

# The Contribution of Neanderthals to Phenotypic Variation in Modern Humans

Michael Dannemann<sup>1</sup> and Janet Kelso<sup>1,\*</sup>

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.<sup>1–4</sup> 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.<sup>4–11</sup>

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.<sup>12</sup> 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 num-

ber 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.<sup>13</sup> 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.<sup>13</sup> 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.

<sup>1</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany

\*Correspondence: [kelso@eva.mpg.de](mailto: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<sup>14</sup> phase 3), and a different allele in a heterozygous or homozygous state in the genome of the Altai Neanderthal<sup>15</sup> 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<sup>4</sup> 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<sup>16</sup> (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  $p \ll 1.0 \times 10^{-16}$ ), 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  $\text{FDR}_{\text{sex}} = 0.33$ , min  $\text{FDR}_{\text{age}} = 0.28$ ).

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<sup>14</sup> (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.<sup>17</sup> 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.<sup>18</sup> 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  $1.0 \times 10^{-9}$  per base pair per year and applying the approach presented by Huerta-Sánchez et al.<sup>19</sup> and the average recombination rates<sup>20</sup> 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  $\sim 1/1,000$  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<sup>4</sup> and their similarity to the Altai Neanderthal genome<sup>15</sup> (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)	$3.7 \times 10^{-202}$	0.097	categorical	chr16: 89,813,988–90,008,296	<i>SPIRE2</i> , <i>TCF25</i> , <i>MC1R</i> , <i>TUBB3</i> , <i>FANCA</i>	–	<p><i>FANCA</i>: muscle (skeletal), lung, pancreas, esophagus (muscularis), adipose (subcutaneous), nerve (tibial), artery (tibial), whole blood</p> <p><i>SPIRE2</i>: muscle (skeletal), heart (atrial appendage), adipose (visceral; omentum); skin (not sun exposed; suprapubic), minor salivary gland, esophagus (muscularis), esophagus (mucosa), esophagus (gastroesophageal junction), testis, skin (sun exposed; lower leg), adipose (subcutaneous), nerve (tibial), artery (tibial), heart (left ventricle), cells (transformed fibroblasts), artery (aorta), pituitary</p> <p><i>TCF25</i>: uterus, brain (putamen; basal ganglia)</p> <p><i>TUBB3</i>: vagina, esophagus (mucosa)</p> <p><i>MC1R</i>: breast (mammary tissue)</p> <p><i>DBNDD1</i>: breast (mammary tissue), skin (not sun exposed; suprapubic), skin (sun exposed; lower leg), whole blood</p> <p><i>GAS8-AS1</i> (MIM: 605179): testis</p> <p><i>DEF8</i>: skin (sun exposed; lower leg)</p> <p><i>GAS8</i> (MIM: 605178): brain (spinal cord; cervical c-1)</p>	$1.84 \times 10^{-9}$
Skin color	sun exposure	chr6: 45,553,288 (rs115127056)	$4.21 \times 10^{-30}$	0.075	categorical	chr6: 45,533,261–45,680,205	<i>RUNX2</i>	–	<i>RUNX2</i> : brain (cerebellum), brain (hippocampus), brain (cerebellar hemisphere)	$<2.2 \times 10^{-22}$
Ease of skin tanning	sun exposure	chr9: 16,804,167 (rs10962612)	$1.59 \times 10^{-22}$	0.77	categorical	chr9: 16,720,122–16,804,167	<i>BNC2</i>	–	<i>BNC2</i> : muscle (skeletal)	$1.62 \times 10^{-12}$
Hair color (natural before graying)	sun exposure	chr14: 92,793,206 (rs77004437)	$4.56 \times 10^{-21}$	0.089	categorical	chr14: 92,767,097–92,801,297	<i>SLC24A4</i>	–	<i>SLC24A4</i> : muscle (skeletal)	0.008
Skin color	sun exposure	chr9: 16,904,635 (rs62543578)	$1.6 \times 10^{-14}$	0.19	categorical	chr9: 16,891,561–16,915,874	<i>BNC2*</i>	–	–	0.001

(Continued on next page)

**Table 1. Continued**

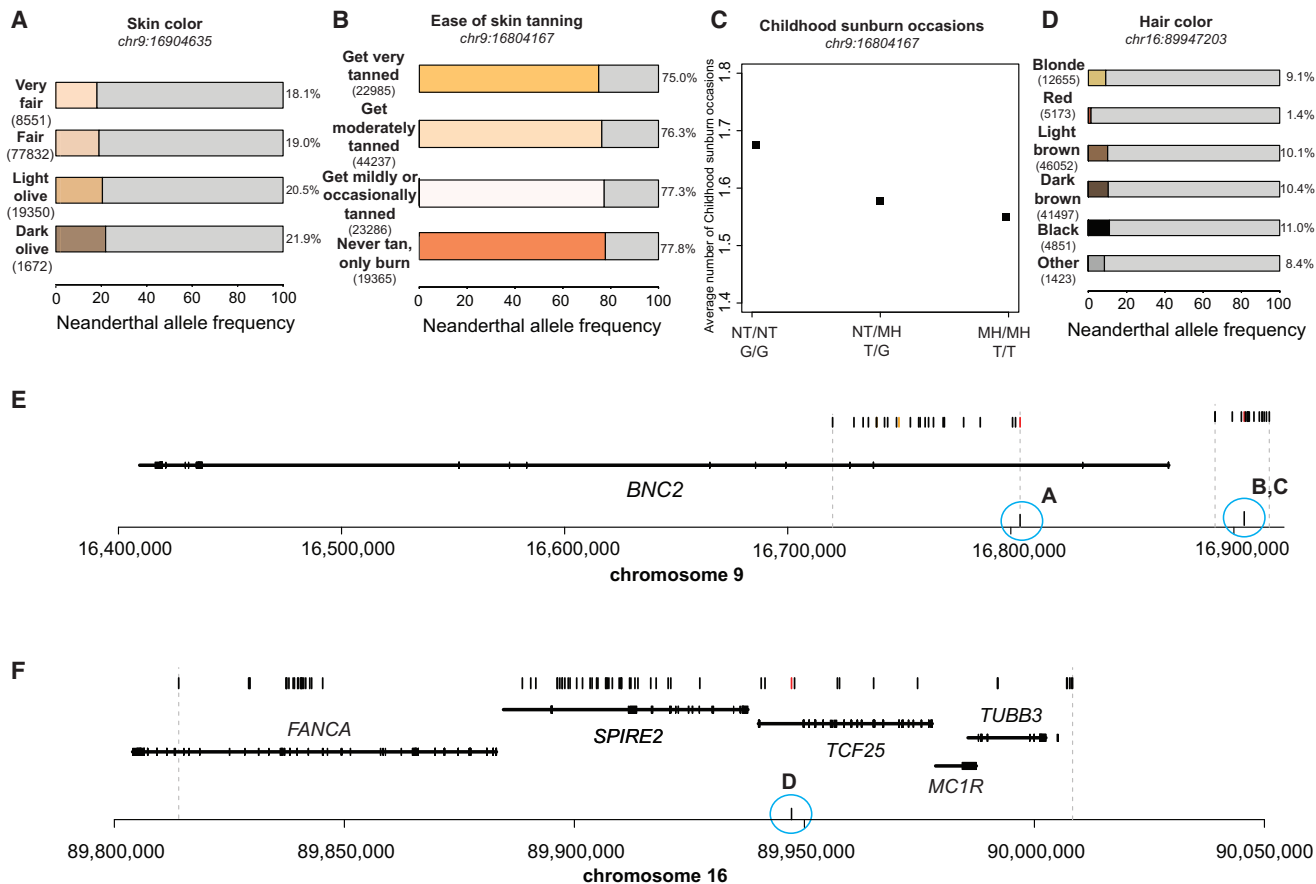
Phenotype	Meta-phenotype	Tag aSNP	Association p Value	Neanderthal Allele Frequency	Data Type	Archaic Haplotype (hg19)	Overlapping Gene(s)	Missense Mutations	Associated eQTLs	FDR ILS Test
Comparative height size at age 10 years	early life factors	chr19: 31,033,240 (rs56199929)	$3.97 \times 10^{-14}$	0.16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

(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)	$1.52 \times 10^{-9}$	0.16	continuous	chr10: 70,009,572–70,059,496	<i>PBLD</i> (MIM: 612189)	–	<i>PBLD</i> : muscle (skeletal), brain (cortex), brain (caudate; basal ganglia), brain (putamen; basal ganglia)  <i>ATOH7</i> (MIM: 609875): artery (coronary), breast (mammary tissue), skin (not sun exposed; suprapubic), minor salivary gland, adrenal gland, pancreas, esophagus (gastroesophageal junction), colon (transverse), adipose (subcutaneous), artery (tibial), brain (cerebellum), artery (aorta), spleen  <i>MYPN</i> (MIM: 608517): brain (putamen; basal ganglia)	0.002
Hair color (natural before graying)	sun exposure	chr6: 503,851 (rs71550011)	$2.91 \times 10^{-9}$	0.07	categorical	chr6: 503,851–544,833	<i>EXOC2</i>	–	<i>EXOC2</i> : cells (transformed fibroblasts)	0.004
Daytime dozing or sleeping (narcolepsy)	sleep	chr10: 94,711,457 (rs112294410)	$4.09 \times 10^{-9}$	0.017	categorical	chr10: 94,574,048–94,756,023	<i>EXOC6</i>	–	–	$<2.2 \times 10^{-22}$
Impedance of leg (right)	impedance measures	chr15: 84,716,986 (rs12902672)	$5.54 \times 10^{-9}$	0.27	continuous	chr15: 84,703,470–85,114,447	<i>ADAMTSL3</i> , <i>GOLGA6L4</i>	<i>ADAMTSL3</i> (chr15: 84,706,461)	<i>NMB</i> : 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  <i>WDR73</i> : heart (atrial appendage), brain (cortex), thyroid, esophagus (muscularis), nerve (tibial), ovary, brain (anterior cingulate cortex; BA24)  <i>SLC28A1</i> : breast (mammary tissue)  <i>ZNF592</i> : lung, pancreas, liver  <i>GOLGA6L4</i> : small intestine (terminal ileum)  <i>SEC11A</i> : brain (anterior cingulate cortex; BA24)  <i>ALPK3</i> : brain (cerebellar hemisphere)  <i>ADAMTSL3</i> : brain (amygdala)	$1.17 \times 10^{-5}$

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 inferred 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.<sup>4,9,11</sup> 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.<sup>21–28</sup> 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.<sup>29</sup> 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 under-represented 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.<sup>30</sup> 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,<sup>11</sup> 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.<sup>12</sup> All of the Neanderthal-like SNPs overlapping *BNC2* on this haplotype have significant scores in a test for recent positive selection in Europeans<sup>31</sup> (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.

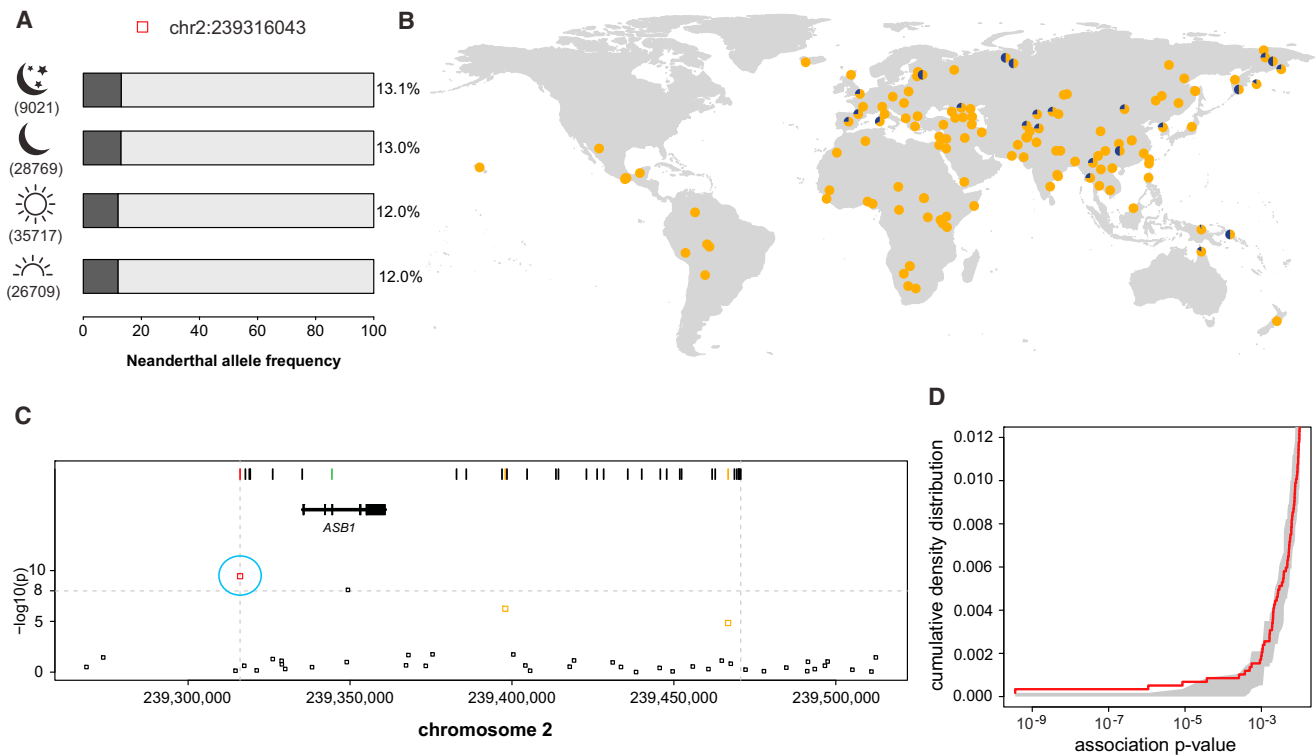
Additionally, it is not clear that phenotypic inference from single variants for which a function is known on

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.<sup>32</sup>

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.<sup>33</sup> Two previous studies of chronotypes identified strongly associated SNPs in the *ASB1* region.<sup>34,35</sup> Of the 540 SNPs with significant genome-wide associations in Hu et al.<sup>34</sup> ( $p < 1.0 \times 10^{-8}$ ), 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.<sup>35</sup> 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.,<sup>34</sup> the association is stronger for the set of aSNPs ( $p$  values ranging from  $3.4 \times 10^{-6}$  to  $2.6 \times 10^{-9}$ ; 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.<sup>35</sup> (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.<sup>36</sup> 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,<sup>37</sup> 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  $-\log_{10}(p)$ ) 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 ( $\sim 1\%–5\%$ ), 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.<sup>4,11,12</sup> Some of the psychiatric and metabolic phenotypes, such as obesity, identified in Simonti et al.<sup>12</sup> 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.<sup>38</sup> 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](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 Neanderthal- 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). Neanderthal 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 Neanderthal 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 Neanderthal 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 Neanderthal 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., Hebringer, S.J., Jarvik, G.P., Kullo, I.J., et al. (2016). The phenotypic legacy of admixture between modern humans and Neanderthals. *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 Neanderthal 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. *Bio-Essays* 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 Mellow, 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., Norderfelt, S., Tan-  
don, 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.