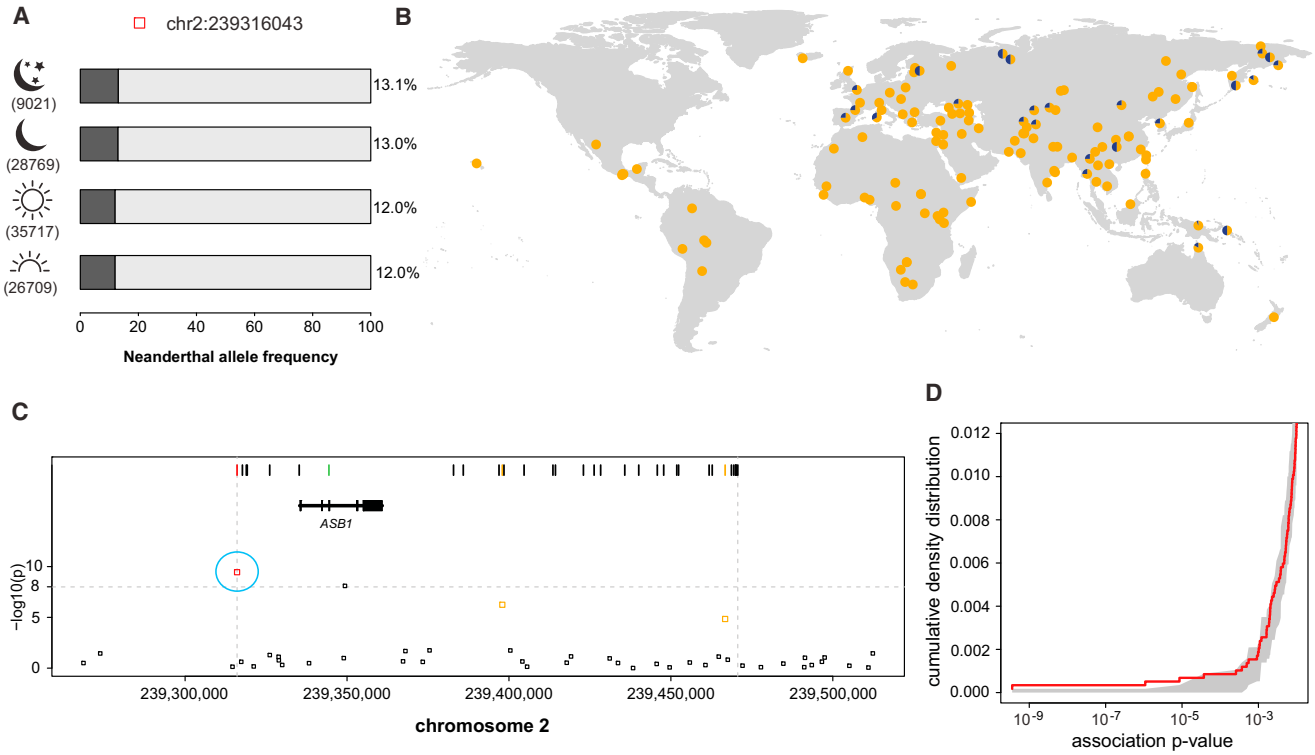


Figure 2. Archaic Haplotype Associated with Chronotype

(A) The Neanderthal allele frequency in percentage (x axis) and the number of individuals in the UK Biobank cohort for the four reported chronotype phenotypes (y axis; from top to bottom: definitely an evening person, more an evening than a morning person, more a morning than an evening person, definitely a morning person) for the archaic tag SNP with the strongest association with chronotype (position chr2: 239,316,043 [rs75804782] near *ASB1*).

(B) Worldwide frequency of the archaic allele (C, blue) and the modern human allele (T, orange) in the Simons Genome Diversity Panel populations.

(C) The association p values (y axis; in the form of $-\log_{10}(p)$) with chronotype for all archaic and non-archaic SNPs (squares) genotyped by the UK Biobank study in the region of the inferred archaic haplotype at chr2: 239,316,043–239,470,654. The tag SNP at chr2: 239,316,043 (rs75804782) is shown in red, other aSNPs are shown in orange, and non-archaic SNPs are shown in black. The genome-wide significance cutoff of $p = 1.0 \times 10^{-8}$ and the extent of the inferred archaic haplotype are illustrated with dashed horizontal and vertical gray lines, respectively. At the top, we show all aSNPs that are within the inferred archaic haplotype and are present in any 1000 Genomes individual. The directly genotyped SNPs from the UK Biobank are illustrated as red (the archaic tag SNP) and orange bars. One archaic allele that leads to a missense mutation in *ASB1* is marked as a green bar.

(D) The cumulative density distribution of p values (zoom in for $p < 0.01$, x axis log scale) for associations between archaic alleles and chronotype (red line) and the 95% confidence interval region for 1,000 cumulative density distributions of associations between non-archaic alleles matched to the Neanderthal allele frequency and chronotype (gray shading).

the minor allele frequency. In the case of archaic alleles, which are generally less frequent (~1%–5%), this is of particular relevance. We therefore tested whether the impact of archaic alleles on particular traits is more or less than that of non-archaic alleles by comparing the contributions of archaic alleles with the contributions of 1,000 similarly sized sets of frequency-matched non-archaic tag SNPs. Phenotypes with an enrichment of low association p values for archaic alleles could indicate a larger-than-expected contribution of introgressed archaic DNA to these phenotypes, whereas an enrichment of low p values for non-archaic alleles suggests a lower contribution from archaic alleles to the phenotype. We note that our frequency matching of archaic and non-archaic alleles does not account for multiple other factors that might differ between these two sets of variants. For example, the longer haplotypes associated with archaic introgression mean that archaic variants might be more likely to occur together. However, it is unclear whether the higher number of archaic alleles on archaic haplotypes would increase or decrease the chance of being significantly associated with phenotypes in modern humans. We believe that further matching of, for example, haplotype length or number of SNPs of a haplotype introduces new potential biases and does not solve this problem. For each phenotype, we selected the lower tail of the p value distributions ($p < 1.0 \times 10^{-4}$) for archaic and non-archaic SNPs and then tested whether the archaic p value distribution was significantly different from 1,000 non-archaic distributions (Material and Methods). For the majority of phenotypes (130/136), we found no difference between the relative contribution of archaic alleles and that of non-archaic alleles, indicating that for most phenotypes measured here, Neanderthal alleles contribute phenotypic

variation proportionally to non-archaic SNPs at similar frequencies (Table S3). We detected six phenotypes where there was a significant difference between the p values distributions for archaic alleles and those for non-archaic alleles (FDR < 0.05). Neanderthal alleles contributed more variation in four behavioral phenotypes influencing sleep, mood, and smoking behaviors, suggesting that Neanderthal alleles contribute more to these traits than expected from their frequency in modern humans. Conversely, for two associations (ease of skin tanning and pork intake), non-archaic alleles showed lower association p values (Table S3), indicating that introgressed Neanderthal alleles contribute less than frequency-matched non-archaic alleles to these traits.

Discussion

Largely on the basis of disease cohorts and signatures of positive selection, a number of immune, skin, metabolic, and behavioral phenotypes have been suggested to be influenced by archaic ancestry. Using the UK Biobank cohort, we have now been able to test the contribution of introgressed Neanderthal alleles to 136 common, largely non-disease phenotypes in present-day Europeans. We found that skin and hair traits are over-represented among the most significant associations with archaic alleles. However, when we compared the contribution of alleles of Neanderthal origin with the contributions of alleles of modern human origin, we found that both archaic and non-archaic variants contribute equally to skin and hair phenotypes, consistent with a neutral contribution from Neanderthals and with the idea that Neanderthals themselves were likely to be variable with respect to these traits. In fact, for most associations, Neanderthal variants do not seem to contribute more than non-archaic variants. However, there are four phenotypes, all behavioral, to which Neanderthal alleles contribute more phenotypic variation than non-archaic alleles: chronotype, loneliness or isolation, frequency of unenthusiasm or disinterest in the last 2 weeks, and smoking status. Of these, the significant association between a Neanderthal variant in *ASB1* and preference for evening activity also shows a correlation between the Neanderthal allele frequency and latitude, suggesting a link to differences in sunlight exposure for this phenotype. Additionally, the phenotype of increased frequencies of unenthusiasm or disinterest in the last 2 weeks was significantly associated with an archaic haplotype (chr5: 29,936,068–29,974,930; nearest gene: *CDH6* [MIM: 603007]), and Neanderthal alleles also contributed more often to this phenotype than non-archaic alleles. A number of the associations we detected, such as dermatological traits, smoking, and mood disorders, overlap associations found in previous studies.[4,11,12] Some of the psychiatric and metabolic phenotypes, such as obesity, identified in Simonti et al.[12] were not replicated in our study. We speculate that this might partially reflect differences in the criteria for cohort selection; individuals in the eMERGE cohort are already undergoing medical treatment, whereas volunteers for the UK Biobank cohort are not.

Multiple phenotypes significantly influenced by Neanderthal introgression have some link to sunlight exposure. Given that Neanderthals had inhabited Eurasia for more than 200,000 years, they were most likely adapted to lower UVB levels and wider variation in sunlight duration than the early modern humans who arrived in Eurasia from Africa around 100,000 years ago.[38] Skin and hair color, circadian rhythms, and mood are all influenced by light exposure. We speculate that their identification in our analysis suggests that sun exposure might have shaped Neanderthal phenotypes and that gene flow into modern humans continues to contribute to variation in these traits today.

Supplemental Data

Supplemental Data include one figure and six tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2017.09.010.

Acknowledgments

This research was conducted with the UK Biobank Resource. We thank Aida Andres, Hernan Burbano, Roger Mundry, Svante Pääbo, Martin Petr, Kay Prüfer, David Reich, Sriram Sankararaman, Joshua Schmidt, and Benjamin Vernot for useful discussions and the multimedia department of the Max Planck Institute for Evolutionary Anthropology for help with figure preparation. Financial support for this study was provided by the Max Planck Society.

Received: June 8, 2017
Accepted: September 5, 2017
Published: October 5, 2017

Web Resources

1000 Genomes, http://browser.1000genomes.org/index.html
dbSNP, https://www.ncbi.nlm.nih.gov/projects/SNP/
Ensembl Genome Browser, http://www.ensembl.org/index.html
GTEx Portal, https://www.gtexportal.org/home/
OMIM, http://www.omim.org/
UK Biobank, http://www.ukbiobank.ac.uk
UK Biobank genotyping and quality controls, https://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_qc.pdf

References

1. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.
2. Harris, K., and Nielsen, R. (2016). The genetic cost of Neanderthal introgression. Genetics *203*, 881–891.
3. Juric, I., Aeschbacher, S., and Coop, G. (2016). The strength of selection against Neanderthal introgression. PLoS Genet. *12*, e1006340.

4. Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., and Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. Nature *507*, 354–357.
5. Dannemann, M., Andrés, A.M., and Kelso, J. (2016). Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human Toll-like receptors. Am. J. Hum. Genet. *98*, 22–33.
6. Gittelman, R.M., Schraiber, J.G., Vernot, B., Mikacenic, C., Wurfel, M.M., and Akey, J.M. (2016). Archaic hominin admixture facilitated adaptation to out-of-Africa environments. Curr. Biol. *26*, 3375–3382.
7. Mendez, F.L., Watkins, J.C., and Hammer, M.F. (2013). Neandertal origin of genetic variation at the cluster of OAS immunity genes. Mol. Biol. Evol. *30*, 798–801.
8. Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic adaptation and Neandertal admixture shaped the immune system of human populations. Cell *167*, 643–656.e17.
9. Racimo, F., Sankararaman, S., Nielsen, R., and Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. *16*, 359–371.
10. Sams, A.J., Dumaine, A., Nédélec, Y., Yotova, V., Alfieri, C., Tanner, J.E., Messer, P.W., and Barreiro, L.B. (2016). Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. Genome Biol. *17*, 246.
11. Vernot, B., and Akey, J.M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. Science *343*, 1017–1021.
12. Simonti, C.N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D.S., Chisholm, R.L., Crosslin, D.R., Hebbring, S.J., Jarvik, G.P., Kullo, I.J., et al. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. Science *351*, 737–741.
13. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779.
14. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
15. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.
16. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
17. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.
18. Dannemann, M., Prüfer, K., and Kelso, J. (2017). Functional implications of Neandertal introgression in modern humans. Genome Biol. *18*, 61.
19. Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., et al. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature *512*, 194–197.
20. Hinds, D.A., McMahon, G., Kiefer, A.K., Do, C.B., Eriksson, N., Evans, D.M., St Pourcain, B., Ring, S.M., Mountain, J.L., Francke, U., et al. (2013). A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat. Genet. *45*, 907–911.
21. Bastiaens, M.T., ter Huurne, J.A., Kielich, C., Gruis, N.A., Westendorp, R.G., Vermeer, B.J., Bavinck, J.N.; and Leiden Skin Cancer Study Team (2001). Melanocortin-1 receptor gene variants determine the risk of nonmelanoma skin cancer independently of fair skin and red hair. Am. J. Hum. Genet. *68*, 884–894.
22. Box, N.F., Wyeth, J.R., O'Gorman, L.E., Martin, N.G., and Sturm, R.A. (1997). Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair. Hum. Mol. Genet. *6*, 1891–1897.
23. Flanagan, N., Healy, E., Ray, A., Philips, S., Todd, C., Jackson, I.J., Birch-Machin, M.A., and Rees, J.L. (2000). Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. Hum. Mol. Genet. *9*, 2531–2537.
24. Harding, R.M., Healy, E., Ray, A.J., Ellis, N.S., Flanagan, N., Todd, C., Dixon, C., Sajantila, A., Jackson, I.J., Birch-Machin, M.A., and Rees, J.L. (2000). Evidence for variable selective pressures at MC1R. Am. J. Hum. Genet. *66*, 1351–1361.
25. Sturm, R.A., Box, N.F., and Ramsay, M. (1998). Human pigmentation genetics: the difference is only skin deep. BioEssays *20*, 712–721.
26. Sturm, R.A., Teasdale, R.D., and Box, N.F. (2001). Human pigmentation genes: identification, structure and consequences of polymorphic variation. Gene *277*, 49–62.
27. Valverde, P., Healy, E., Jackson, I., Rees, J.L., and Thody, A.J. (1995). Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat. Genet. *11*, 328–330.
28. Valverde, P., Healy, E., Sikkink, S., Haldane, F., Thody, A.J., Carothers, A., Jackson, I.J., and Rees, J.L. (1996). The Asp84Glu variant of the melanocortin 1 receptor (MC1R) is associated with melanoma. Hum. Mol. Genet. *5*, 1663–1666.
29. Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., et al. (2007). A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. Science *318*, 1453–1455.
30. Jacobs, L.C., Wollstein, A., Lao, O., Hofman, A., Klaver, C.C., Uitterlinden, A.G., Nijsten, T., Kayser, M., and Liu, F. (2013). Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. Hum. Genet. *132*, 147–158.
31. Field, Y., Boyle, E.A., Telis, N., Gao, Z., Gaulton, K.J., Golan, D., Yengo, L., Rocheleau, G., Froguel, P., McCarthy, M.I., and Pritchard, J.K. (2016). Detection of human adaptation during the past 2000 years. Science *354*, 760–764.
32. Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E., and Visscher, P.M. (2013). Pitfalls of predicting complex traits from SNPs. Nat. Rev. Genet. *14*, 507–515.
33. Roenneberg, T., Wirz-Justice, A., and Merrow, M. (2003). Life between clocks: daily temporal patterns of human chronotypes. J. Biol. Rhythms *18*, 80–90.
34. Hu, Y., Shmygelska, A., Tran, D., Eriksson, N., Tung, J.Y., and Hinds, D.A. (2016). GWAS of 89,283 individuals identifies

genetic variants associated with self-reporting of being a morning person. Nat. Commun. *7*, 10448.

35. Lane, J.M., Vlasac, I., Anderson, S.G., Kyle, S.D., Dixon, W.G., Bechtold, D.A., Gill, S., Little, M.A., Luik, A., Loudon, A., et al. (2016). Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. Nat. Commun. *7*, 10889.
36. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.
37. Adan, A., Archer, S.N., Hidalgo, M.P., Di Milia, L., Natale, V., and Randler, C. (2012). Circadian typology: a comprehensive review. Chronobiol. Int. *29*, 1153–1175.
38. Jablonski, N.G., and Chaplin, G. (2010). Colloquium paper: human skin pigmentation as an adaptation to UV radiation. Proc. Natl. Acad. Sci. USA *107* (*Suppl 2*), 8962–8968.

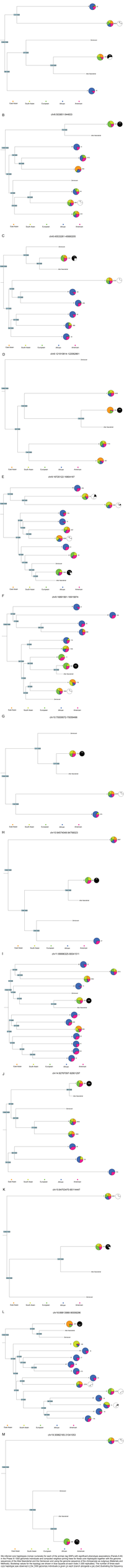
The American Journal of Human Genetics, Volume *101*

Supplemental Data

The Contribution of Neanderthals to Phenotypic Variation in Modern Humans

Michael Dannemann and Janet Kelso

Figure S1.Divergence of the core haplotypes inferred for each of the archaic tag SNPs with significant phenotype associations.



We inferred core haplotypes (roman numerals) for each of the archaic tag SNPs with significant phenotype associations (Panels A-M) at the Phase III 1000 genomes introgressed and ancestral neighbor joining trees for these core haplotypes together with the genome wide maximum parsimony trees for the Denisovan, Alle Neandertal, and the ancestral human haplotype (Methods). Bootstrap values for the topology are shown in blue accurate at each node (1,000 replicates). The number of times each core haplotype was observed in the Phase III 1000 genomes is shown in the pie charts at the tips of the trees, with the relative frequency of the core haplotype in each of the five continental super-populations. Core haplotype clusters in which haplotype carrying the tag allele for the archaic tag SNP are assigned are displayed with an additional pie chart showing the proportion of haplotypes in the cluster that are carriers of the tag SNP. Branch lengths are proportional to the number of haplotypes in the core haplotype cluster. Core haplotypes with fewer than 10 haplotypes assigned are not shown.

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
Alcohol drinker status	Alcohol	categorical	112338	
Alcohol intake frequency.	Alcohol	categorical	112338	
Absence of notch position in the pulse waveform	Arterial stiffness		36580	
Arterial pulse-wave stiffness device ID	Arterial stiffness		2014	
Arterial stiffness device ID	Arterial stiffness		36982	
Position of pulse wave notch	Arterial stiffness		36580	
Position of the pulse wave peak	Arterial stiffness		36580	
Position of the shoulder on the pulse waveform	Arterial stiffness		36580	
Pulse rate	Arterial stiffness		36580	
Pulse wave Arterial Stiffness index	Arterial stiffness		36491	
Pulse wave peak to peak time	Arterial stiffness		36563	
Pulse wave pressure versus time response curve	Arterial stiffness		36580	
Pulse wave reflection index	Arterial stiffness		36580	
Pulse wave velocity (manual entry)	Arterial stiffness		39	
Reason for skipping arterial stiffness	Arterial stiffness		185	
Stiffness method	Arterial stiffness		36982	
Auto-refraction method (right)	Autorefraction		24042	
Reason for skipping refractometry (left)	Autorefraction		304	
Reason for skipping refractometry (right)	Autorefraction		263	
Age at recruitment	Baseline characteristics		112338	
Month of birth	Baseline characteristics		112338	
Sex	Baseline characteristics		112338	
Townsend deprivation index at recruitment	Baseline characteristics		112192	
Year of birth	Baseline characteristics		112338	
Blood pressure device ID	Blood pressure		112242	
Blood pressure manual sphygmomanometer device ID	Blood pressure		2119	
Diastolic blood pressure, automated reading	Blood pressure	continuous	104611	
Diastolic blood pressure, manual reading	Blood pressure			X
Method of measuring blood pressure	Blood pressure			X
Pulse rate (during blood-pressure measurement)	Blood pressure			X
Pulse rate, automated reading	Blood pressure	continuous	104611	
Systolic blood pressure, automated reading	Blood pressure	continuous	104611	
Systolic blood pressure, manual reading	Blood pressure			X
Time since interview start at which blood pressure screen(s) shown	Blood pressure			X
Arm fat mass (left)	Body composition by DXA/Impedance measures	continuous	110379	
Arm fat mass (right)	Body composition by DXA/Impedance measures	continuous	110407	
Leg fat mass (left)	Body composition by DXA/Impedance measures	continuous	110420	
Leg fat mass (right)	Body composition by DXA/Impedance measures	continuous	110426	
Trunk fat mass	Body composition by DXA/Impedance measures	continuous	110362	
Hip circumference	Body size measures	continuous	112173	
Reason for skipping hip measurement	Body size measures		92	
Reason for skipping sitting height	Body size measures		119	
Reason for skipping standing height	Body size measures		118	
Reason for skipping waist	Body size measures		84	
Reason for skipping weight	Body size measures		198	
Sitting height	Body size measures	continuous	112063	
Standing height	Body size measures	continuous	112147	
Waist circumference	Body size measures	continuous	112181	
Weight method	Body size measures		112265	
Weight, manual entry	Body size measures		1630	
Body mass index (BMI)	Body size measures/Impedance measures	continuous	112027	
Weight	Body size measures/Impedance measures	continuous	112065	

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
Ankle spacing width	Bone-densitometry of heel		64971	
Ankle spacing width (left)	Bone-densitometry of heel		35265	
Ankle spacing width (right)	Bone-densitometry of heel		35266	
Foot measured for bone density	Bone-densitometry of heel		74584	
Fractured heel	Bone-densitometry of heel		74584	
Fractured heel (left)	Bone-densitometry of heel		35565	
Fractured heel (right)	Bone-densitometry of heel		35580	
Heel bone mineral density (BMD)	Bone-densitometry of heel		64940	
Heel bone mineral density (BMD) (left)	Bone-densitometry of heel		35247	
Heel bone mineral density (BMD) (right)	Bone-densitometry of heel		35248	
Heel bone mineral density (BMD) T-score, automated	Bone-densitometry of heel		64971	
Heel bone mineral density (BMD) T-score, automated (left)	Bone-densitometry of heel		35265	
Heel bone mineral density (BMD) T-score, automated (right)	Bone-densitometry of heel		35266	
Heel bone mineral density (BMD) T-score, manual entry (left)	Bone-densitometry of heel		299	
Heel bone mineral density (BMD) T-score, manual entry (right)	Bone-densitometry of heel		313	
Heel bone mineral density (BMD), manual entry	Bone-densitometry of heel		9607	
Heel bone mineral density (BMD), manual entry (left)	Bone-densitometry of heel		299	
Heel bone mineral density (BMD), manual entry (right)	Bone-densitometry of heel		313	
Heel bone ultrasound T-score, manual entry	Bone-densitometry of heel		9607	
Heel Broadband ultrasound attenuation (BUA), manual entry	Bone-densitometry of heel		9607	
Heel broadband ultrasound attenuation (BUA), manual entry (left)	Bone-densitometry of heel		299	
Heel broadband ultrasound attenuation (BUA), manual entry (right)	Bone-densitometry of heel		313	
Heel broadband ultrasound attenuation (left)	Bone-densitometry of heel		35259	
Heel broadband ultrasound attenuation (right)	Bone-densitometry of heel		35265	
Heel Broadband ultrasound attenuation, direct entry	Bone-densitometry of heel		64971	
Heel quantitative ultrasound index (QUI), direct entry	Bone-densitometry of heel		64971	
Heel quantitative ultrasound index (QUI), direct entry (left)	Bone-densitometry of heel		35265	
Heel quantitative ultrasound index (QUI), direct entry (right)	Bone-densitometry of heel		35266	
Heel quantitative ultrasound index (QUI), manual entry	Bone-densitometry of heel		9607	
Heel quantitative ultrasound index (QUI), manual entry (left)	Bone-densitometry of heel		299	
Heel quantitative ultrasound index (QUI), manual entry (right)	Bone-densitometry of heel		313	
Heel ultrasound method	Bone-densitometry of heel		75236	
Heel ultrasound method (left)	Bone-densitometry of heel		36946	
Heel ultrasound method (right)	Bone-densitometry of heel		36946	
Shortness of breath walking on level ground	Breathing		36987	
Wheeze or whistling in the chest in last year	Breathing	continuous	112337	
Age at cancer diagnosis	Cancer register		14915	
Behaviour of cancer tumour	Cancer register		14707	
Date of cancer diagnosis	Cancer register		14915	
Histology of cancer tumour	Cancer register		14558	
Reported occurrences of cancer	Cancer register		14915	
Type of cancer: ICD10	Cancer register		12595	
Type of cancer: ICD9	Cancer register		2320	
Ever had bowel cancer screening	Cancer screening		112336	
Ever had prostate specific antigen (PSA) test	Cancer screening		53326	
Most recent bowel cancer screening	Cancer screening		35401	
Time since last prostate specific antigen (PSA) test	Cancer screening		15773	
Chest pain due to walking ceases when standing still	Chest pain		7166	
Chest pain or discomfort	Chest pain	continuous	112336	
Chest pain or discomfort walking normally	Chest pain		17962	
Chest pain or discomfort when walking uphill or hurrying	Chest pain		14633	
Leg pain in calf/calves	Claudication and peripheral artery disease		8286	
Leg pain on walking	Claudication and peripheral artery disease		36987	

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
Leg pain on walking : action taken	Claudication and peripheral artery disease		8286	
Leg pain on walking : effect of standing still	Claudication and peripheral artery disease		8286	
Leg pain when standing still or sitting	Claudication and peripheral artery disease		8286	
Leg pain when walking ever disappears while walking	Claudication and peripheral artery disease		4682	
Leg pain when walking normally	Claudication and peripheral artery disease		8286	
Leg pain when walking uphill or hurrying	Claudication and peripheral artery disease		8286	
Surgery on leg arteries (other than for varicose veins)	Claudication and peripheral artery disease		8286	
Surgery/amputation of toe or leg	Claudication and peripheral artery disease		8286	
Age at death	Death register		2049	
Contributory (secondary) causes of death: ICD10	Death register			X
Date of death	Death register		2049	
Description of cause of death	Death register		1733	
Underlying (primary) cause of death: ICD10	Death register		2048	
Age when last ate meat	Diet		3954	
Bread intake	Diet	continuous	112254	
Bread type	Diet	categorical	108831	
Bread type/intake (pilot)	Diet		83	
Cereal intake	Diet	continuous	112337	
Cereal type	Diet	categorical	93149	
Cheese intake	Diet	categorical	109960	
Coffee intake	Diet	continuous	112338	
Coffee type	Diet	categorical	88164	
Cooked vegetable intake	Diet	continuous	112338	
Fresh fruit intake	Diet	continuous	112338	
Hot drink temperature	Diet	categorical	112338	
Lamb/mutton intake	Diet	categorical	112338	
Major dietary changes in the last 5 years	Diet	categorical	112338	
Milk type used	Diet	categorical	112336	
Never eat eggs, dairy, wheat, sugar	Diet			X
Never eat eggs, dairy, wheat, sugar (pilot)	Diet			X
Non-butter spread type details	Diet		59603	
Non-oily fish intake	Diet	categorical	112338	
Processed meat intake	Diet	categorical	112338	
Salad / raw vegetable intake	Diet	continuous	112338	
Salt added to food	Diet	categorical	112338	
Spread type	Diet	categorical	112254	
Spread type (pilot)	Diet		84	
Tea intake	Diet	continuous	112338	
Variation in diet	Diet	categorical	112254	
Water intake	Diet	continuous	112338	
Dried fruit intake	Diet/Fruit/vegetables yesterday	continuous	112338	
Beef intake	Diet/Meat/fish yesterday	categorical	112338	
Oily fish intake	Diet/Meat/fish yesterday	categorical	112338	
Pork intake	Diet/Meat/fish yesterday	categorical	112338	
Poultry intake	Diet/Meat/fish yesterday	categorical	112338	
Comparative body size at age 10	Early life factors	categorical	112338	
Comparative height size at age 10	Early life factors	categorical	112338	
Handedness (chirality/laterality)	Early life factors	categorical	112338	
Part of a multiple birth	Early life factors	categorical	110757	
Able to walk or cycle unaided for 10 minutes	ECG during exercise		16527	
Chest pain felt during physical activity	ECG during exercise		16527	

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
Chest pain felt outside physical activity	ECG during exercise		16527	
Completion status of test	ECG during exercise		16259	
Doctor restricts physical activity due to heart condition	ECG during exercise		16527	
Duration of fitness test	ECG during exercise		16254	
ECG, heart rate	ECG during exercise			X
ECG, load	ECG during exercise			X
ECG, number of stages in a phase	ECG during exercise			X
ECG, phase duration	ECG during exercise			X
ECG, phase name	ECG during exercise			X
ECG, phase time	ECG during exercise			X
ECG, stage duration	ECG during exercise		0	
ECG, stage name	ECG during exercise			X
ECG, trend phase name	ECG during exercise			X
ECG/bike method for fitness test	ECG during exercise		16475	
Maximum heart rate during fitness test	ECG during exercise		16249	
Maximum workload during fitness test	ECG during exercise		16254	
Number of trend entries	ECG during exercise		16254	
Program category	ECG during exercise		16527	
Reason at-rest ECG performed without bicycle	ECG during exercise		191	
Reason ECG not completed	ECG during exercise		44	
Reason for skipping ECG	ECG during exercise		79	
Target heart rate achieved	ECG during exercise		16254	
Qualifications	Education			X
Length of mobile phone use	Electronic device use		112338	
Ethnic background	Ethnicity		112338	
Both eyes present	Eye surgery/complications		24334	
Current eye infection	Eye surgery/complications		24333	
Ever had cataract surgery	Eye surgery/complications		1714	
Ever had corneal graft surgery	Eye surgery/complications		1714	
Ever had eye surgery	Eye surgery/complications		24334	
Ever had laser treatment for glaucoma or high eye pressure	Eye surgery/complications		1714	
Ever had refractive laser eye surgery	Eye surgery/complications		1714	
Ever had surgery for glaucoma or high eye pressure	Eye surgery/complications		1714	
Age at first live birth	Female-specific factors		40126	
Age of primiparous women at birth of child	Female-specific factors		7907	
Ever had stillbirth, spontaneous miscarriage or termination	Female-specific factors		59010	
Ever taken oral contraceptive pill	Female-specific factors		59010	
Length of menstrual cycle	Female-specific factors		12883	
Number of live births	Female-specific factors		59010	
Number of spontaneous miscarriages	Female-specific factors		18626	
Number of stillbirths	Female-specific factors		18626	
Attempted fluid intelligence (FI) test.	Fluid intelligence test		36539	
FI1 : numeric addition test	Fluid intelligence test/Fluid intelligence test		36093	
FI10 : arithmetic sequence recognition	Fluid intelligence test/Fluid intelligence test		7577	
FI11 : antonym	Fluid intelligence test/Fluid intelligence test		3921	
FI12 : square sequence recognition	Fluid intelligence test/Fluid intelligence test		2732	
FI13 : subset inclusion logic	Fluid intelligence test/Fluid intelligence test		981	
FI2 : identify largest number	Fluid intelligence test/Fluid intelligence test		36029	
FI3 : word interpolation	Fluid intelligence test/Fluid intelligence test		36013	
FI4 : positional arithmetic	Fluid intelligence test/Fluid intelligence test		35982	
FI5 : family relationship calculation	Fluid intelligence test/Fluid intelligence test		35339	

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
FI6 : conditional arithmetic	Fluid intelligence test/Fluid intelligence test		31902	
FI7 : synonym	Fluid intelligence test/Fluid intelligence test		26207	
FI8 : chained arithmetic	Fluid intelligence test/Fluid intelligence test		23796	
FI9 : concept interpolation	Fluid intelligence test/Fluid intelligence test		11078	
Fluid intelligence score	Fluid intelligence test/Fluid intelligence test		36093	
Falls in the last year	General health	categorical	112337	
Long-standing illness, disability or infirmity	General health	categorical	112337	
Overall health rating	General health	categorical	112337	
Weight change compared with 1 year ago	General health	categorical	112337	
Hand grip strength (left)	Hand grip strength	continuous	111875	
Hand grip strength (right)	Hand grip strength	continuous	111876	
Reason for skipping grip strength (left)	Hand grip strength		390	
Reason for skipping grip strength (right)	Hand grip strength		389	
Speech-reception-threshold (SRT) estimate (left)	Hearing test		35290	
Speech-reception-threshold (SRT) estimate (right)	Hearing test		35279	
Arm fat percentage (left)	Impedance measures	continuous	110394	
Arm fat percentage (right)	Impedance measures	continuous	110413	
Arm fat-free mass (left)	Impedance measures	continuous	110380	
Arm fat-free mass (right)	Impedance measures	continuous	110398	
Arm predicted mass (left)	Impedance measures	continuous	110378	
Arm predicted mass (right)	Impedance measures	continuous	110397	
Basal metabolic rate	Impedance measures	continuous	110432	
Body fat percentage	Impedance measures	continuous	110365	
Impedance of arm (left)	Impedance measures	continuous	110426	
Impedance of arm (right)	Impedance measures	continuous	110422	
Impedance of leg (left)	Impedance measures	continuous	110428	
Impedance of leg (right)	Impedance measures	continuous	110429	
Impedance of whole body	Impedance measures	continuous	110426	
Leg fat percentage (left)	Impedance measures	continuous	110421	
Leg fat percentage (right)	Impedance measures	continuous	110428	
Leg fat-free mass (left)	Impedance measures	continuous	110416	
Leg fat-free mass (right)	Impedance measures	continuous	110424	
Leg predicted mass (left)	Impedance measures	continuous	110414	
Leg predicted mass (right)	Impedance measures	continuous	110424	
Trunk fat percentage	Impedance measures	continuous	110368	
Trunk fat-free mass	Impedance measures	continuous	110343	
Trunk predicted mass	Impedance measures	continuous	110327	
Whole body fat mass	Impedance measures	continuous	110243	
Whole body fat-free mass	Impedance measures	continuous	110423	
Whole body water mass	Impedance measures	continuous	110434	
Applanation curve (left)	Intraocular pressure		23181	
Applanation curve (right)	Intraocular pressure		23240	
Corneal hysteresis (left)	Intraocular pressure		23181	
Corneal hysteresis (right)	Intraocular pressure		23240	
Corneal resistance factor (left)	Intraocular pressure		23181	
Corneal resistance factor (right)	Intraocular pressure		23240	
Intra-ocular pressure (IOP) method (left)	Intraocular pressure		24040	
Intra-ocular pressure (IOP) method (right)	Intraocular pressure		24042	
Intra-ocular pressure, corneal-compensated (left)	Intraocular pressure		23181	
Intra-ocular pressure, corneal-compensated (right)	Intraocular pressure		23240	
Intra-ocular pressure, Goldmann-correlated (left)	Intraocular pressure		23181	
Intra-ocular pressure, Goldmann-correlated (right)	Intraocular pressure		23240	

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
Pressure curve (left)	Intraocular pressure		23181	
Pressure curve (right)	Intraocular pressure		23240	
Reason for skipping IOP (left)	Intraocular pressure		496	
Reason for skipping IOP (right)	Intraocular pressure		457	
Hair/balding pattern	Male-specific factors		53283	
Number of children fathered	Male-specific factors		53283	
Relative age of first facial hair	Male-specific factors		53283	
Relative age voice broke	Male-specific factors		53283	
Cancer code, self-reported	Medical conditions			X
Cancer year/age first occurred	Medical conditions			X
Interpolated Age of participant when cancer first diagnosed	Medical conditions	continuous	84969	
Interpolated Age of participant when non-cancer illness first diagnosed	Medical conditions	continuous	84969	
Interpolated Year when cancer first diagnosed	Medical conditions			X
Interpolated Year when non-cancer illness first diagnosed	Medical conditions			X
Method of recording time when cancer first diagnosed	Medical conditions			X
Method of recording time when non-cancer illness first diagnosed	Medical conditions			X
Non-cancer illness code, self-reported	Medical conditions			X
Non-cancer illness year/age first occurred	Medical conditions	continuous	84969	
Number of self-reported cancers	Medical conditions	continuous	112327	
Number of self-reported non-cancer illnesses	Medical conditions	continuous	112327	
Pregnant	Medical conditions		58968	
Bipolar and major depression status	Mental health		26838	
Bipolar disorder status	Mental health		333	
Ever depressed for a whole week	Mental health		36987	
Ever highly irritable/argumentative for 2 days	Mental health		36987	
Ever manic/hyper for 2 days	Mental health		36987	
Ever unenthusiastic/disinterested for a whole week	Mental health		36987	
Family relationship satisfaction	Mental health		36987	
Fed-up feelings	Mental health	categorical	112337	
Financial situation satisfaction	Mental health		36987	
Frequency of depressed mood in last 2 weeks	Mental health	categorical	112337	
Frequency of tenseness / restlessness in last 2 weeks	Mental health	categorical	112337	
Frequency of tiredness / lethargy in last 2 weeks	Mental health	categorical	112337	
Frequency of unenthusiasm / disinterest in last 2 weeks	Mental health	categorical	112337	
Friendships satisfaction	Mental health		36987	
Guilty feelings	Mental health	categorical	112337	
Happiness	Mental health		36987	
Health satisfaction	Mental health		36987	
Illness, injury, bereavement, stress in last 2 years	Mental health	categorical	112253	
Illness, injury, bereavement, stress in last 2 years (pilot)	Mental health			X
Irritability	Mental health	categorical	112337	
Length of longest manic/irritable episode	Mental health		7593	
Loneliness, isolation	Mental health	categorical	112337	
Longest period of depression	Mental health		19426	
Longest period of unenthusiasm / disinterest	Mental health		13101	
Manic/hyper symptoms	Mental health			X
Miserableness	Mental health	categorical	112337	
Mood swings	Mental health	categorical	112337	
Nervous feelings	Mental health	categorical	112337	
Neuroticism score	Mental health	continuous	91620	
Number of depression episodes	Mental health		19426	
Number of unenthusiastic/disinterested episodes	Mental health		13101	
Probable recurrent major depression (moderate)	Mental health		3432	
Probable recurrent major depression (severe)	Mental health		2040	

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
Risk taking	Mental health	categorical	112337	
Seen a psychiatrist for nerves, anxiety, tension or depression	Mental health	categorical	112337	
Seen doctor (GP) for nerves, anxiety, tension or depression	Mental health	categorical	112337	
Sensitivity / hurt feelings	Mental health	categorical	112337	
Severity of manic/irritable episodes	Mental health		7593	
Single episode of probable major depression	Mental health		2250	
Suffer from 'nerves'	Mental health	categorical	112337	
Tense / 'highly strung'	Mental health	categorical	112337	
Work/job satisfaction	Mental health		36987	
Worrier / anxious feelings	Mental health	categorical	112337	
Worry too long after embarrassment	Mental health	categorical	112337	
Mouth/teeth dental problems	Mouth	categorical	112253	
Mouth/teeth dental problems (pilot)	Mouth			X
Back pain for 3+ months	Pain		28709	
Facial pains for 3+ months	Pain		2023	
General pain for 3+ months	Pain		1769	
Headaches for 3+ months	Pain		22091	
Hip pain for 3+ months	Pain		12995	
Knee pain for 3+ months	Pain		24373	
Neck/shoulder pain for 3+ months	Pain		25998	
Pain type(s) experienced in last month	Pain	categorical	112336	
Stomach/abdominal pain for 3+ months	Pain		9556	
Number of incorrect matches in round	Pairs matching test/Pairs matching test			X
Frequency of heavy DIY in last 4 weeks	Physical activity		48818	
Frequency of light DIY in last 4 weeks	Physical activity		57730	
Frequency of other exercises in last 4 weeks	Physical activity		53000	
Frequency of stair climbing in last 4 weeks	Physical activity	categorical	111878	
Frequency of strenuous sports in last 4 weeks	Physical activity		11008	
Frequency of walking for pleasure in last 4 weeks	Physical activity	categorical	79141	
Number of days/week walked 10+ minutes	Physical activity	categorical	112338	
Types of physical activity in last 4 weeks	Physical activity	categorical	111935	
Usual walking pace	Physical activity	categorical	111962	
Duration screen displayed	Prospective memory test		36850	
Final attempt correct	Prospective memory test		36850	
History of attempts	Prospective memory test		36850	
Number of attempts	Prospective memory test		36850	
PM: final answer	Prospective memory test		36850	
PM: initial answer	Prospective memory test		36850	
Prospective memory result	Prospective memory test		36850	
Test completion status	Prospective memory test		36850	
Time screen exited	Prospective memory test		36850	
Time to answer	Prospective memory test		36850	
Time when initial screen shown	Prospective memory test		36850	
Age when attended assessment centre	Reception		112338	
Date of attending assessment centre	Reception		112338	
UK Biobank assessment centre	Reception		112338	
OCT measured (left)	Retinal optical coherence tomography		16533	
OCT measured (right)	Retinal optical coherence tomography		16531	
Reason for skipping OCT (left)	Retinal optical coherence tomography		716	
Reason for skipping OCT (right)	Retinal optical coherence tomography		675	
Age first had sexual intercourse	Sexual factors	continuous	103379	
Answered sexual history questions	Sexual factors		112337	
Ever had same-sex intercourse	Sexual factors	categorical	102415	
Lifetime number of same-sex sexual partners	Sexual factors		3374	

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
Lifetime number of sexual partners	Sexual factors	continuous	102415	
Daytime dozing / sleeping (narcolepsy)	Sleep	categorical	112338	
Getting up in morning	Sleep	categorical	112254	
Morning/evening person (chronotype)	Sleep	categorical	112254	
Nap during day	Sleep	categorical	112338	
Sleep duration	Sleep	continuous	112338	
Sleeplessness / insomnia	Sleep	categorical	112338	
Snoring	Sleep	categorical	112338	
Smoking status	Smoking	categorical	112338	
Episodes containing "Diagnoses - main ICD10 - addendum" data	Spell and Episode Data (diagnoses)		458	
Episodes containing "Diagnoses - main ICD10" data	Spell and Episode Data (diagnoses)		79357	
Episodes containing "Diagnoses - secondary ICD10 addendum" data	Spell and Episode Data (diagnoses)		550	
Episodes containing "Diagnoses - secondary ICD10" data	Spell and Episode Data (diagnoses)		59042	
Episodes containing "External cause - ICD10 addendum" data	Spell and Episode Data (diagnoses)		56	
Episodes containing "External cause - ICD10" data	Spell and Episode Data (diagnoses)		10661	
Acceptability of each blow result (text)	Spirometry			X
Acceptability of each blow result (text) (pilot)	Spirometry			X
Caffeine drink within last hour	Spirometry	categorical	105523	
Contra-indications for spirometry	Spirometry	categorical	112265	
Data points for blow	Spirometry			X
Data points for blow (pilot)	Spirometry			X
Forced expiratory volume in 1-second (FEV1)	Spirometry			X
Forced expiratory volume in 1-second (FEV1) (pilot)	Spirometry			X
Forced expiratory volume in 1-second (FEV1), Best measure	Spirometry	continuous	91907	
Forced expiratory volume in 1-second (FEV1), predicted	Spirometry		56611	
Forced expiratory volume in 1-second (FEV1), predicted percentage	Spirometry		56611	
Forced vital capacity (FVC)	Spirometry			X
Forced vital capacity (FVC) (pilot)	Spirometry			X
Forced vital capacity (FVC), Best measure	Spirometry	continuous	91907	
Number of measurements made	Spirometry		104956	
Ordering of blows	Spirometry			X
Peak expiratory flow (PEF)	Spirometry			X
Peak expiratory flow (PEF) (pilot)	Spirometry			X
Reason for skipping spirometry	Spirometry		339	
Reproduciblity of spirometry measurement using ERS/ATS criteria	Spirometry		91907	
Result ranking	Spirometry			X
Result ranking (pilot)	Spirometry			X
Smoked cigarette or pipe within last hour	Spirometry		12758	
Spirometry method	Spirometry		105444	
Spirometry method (pilot)	Spirometry		79	
Used an inhaler for chest within last hour	Spirometry		105523	
Diagnoses - main ICD10	Summary Information (diagnoses)			X
Diagnoses - secondary ICD10	Summary Information (diagnoses)			X
External causes	Summary Information (diagnoses)			X
History of psychiatric care on admission	Summary Information (psychiatric)			X
Mental categories	Summary Information (psychiatric)			X
Childhood sunburn occasions	Sun exposure	continuous	112254	
Ease of skin tanning	Sun exposure	categorical	112254	
Facial ageing	Sun exposure	categorical	112253	
Hair colour (natural, before greying)	Sun exposure	categorical	112338	
Skin colour	Sun exposure	categorical	112338	
Time spend outdoors in summer	Sun exposure	continuous	112254	
Time spent outdoors in winter	Sun exposure	continuous	112254	
Glasses worn/required (left)	Visual acuity		16640	

Table S1: Table of phenotypes from the UK biobank.

Phenotype	Meta phenotype	Data type	Number of individuals	Complex phenotype
Glasses worn/required (right)	Visual acuity		16656	
logMAR, final (left)	Visual acuity		23991	
logMAR, final (right)	Visual acuity		24002	
Reason for skipping visual acuity (left)	Visual acuity		96	
Reason for skipping visual acuity (right)	Visual acuity		78	
Visual acuity measured (left)	Visual acuity		24290	
Visual acuity measured (right)	Visual acuity		24292	

Table includes all UK Biobank phenotypes to which we had access (column 1, with meta phenotype provided by UK biobank in column 2). The 136 phenotypes included in this study are highlighted in bold. Column 3 provides format in which the 136 phenotype we included are represented and column 4 the number of individuals selected for this study because both genotype and phenotype information were available (Material and Methods). In Column 5 phenotypes with complex measurements (i.e. phenotypes that do not have one single measurement) are marked with an ‘X’.

Table S3: Significantly associated phenotypes that show an over- or under-representation of Neandertal alleles.

phenotype	P-value	FDR	maximum enrichment	OR	direction
Pork intake	<0.001	0.023	2	0.222	under-representation
Ease of skin tanning	<0.001	0.023	5	0.364	under-representation
Smoking status	<0.001	0.023	5	4.629	over-representation
Morning/evening person (chronotype)	<0.001	0.023	3	22.22	over-representation
Loneliness, isolation	<0.001	0.023	2	487.388	over-representation
Frequency of unenthusiasm / disinterest in last 2 weeks	0.002	0.039	3	14.164	over-representation
Pain type(s) experienced in last month	0.004	0.052	4	0.399	under-representation
Number of days/week walked 10+ minutes	0.004	0.052	5	5.39	over-representation
Alcohol intake frequency.	0.004	0.052	4	7.162	over-representation
Water intake	0.008	0.077	3	13.83	over-representation
Body mass index (BMI)	0.01	0.077	13	2.268	over-representation
Handedness (chirality/laterality)	0.01	0.077	5	4.516	over-representation
Salad / raw vegetable intake	0.01	0.077	5	4.616	over-representation
Non-cancer illness year/age first occurred	0.01	0.077	4	5.814	over-representation
Frequency of tiredness / lethargy in last 2 weeks	0.01	0.077	4	6.008	over-representation
Non-oily fish intake	0.012	0.082	1	0.28	under-representation
Coffee intake	0.012	0.082	1	163.265	over-representation
Bread intake	0.014	0.09	2	13.976	over-representation
Tea intake	0.016	0.095	1	122.401	over-representation
Processed meat intake	0.018	0.095	1	0.279	under-representation
Ever had same-sex intercourse	0.018	0.095	2	0.316	under-representation
Pulse rate, automated reading	0.018	0.095	10	2.339	over-representation
Alcohol drinker status	0.02	0.097	2	0.354	under-representation
Coffee type	0.02	0.097	3	6.111	over-representation
Daytime dozing / sleeping (narcolepsy)	0.022	0.102	3	6.412	over-representation
Caffeine drink within last hour	0.024	0.103	7	2.778	over-representation
Miserableness	0.024	0.103	1	81.376	over-representation
Seen a psychiatrist for nerves, anxiety, tension or depression	0.028	0.112	4	4.367	over-representation
Frequency of depressed mood in last 2 weeks	0.028	0.112	1	69.875	over-representation
Whole body fat-free mass	0.032	0.12	1	0.306	under-representation
Basal metabolic rate	0.032	0.12	2	0.376	under-representation
Trunk fat-free mass	0.036	0.13	1	0.292	under-representation
Trunk predicted mass	0.042	0.148	2	0.374	under-representation
Lifetime number of sexual partners	0.046	0.157	1	42.462	over-representation
Hand grip strength (left)	0.048	0.159	2	7.322	over-representation
Major dietary changes in the last 5 years	0.066	0.213	2	6.474	over-representation
Interpolated Year when non-cancer illness first diagnosed	0.074	0.232	1	25.718	over-representation
Standing height	0.076	0.232	1	0.324	under-representation
Childhood sunburn occasions	0.092	0.274	4	0.545	under-representation
Mouth/teeth dental problems	0.096	0.278	3	3.906	over-representation
Cheese intake	0.112	0.315	2	5.652	over-representation
Nap during day	0.114	0.315	3	3.568	over-representation

Table S3: Significantly associated phenotypes that show an over- or under-representation of Neandertal alleles.

phenotype	P-value	FDR	maximum enrichment	OR	direction
Types of physical activity in last 4 weeks	0.12	0.324	1	0.447	under-representation
Comparative height size at age 10	0.128	0.337	5	2.261	over-representation
Frequency of walking for pleasure in last 4 weeks	0.134	0.341	2	4.687	over-representation
Frequency of tenseness / restlessness in last 2 weeks	0.136	0.341	3	3.101	over-representation
Impedance of arm (left)	0.144	0.341	1	0.498	under-representation
Diastolic blood pressure, automated reading	0.146	0.341	3	0.568	under-representation
Arm fat percentage (right)	0.146	0.341	4	2.515	over-representation
Contra-indications for spirometry	0.148	0.341	1	0.478	under-representation
Impedance of leg (right)	0.15	0.341	5	0.595	under-representation
Leg predicted mass (right)	0.16	0.356	2	0.507	under-representation
Number of self-reported cancers	0.166	0.356	3	0.601	under-representation
Trunk fat mass	0.17	0.356	1	0.478	under-representation
Fed-up feelings	0.17	0.356	1	11.109	over-representation
Hair colour (natural, before greying)	0.172	0.356	9	0.706	under-representation
Leg fat-free mass (right)	0.178	0.359	2	0.513	under-representation
Time spend outdoors in summer	0.182	0.359	3	2.712	over-representation
Impedance of arm (right)	0.188	0.359	2	0.536	under-representation
Number of self-reported non-cancer illnesses	0.19	0.359	3	2.662	over-representation
Skin colour	0.192	0.359	4	0.662	under-representation
Forced expiratory volume in 1-second (FEV1), Best measure	0.192	0.359	2	3.833	over-representation
Leg fat mass (right)	0.204	0.375	5	1.986	over-representation
Sleeplessness / insomnia	0.208	0.375	2	3.675	over-representation
Interpolated Age of participant when non-cancer illness first diagnosed	0.214	0.375	3	2.366	over-representation
Frequency of stair climbing in last 4 weeks	0.214	0.375	1	8.652	over-representation
Trunk fat percentage	0.218	0.375	3	2.465	over-representation
Leg fat-free mass (left)	0.22	0.375	2	0.561	under-representation
Leg predicted mass (left)	0.224	0.377	2	0.57	under-representation
Arm fat mass (left)	0.232	0.382	9	1.527	over-representation
Milk type used	0.234	0.382	1	0.581	under-representation
Lamb/mutton intake	0.238	0.383	3	0.595	under-representation
Facial ageing	0.244	0.388	5	1.772	over-representation
Poultry intake	0.26	0.402	4	1.951	over-representation
Overall health rating	0.26	0.402	2	3.076	over-representation
Impedance of whole body	0.266	0.406	1	0.736	under-representation
Fresh fruit intake	0.284	0.428	1	6.608	over-representation
Bread type	0.288	0.428	2	2.952	over-representation
Impedance of leg (left)	0.294	0.432	8	0.724	under-representation
Beef intake	0.298	0.432	3	2.231	over-representation
Whole body fat mass	0.304	0.435	2	0.626	under-representation
Arm fat mass (right)	0.32	0.447	6	1.568	over-representation
Never eat eggs, dairy, wheat, sugar	0.324	0.447	1	0.539	under-representation
Cereal intake	0.324	0.447	2	2.604	over-representation

Table S3: Significantly associated phenotypes that show an over- or under-representation of Neandertal alleles.

phenotype	P-value	FDR	maximum enrichment	OR	direction
Hot drink temperature	0.328	0.448	1	5.228	over-representation
Illness, injury, bereavement, stress in last 2 years	0.348	0.469	2	2.643	over-representation
Hip circumference	0.364	0.485	2	0.703	under-representation
Mood swings	0.402	0.529	2	2.276	over-representation
Weight	0.41	0.529	3	0.714	under-representation
Sitting height	0.422	0.529	2	0.7	under-representation
Systolic blood pressure, automated reading	0.422	0.529	2	0.705	under-representation
Snoring	0.422	0.529	1	4.25	over-representation
Sleep duration	0.424	0.529	2	2.305	over-representation
Spread type	0.442	0.545	2	2.23	over-representation
Usual walking pace	0.448	0.547	1	1.012	under-representation
Neuroticism score	0.472	0.564	1	0.7	under-representation
Hand grip strength (right)	0.472	0.564	2	2.174	over-representation
Comparative body size at age 10	0.486	0.575	3	0.737	under-representation
Part of a multiple birth	0.512	0.6	2	1.973	over-representation
Leg fat mass (left)	0.522	0.605	1	0.756	under-representation
Forced vital capacity (FVC), Best measure	0.53	0.605	5	1.394	over-representation
Age first had sexual intercourse	0.532	0.605	1	3.155	over-representation
Waist circumference	0.538	0.605	1	3.292	over-representation
Cereal type	0.544	0.605	2	1.897	over-representation
Sensitivity / hurt feelings	0.548	0.605	1	3.064	over-representation
Arm fat percentage (left)	0.562	0.615	3	1.571	over-representation
Leg fat percentage (right)	0.568	0.616	3	1.612	over-representation
Cooked vegetable intake	0.6	0.641	2	1.865	over-representation
Wheeze or whistling in the chest in last year	0.602	0.641	2	1.814	over-representation
Dried fruit intake	0.642	0.675	2	0.9	under-representation
Salt added to food	0.646	0.675	1	2.42	over-representation
Body fat percentage	0.68	0.704	2	1.596	over-representation
Falls in the last year	0.686	0.704	1	0.905	under-representation
Risk taking	0.702	0.714	2	0.892	under-representation
Leg fat percentage (left)	0.714	0.72	2	1.625	over-representation
Long-standing illness, disability or infirmity	0.734	0.734	1	1.055	under-representation

For each phenotype (column 1) we show the empirical P value (column 2) for the comparison of the cumulative association P-value distribution (for P<10-4) for archaic alleles and 1,000 cumulative non-archaic allele association P-value distributions. The corresponding false discovery rates (column 3), the rank of the sorted set of archaic SNPs (by P value) for which the archaic distribution has the largest distance to the mean non-archaic distributions (column 4) and the corresponding largest odds ratio between the archaic association P-value distributions (column 5) are provided. Column 6 identifies whether the Neandertal alleles are under- or over-represented compared to non-archaic alleles based on the direction of enrichment.

Table S5: Amino acid variants associated with hair color variation at MC1R.

AA variant	chromosomal position (chr 16)	rsID	Vindija 33.19	Altai	ref	alt	publication(s)
Val60Leu	89985844	*rs1805005*	G	G	G	T	1,2,3,4,6,7
Ala64Ser	89985856		G	G	G	T	5
Arg67Gln	89985866	*rs34090186*	G	G	G	A	1,8
Phe76Tyr	89985893		T	T	T	A	5
Ala81Pro	89985907		G	G	G	C	7
Asp84Glu	89985918	*rs1805006*	C	C	C	A	1,2,3,4,5,6,7
Val92Met	89985940	*rs2228479*	G	G	G	A	1,2,3,4,5,6,7
Thr95Met	89985950		C	C	C	T	4,5,7
Val97Ile	89985955		G	G	G	A	5
Ala103Val	89985974		C	C	C	T	5
Gly104Ser	89985976	*rs2229617*	G	G	G	A	7
Leu106Gln	89985983		T	T	T	A	5
Arg142His	89986091	*rs11547464*	G	G	G	A	1,2,3,4,7
Arg151Cys	89986117	*rs1805007*	C	C	C	T	1,2,3,4,6,7
Ile155Thr	89986130	*rs1110400*	T	T	T	C	1,2,3,4,6,7
Arg160Trp	89986144	*rs1805008*	C	C	C	T	1,2,3,4,6,7
Arg163Gln	89986154	*rs885479*	G	G	G	A	1,2,3,4,6,7,8
Val174Ile	89986186		G	G	G	A	7
Pro230Leu	89986355	*rs368714912*	C	C	C	T	7
His260Pro	89986445		A	A	A	A	7
Val265Ile	89986459	*rs587783376*	G	A	G	A	
Lys278Glu	89986498	*rs201171524*	A	A	A	G	7
Asn279Ser	89986502	*rs376692024*	A	A	A	G	7
Ile287Met	89986527	*rs373957223*	C	C	C	G	1,8
Asp294His	89986546	*rs1805009*	G	G	G	C	1,2,3,4,5,6,7
Arg307Gly	89986586		G	G	G		9

Table includes previously reported non-synonymous variants which have been significantly associated with hair color variation and their genotype in two high-coverage Neandertals (columns 4,5).

(1) Box et al., 1997
(2) Sturm et al., 1998
(3) Box et al., 2001b
(4) Flanagan et al., 2000
(5) Valverde et al., 1995
(6) Harding et al., 2000
(7) Bastiaens et al., 2001
(8) Rana et al., 1999
(9) Lalueza-Fox et al. 2010

Table S6: : Frequency of significantly associated Neandertal alleles in the UK biobank cohort and 1,000 Genomes (phase III) populations

	UK biobank	CEU	TSI	FIN	GBR	IBS	CHB	JPT	CHS	CDX	KHV	GIH	PJL	BEB	STU	ITU	YRI	LWK	GWD	MSL	ESN	ASW	ACB	MXL	PUR	CLM	PEL
chr2:239316043	12.4	13.6	8.3	18.2	13.6	10.3	7.8	2.4	5.1	4.5	5	8	5.7	2.3	3.4	3.4	0	0	0	0	0	0.8	1.6	3.7	7.6	4.3	2.3
chr6:121947984	2.9	3	4.2	3.5	3.3	3.3	10.7	9.1	14.4	17.2	16.8	0.9	1	1.7	0.5	0.5	0	1.5	0	0	0	0	0	0.7	2.4	1.1	1.7
chr6:45553288	7.5	7.1	4.6	8.1	7.6	5.1	0.5	0	0	0.5	0.5	1.4	2.1	0.6	1.5	3.9	0	0	0	0	0	0.8	1	4.5	3.3	4.3	1.2
chr6:503851	7	5.6	1.4	8.1	5.4	6.5	1	0	0	0	0	5.2	3.1	3.5	3.4	1.5	0	0	0	0	0	2.3	1	3.7	3.3	2.1	1.2
chr9:16804167	76.5	80.3	69	74.2	75	66.4	0	0	0	0	0	22.6	26	13.4	12.1	14.1	0	1.5	0	0	0	9.8	8.9	29.1	48.6	46.3	12.2
chr9:16904635	19.3	22.7	17.6	17.7	19	17.8	0	0	0	0	0	4.7	6.2	5.2	2.9	4.9	0	0.5	0	0	0	3.8	3.6	9	12.9	9.6	4.1
chr10:70019371	15.9	12.6	13.4	23.7	15.2	15.9	16.5	12	16.7	15.7	10.4	22.6	22.4	27.9	22.8	18.9	0	0	0.9	0	0	3	6.2	23.9	11	21.8	48.8
chr10:94711457	1.7	2	1.9	1	2.2	2.8	0	0	0	0	0	0.9	0.5	2.9	0.5	1.5	0	0	0	0	0	0.8	0.5	2.2	0	2.7	2.3
chr11:89996325	4.1	3.5	1.9	6.1	4.3	2.8	1.5	1.4	1.4	0.5	0.5	4.7	5.7	3.5	3.4	1.5	0	0	0	0	0	0.8	0.5	0.7	3.3	2.1	2.9
chr14:92793206	8.9	12.1	14.8	5.6	8.7	10.7	3.9	4.3	4.2	5.6	4	3.3	6.2	2.9	3.9	4.9	0	0	0.4	0	0	0.8	2.1	9	14.8	8	5.2
chr15:84716986	26.6	28.8	30.1	27.8	24.5	24.8	6.3	2.4	6.5	5.6	6.9	16	25	19.8	17.5	21.8	0	1.5	0.4	0	0	4.5	3.1	16.4	21.4	17.6	15.1
chr16:89947203	9.7	6.6	6	5.6	11.4	5.1	20.4	10.6	32.9	42.9	39.6	2.8	3.6	2.3	0	0.5	0	0	0	0	0	2.3	2.1	5.2	2.9	1.1	1.2
chr19:31033240	16.2	18.7	13	19.7	15.8	17.8	0	0	0	0	0	4.2	1	1.2	6.8	2.4	0	0	0.4	0	0	3.8	2.1	7.5	12.9	12.8	5.2