The Population-Specific Impact of Neandertal Introgression on Human Disease

Michael Dannemann ⓘ *

Institute of Genomics, University of Tartu, Estonia

*Corresponding author: E-mail: michael.dannemann@ut.ee.

Abstract

Since the discovery of admixture between modern humans and Neandertals, multiple studies investigated the effect of Neandertal-derived DNA on human disease and nondisease phenotypes. These studies have linked Neandertal ancestry to skin- and hair-related phenotypes, immunity, neurological, and behavioral traits. However, these inferences have so far been limited to cohorts with participants of European ancestry. Here, I analyze summary statistics from 40 disease GWAS (genome-wide association study) cohorts of ~212,000 individuals provided by the Biobank Japan Project for phenotypic effects of Neandertal DNA. I show that Neandertal DNA is associated with autoimmune diseases, prostate cancer and type 2 diabetes. Many of these disease associations are linked to population-specific Neandertal DNA, highlighting the importance of studying a wider range of ancestries to characterize the phenotypic legacy of Neandertals in people today.

Key words: Neandertal admixture, disease, population genetics, genome-wide association studies, human evolution.

Significance

The growing number of non-European cohorts with genotype and phenotype data has enabled researchers to expand the annotation of genetic variants to a larger set of ancestries. This study takes advantage of disease GWAS (genome-wide association study) data from the Biobank Japan and the UK Biobank to compare the phenotypic effects of Neandertal DNA in Asian and European cohorts. The output of this work shows that several Neandertal DNA phenotype associations are linked to population-specific variants, highlighting the importance that phenotype data from more diverse phenotype cohorts are needed to complete our understanding of the phenotypic legacy of Neandertal admixture.

Introduction

When modern humans expanded out of Africa >60,000 years ago, they met and interbred with the archaic humans groups of Neandertals and Denisovans. These admixture events left traces of archaic ancestry in the genomes of people today, with ~2% of the genomes of present-day non-Africans deriving from Neandertals. An additional ~5% of the genomes of present-day Oceanians and <0.5% of the Genomes of East Asians are derived from Denisovans (Wall and Yoshihara Caldeira Brandt 2016). Some of this introgressed archaic DNA likely evolved under negative selection and was rapidly removed following the admixture (Harris and Nielsen 2016; Juric et al. 2016; Petr et al. 2019). Still, at least 40% of the Neandertal genome can be reconstructed from the Neandertal ancestry in people today (Sankararaman et al. 2014; Vernot and Akey 2014), some of which has been suggested to be a target of positive selection (Racimo et al. 2015; Gittelman et al. 2016; Dannemann and Racimo 2018). Since most Neandertal variants segregate at low frequencies in present-day populations, large cohorts with genotype and phenotype data are needed to study their phenotypic effect. Using such cohorts, it has been shown that Neandertal DNA

influences several disease traits, such as immunity, or neurological traits, but also other nondisease phenotypes, like skull morphology, skin- and hair-related phenotypes, or behavioral phenotypes (Quach et al. 2016; Simonti et al. 2016; Dannemann and Kelso 2017; Gunz et al. 2019; Skov et al. 2020). However, the effects were tested in cohorts with participants of largely European ancestry. This reduced diversity in phenotype cohorts is a general pattern in genome-wide association studies (GWAS) today and limits the study of the phenotypic effects of variants specific to non-European populations (Mills and Rahal 2019; Sirugo et al. 2019). It has been noted that analyzing different ancestries can yield new insights into phenotypes (SIGMA Type 2 Diabetes Consortium et al. 2014; Wojcik et al. 2019; Spracklen et al. 2020). Since almost half of the Neandertal variants in the 1,000 Genomes data are found exclusively in Asian populations, it stands to reason that many phenotypic effects of Neandertal variants are yet to be discovered (Dannemann et al. 2020) (fig. 1A and B).

The Biobank Japan Project recently released GWAS maps for multiple disease traits (Nagai et al. 2017; Ishigaki et al. 2020). Here, I analyzed these association maps for their Neandertal DNA content and show that Neandertal variants are strongly associated with multiple autoimmune diseases, prostate cancer and type 2 diabetes. I compared Neandertal-associated phenotypes in the Asian cohort to those discovered in a European cohort, and tested whether Neandertal DNA contributes disproportionately to disease traits by comparing its phenotypic effects to the effects of other nonarchaic variants.

Results

Detection of Significant Disease Associations for Neandertal Variants in the Japan Biobank

I analyzed 40 disease phenotypes for which genome-wide association maps have been made available by the Biobank Japan Project (Ishigaki et al. 2020) (supplementary table S1, Supplementary Material online). The cohort consists of disease-specific subsets of a total of 212,453 Japanese individuals, of which 179,660 individuals were treated for a disease and an additional 32,793 were healthy controls. Between 442 and 40,250 patients were available per disease (supplementary table S1, Supplementary Material online). Genotype data have been generated using three custom genotype arrays and subsequent imputation based on 1000 Genomes Project East Asian haplotype reference haplotypes (see Materials and Methods). The analyzed association maps included individuals and variants that have passed quality control filters (see Materials and Methods).

For each of these association maps, I tested for the presence of association scores for variants that are likely markers for Neandertal admixture in the 1,000 Genomes (1000

Genomes Project Consortium 2012). Variants were defined as potential archaic-like SNPs (aSNPs) based on a previously described method that identified Neandertal introgressed haplotypes in 1,000 Genomes populations (Dannemann et al. 2020) (see Materials and Methods). Such aSNPs are SNPs that are found in non-Africans and Neandertals but are absent in the African Yoruba population. These SNPs are further required to be part of a haplotype that, based on its length, is unlikely to be the result of incomplete lineage sorting (ILS) (see Materials and Methods).

Between 142,102 and 155,479, aSNPs with an allele frequency $>1\%$ were detected across the 40 disease genome-wide association maps. Among them, 18 genomic clusters of aSNPs with high levels of linkage disequilibrium (LD, with $r^2 > 0.5$) showed significant associations ($P < 10^{-8}$) with eight diseases (supplementary table S2 and figs. S1 and S2, Supplementary Material online). Eleven of these clusters harbor non-aSNPs in LD ($r^2 > 0.2$ within ± 1 Mb of the strongest associated aSNP) with stronger disease associations than aSNPs. Thus, we classified these as low confidence clusters and excluded them from further analysis (see Materials and Methods; table 1; supplementary table S2 and fig. S2, Supplementary Material online). Of the remaining six aSNP clusters, four were associated with Graves' disease, dermatitis, rheumatoid arthritis, and prostate cancer, respectively. Neandertal variants linked to prostate cancer on chromosome 2 (rs12621278) were predicted to be protective (table 1). This association has been detected in other GWAS, where it showed a similar risk-decreasing effect on prostate cancer across multiple populations (Eeles et al. 2009). The Neandertal variants associated with Graves' disease, dermatitis, and rheumatoid arthritis, all three autoimmune diseases, were more prevalent in patients than controls (table 1). An aSNP increasing risk for rheumatoid arthritis on chromosome 6 (rs2233433) has been linked to a similar risk-increasing effect for rheumatoid arthritis in another Japanese cohort (Myouzen et al. 2012). This particular aSNP also modifies the amino acid sequence of NFKBIE, a gene that has been shown to be involved in the NF-κB pathway, which is for example responsible for regulating inflammatory processes, cell growth and cell death (Barkett and Gilmore 1999). Another aSNP (rs61742784) that is in high LD to the top-associated aSNP for rheumatoid arthritis in the Japan Biobank is also amino acid changing for TCTE1, a regulator of ciliary and flagellar motility (Castaneda et al. 2017).

The last two aSNP clusters were significantly associated with type 2 diabetes (table 1; fig. 2A and B), and were the strongest associations among the six high-confidence aSNP clusters (table 1 and supplementary fig. S1, Supplementary Material online). One aSNP cluster, which resides on chromosome 6, showed a risk-decreasing effect for the Neandertal variants, whereas the other associated aSNP cluster, on chromosome 17, showed a risk-increasing effect. The top associated aSNP on chromosome 17 (rs2292351) shows high LD

Fig. 1.—Neandertal DNA and its associations with disease phenotypes in Asians and European cohorts. (A) Percentage of Neandertal DNA segregating in present-day populations. Previous work has shown that ~36% of the Neandertal genome can be recovered in 1,000 Genomes Asians and Europeans (Dannemann et al. 2020). Around 16% of that reconstructed Neandertal genome is exclusively found in Asians, a substantially larger fraction than the 2% of the Neandertal genome found only in Europeans. (B) Major cohorts that have been used to predict the phenotypic effect of Neandertal DNA. To date, three large cohorts with genotype and phenotype information have been used to associate Neandertal DNA to effects on disease and nondisease phenotypes (eMERGE network, UK biobank, deCode). (C) Frequency distribution of aSNPs that tag clusters harboring sets of aSNPs in high LD ($r^2 > 0.5$) in the 1,000 Genomes Japanese (y axis) and British (x axis) populations is displayed in gray. The population-specific frequencies of disease-associated aSNPs are highlighted

Table 1

Significantly Associated aSNPs in the Japan and UK Biobanks

	Top Associated aSNP		P value		Beta		Frequency	
	BBJ	UKB	BBJ	UKB	BBJ	UKB	BBJ	UKB
Graves' disease	chr1:160419940_A/**G** rs117201373	—	3.29E-13	—	0.5	—	6.5	0
Prostate cancer	chr2:173321791_T/**G** rs116532339	chr2:173309330_G/**A** rs148644163	5.07E-12	1.06E-09	−0.2	−5.4	21.4	6.2
Dermatitis	chr3:112394029_T/**C** rs75024669	chr3:112394029_T/**C** rs75024669	1.56E-12	—	0.2	—	32.9	0
Type 2 diabetes	chr6:39037662_G/**C** rs9380826	chr6:31003437_G/**A**	1.09E-17	0.99	−0.1	−0.01	18	0.5
Rheumatoid arthritis	chr6:44234621_G/**T** rs28362855	chr6:33768726_A/**T** rs34884883	4.30E-12	0.12	0.2	−0.5	22.1	17.1
Type 2 diabetes	chr17:6946921_C/**G** rs2292351	chr17:6946921_C/**G** rs2292351	6.30E-18	—	0.1	—	7.8	0.5
Dermatitis	chr10:6123716_C/**T** rs61839680	chr10:6115639_T/**G** rs62626322	0.93	2.57E-14	0.003	4.5	21.9	11.3
Prostate cancer	chr19:51361757_T/**C** rs17632542	chr19:51361757_T/**C** rs17632542	—	1.88E-14	—	−6.3	0	7.4

NOTE.—Top associated aSNPs, together with summary statistics (association P value, effect size [in 10^{-3} for the UK biobank] and Neandertal allele frequency in given biobank) in both biobanks (—, if no association data available) are displayed. Neandertal variants are for each of the SNP the alternative allele (highlighted in bold, columns 2 and 3) and summary statistics for the significant associations with a gray background.

($r^2 > 0.8$ in 1,000 Genomes JPT) with three amino acid changing aSNPs in *SLC16A11* (*rs13342692*, *rs75418188*, and *rs75493593*). These variants have previously been linked to a Neandertal haplotype that showed a significant association with type 2 diabetes in Mexicans (SIGMA Type 2 Diabetes Consortium et al. 2014) and likely represents the same association as the one found here in the Japan biobank cohort. This analysis shows that these and other Neandertal variants are linked to effects on type 2 diabetes in Asians, and that Neandertal DNA is associated with both increased and decreased diabetes risks.

Comparison of Neandertal-Linked Associations between Japan and UK Biobanks

In order to determine whether the six aSNP disease associations in Japanese are population-specific or show similar effects in Europeans, I analyzed genome-wide association maps of 26 diseases for which matching GWAS data in the UK biobank were available (supplementary table S1, Supplementary Material online) (Sudlow et al. 2015; Bycroft et al. 2018). The GWAS have been provided by the Neale lab (http://www.nealelab.is/uk-biobank/) and are based on 361,194 individuals and ∼10.8 million SNPs that pass quality filters (see Materials and Methods). The genotype data have been generated using two custom arrays and imputation based on the Haplotype Reference Consortium, the 1000 Genomes Project, and UK10K. The number of cases ranged from 121 to 9,321, which was for all but three diseases lower than the case number in the Japan biobank. The association analyses have been conducted using the phenome scan software (Millard et al. 2018) that incorporates various covariates controlling for ancestry, sex, and age (see Materials and Methods).

In this study, I analyzed 168,511 aSNPs with a minimum allele frequency of 1% for disease association in the UK biobank cohort. In this European cohort and across all 26 diseases, a total of four genomic clusters with significantly associated aSNPs were detected (supplementary table S2 and figs. S1 and S2, Supplementary Material online). Following the same criteria as for the Japanese associations, one of the four clusters was classified as low confidence with

FIG. 1.—Continued

in brown (Japan biobank), blue (UK biobank), and green (both biobanks). (*D* and *E*) The association of genomic variants with prostate cancer in the Japan (*D*) and UK (*E*) biobanks on chromosome 2 are displayed (*y* axis represents −log$_{10}$(*P* value)). Significantly associated aSNPs (darkred) and other aSNPs (red) are highlighted. (*F*) Presence (dark gray) and absence (light gray) of the archaic allele on haplotypes in 1,000 Genomes individuals (*y* axis) carrying aSNPs (*x* axis) across chr2:173,300,939–173,397,690. The genotype status (0: Yoruba allele, 1: archaic allele) of the Altai and Vindija Neandertals as well as the Denisovan across 18 aSNPs are shown on top of the image. (*G*) The frequency of two putative Neandertal haplotypes spanning chr2:173,300,939–173,397,690 in 1,000 Genomes populations. The frequencies of the archaic haplotypes with (purple) and without (green) the archaic alleles at chr2:173,331,972, chr2:173,341,396, and chr2:173,344,846 are displayed together with the frequencies of all other nonarchaic haplotypes (gray).
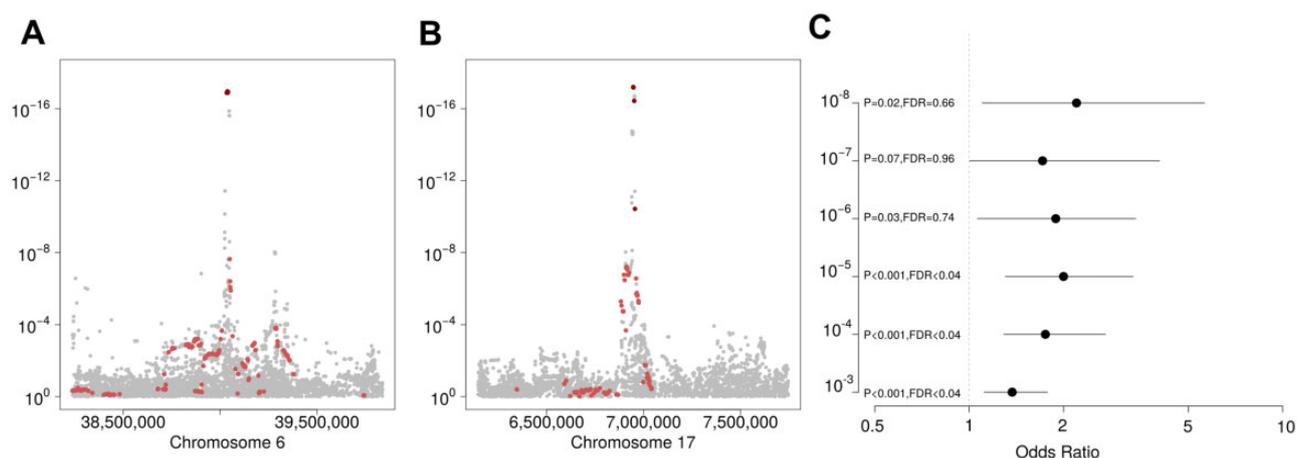
FIG. 2.—Neandertal-linked association with type 2 diabetes in a Japanese cohort. (A and B) Association P values (y axis, $-\log_{10}$-transformed) of two sets of aSNPs on chromosomes 6 (A) and 17 (B) that are significantly associated with type 2 diabetes in the Japan biobank cohort. Significantly associated aSNPs ($P < 10^{-8}$) are shown in darkred, other aSNPs in red, and nonarchaic variants in gray. (C) For six significance cutoffs (y axis), the distribution for the $\log_2$ odds ratios (median, and 95% confidence intervals) calculated as the number of associated tag aSNPs divided by the 1,000 numbers of associated nonarchaic tag SNPs is displayed. The number of tag aSNPs representing sets of aSNPs with LD values of $r^2 > 0.5$ and with association P values below a given cutoff has been compared with the numbers for 1,000 random sets of frequency-matched nonarchaic tag SNP with P value <cutoff.

nonarchaic variants showing stronger associations in the genomic region than the local aSNP cluster. One of the three high-confidence aSNP clusters is located on chromosome 2 and includes aSNPs that show significantly decreased risk of prostate cancer and match the cancer-associated aSNPs in the Japan biobank (table 1 and fig. 1D). A second cluster of aSNPs is on chromosome 19 and is associated with decreased risk of developing prostate cancer in the UK cohort. The aSNPs in this region overlap with variants that have been linked to reduced levels of prostate-specific antigens in an Icelandic cohort (Gudmundsson et al. 2010) and identified to originate from Neandertals (Skov et al. 2020). The third significant aSNP association in the UK biobank is with dermatitis, a disease for which aSNPs have also shown significant association in the Japan Biobank. Although the loci between Japanese and UK populations differ, it is reassuring the Neandertal variants show a risk-increasing effect in both cohorts (table 1).

An explanation for some of the cohort-specific associations may be that the associated aSNPs have a frequency that is too low in the nonsignificant cohort to allow detection. Four of the seven cohort-specific associations show such low allele frequencies (<1%) in the nonsignificant cohort (table 1 and fig. 1C). Conversely, only one out of the four other associations with aSNP frequencies >5% in both biobank cohorts could be replicated. Several genomic factors could be responsible for the lack of replication between cohorts. For example, causal variants or other interacting genetic variants might be missing in one population. In addition, differences in LD are another factor possibly contributing to the missing replication. Given the potential impact of these factors, it is encouraging to see that the shared prostate cancer association was found in spite of the population-specific genomic backgrounds.

Inference of the Likely Archaic Source of Significantly Associated aSNPs

Previous studies have demonstrated that East Asians carry low levels of Denisovan DNA (<0.5%) in addition to their Neandertal ancestry (Meyer et al. 2012; Sankararaman et al. 2016; Vernot et al. 2016). Among Asian populations, Japanese have been shown to carry the lowest amount of Denisovan DNA (Qin and Stoneking 2015). However, despite the low amount of Denisovan DNA, some individual Denisovan variants have reached large frequencies in non-Oceanian populations (Huerta-Sánchez et al. 2014; Racimo et al. 2016). In order to test for the most likely source of the introgressed haplotypes, I calculated the sequence similarity of haplotypes in 1,000 Genomes Populations that carry the associated aSNPs to high-coverage archaic hominin genomes (two Neandertals and one Denisovan; Meyer et al. 2012; Prüfer et al. 2014, 2017) (see Materials and Methods). The haplotypes comprise regions in which other aSNPs show LD of $r^2 > 0.5$ with the top associated aSNP (see Materials and Methods). For seven of the eight associated aSNP clusters, the haplotypes in the 1,000 Genomes carrying the top associated aSNP showed a closer sequence relationship with Neandertals than with the Denisovan (supplementary fig. S3, Supplementary Material online). One region showed almost equal sequence similarity between all three archaic genomes. It has been suggested that the large genetic distance of the sequenced Denisovan individual to the introgressing Denisovan populations can lead to false classifications of Denisovans haplotype as Neandertal (Vernot et al. 2016). However, it is difficult to resolve that problem for individual haplotypes and given the observations from this analysis it is unlikely

that any of the significantly associated aSNPs are of Denisovan ancestry.

In addition, a sequence analysis of haplotypes surrounding significantly associated aSNPs in 1,000 Genomes individuals revealed that the genomic regions at chr2:173,300,939–173,397,690, encompassing aSNPs with significant associations with prostate cancer in both biobank cohorts, are likely composed of two distinct archaic haplotypes in present-day people. In this genomic region, 18 aSNPs can be observed in 1,000 Genomes individuals. Three of the aSNPs in the region (chr2:173,331,972, chr2:173,341,396, and chr2:173,344,846) are in perfect LD with one another, but show only low levels of LD ($r^2 < 0.5$) with the remaining aSNPs in the region. The three variants are found on 63 of the 533 haplotypes in the 1,000 Genomes with at least one archaic allele across the 18 aSNPs (fig. 1F). The carriers of the haplotypes with these three aSNPs are exclusively found in Asia and America, with frequencies of up to 6% (CHB) (fig. 1G). The more common archaic haplotype, without the three aSNPs, can be found in all non-African populations and is particularly prevalent in East Asia where it reaches frequencies of between 14% and 25% and is segregating at <11% in other non-African populations (fig. 1G). The sequence similarity of both haplotypes to the Altai and Vindija Neandertals is almost identical and shows a higher relatedness than to the Denisovan individual (supplementary fig. S3B, Supplementary Material online).

Testing for a Dis-Proportional Contribution of Neandertal-Linked Variants to Disease Phenotypes

In this study, a total of 254,191 aSNPs with a frequency of at least 1% in one of the two biobanks were tested for disease association. Although 27.3% of these aSNPs reached frequencies >1% in both cohorts, 39.0% (UK Biobank) and 33.7% (Biobank Japan) were detected at frequencies of >1% in only one of the two biobank cohorts, respectively. Overall 13% (5/40) and 8% (2/26) of the tested disease phenotypes in the Japan and UK biobanks, respectively, showed significant associations with aSNPs. Although some diseases like prostate cancer and type 2 diabetes showed multiple associations, the absolute number of associations with a given disease is not necessarily informative about the genome-wide impact of Neandertal DNA. For example, the number of cases will likely impact the power of detecting associations, particularly when testing low-frequency variants such as many Neandertal variants. In addition, the level of penetrance and the number of genetic factors will significantly influence the ability to link genetic variants to phenotypic effects. For example, negative selection has also been suggested to play a crucial role in determining the level of how polygenic a trait is (Boyle et al. 2017; Liu et al. 2019). Consequently, higher levels of heritability and lower levels of negative selection on a disease could result in an increased number of strongly disease-associated variants, an effect that needs to be accounted for when assessing the relative contribution of Neandertal DNA on different diseases.

In order to test the proportional phenotypic impact of aSNPs to other genomic variants, I applied a previously used approach that is designed to correct for these phenotype-specific effects (Quach et al. 2016; Simonti et al. 2016; Dannemann and Kelso 2017). Following this methodology, I compared for each disease and each biobank the number of associations of aSNPs with the number of associations with frequency-matched nonarchaic variants. The frequency matching step will control for an equal association detection power when comparing archaic and nonarchaic variants. Additionally, the structure of Neandertal haplotypes in present-day genomes, with multiple introgressed variants found in high LD on them, can likely inflate the number of associated aSNPs with a given trait. In order to correct for that phenomena, all SNPs that show LD values of $r^2 > 0.5$ among them were collapsed and a random tag SNP was chosen among these variants to represent them. If the set of SNPs in LD contained aSNPs the representing tag SNP had to be a random aSNP from this set (see Materials and Methods). SNPs without other SNPs in LD were defined as their own representing tag SNP. The selection of a random SNP from each set will make this analysis independent of this difference of SNPs per set and allow for an unbiased comparison between archaic and nonarchaic sets. Conversely, selecting the top associated SNP per set would statistically favor archaic sets to have tag SNPs with on average stronger associations, due to their on average higher number of SNPs in LD in a given set (Dannemann et al. 2017). Additionally, this procedure also does not implement additional steps to match archaic and nonarchaic sets. For example, no matching for the number of SNPs is implemented. Only a few genomic nonarchaic sets will exist to match archaic sets in the number of SNPs in LD and those nonarchaic sets that likely are not a random subset of nonarchaic genomic variants.

Next, for each disease, I calculated the number of tag aSNPs below a P value cutoff and compared it with 1,000 sets of frequency-matched sets of nonarchaic tag SNPs. In order to assure a robust analysis, that is not dependent on one specific P value cutoff I have repeated this step for six P value cutoffs (10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}). When testing all diseases in the Japan and UK biobanks for a disproportional contribution of Neandertal variants, only one disease showed a significant difference in the number of associated aSNPs compared with nonarchaic SNPs (FDR < 0.05). Dependent on the cutoff, between 11 and 49 significant Neandertal-linked variants were detected, resulting in a between 1.4 and 2.2 times higher number of associations with type 2 diabetes when compared with their nonarchaic counterparts in the Japan biobank (5/6 P values and 3/6 FDRs associated with type 2 diabetes were <0.05, supplementary table S3, Supplementary Material online, fig. 2C). The

contribution of aSNPs to type 2 diabetes in the UK biobank was not different from those of frequency-matched non-archaic variants (FDR = 1, supplementary table S3, Supplementary Material online).

When looking at the initial number of genome-wide significant associations, both type 2 diabetes and prostate cancer showed two such associations in the Japan and UK biobanks, respectively. The enrichment results for prostate cancer in the UK biobank were more moderate than the type 2 diabetes results in the Biobank Japan, with FDR > 0.15 for all six tested P value cutoffs. This analysis demonstrates that only the associations with type 2 diabetes in Japanese populations are unexpectedly high, whereas the associations with prostate cancer in the UK biobank are not larger than expected from other frequency-matched variants in the genome. These results highlight the importance of choosing a proper background set in order to assess the relative contribution of Neandertal DNA to phenotypes.

Discussion

In this study, I analyze the Biobank Japan cohort for significant disease associations with variants that are likely of Neandertal ancestry. Among the six significantly associated aSNPs, three are associated with dermatitis, Graves' disease, and rheumatoid arthritis, which all have been linked to autoimmune processes and all show a higher disease risk for the respective aSNP carriers. These observations are in line with previous studies in European cohorts that have also linked several other Neandertal variants to an effect on immune processes, with some of these effects hypothesized to have been beneficial to modern humans entering new environments (Dannemann et al. 2016; Quach et al. 2016; Sams et al. 2016; Enard and Petrov 2018). This study provides evidence that the impact of Neandertal DNA on the immune system has not been population-specific but influenced both, European and Asian populations. This observation is further supported by the significant dermatitis associations of another risk-increasing aSNP in the UK biobank. However, given the different associated archaic variants in both biobanks, the results of this work also suggest that the effects of how Neandertal DNA influences immunity might be population-specific.

Another disease for which associations were found in both populations was prostate cancer. One association was found in the Biobank Japan, identical to one detected in the UK biobank, together with a second aSNP association identified only in this European cohort. Conversely to the autoimmune disease aSNP associations, for prostate cancer all the significantly associated aSNPs had a protective effect. The region spanning the aSNPs with the shared significant prostate cancer associations in both biobank cohorts show evidence for the presence of two independent archaic haplotypes. Both haplotypes show a higher sequence similarity to the Altai and Vindija Neandertals than to the Denisovan individual and

reach combined allele frequencies of between 19% and 29% in East Asians. One of the haplotypes is absent from Europeans and is in general found at lower allele frequencies of <11%. It is intriguing to observe that the presence of multiple archaic haplotypes, their elevated frequencies in some present-day populations, and their disease-protective associations resemble the characteristics of other archaic variants that have previously been linked to positive selection (Dannemann et al. 2016; Deschamps et al. 2016).

The only disease with multiple aSNP associations in the Japanese cohort was type 2 diabetes, with two significantly associated aSNP clusters. In contrast to the other aSNP disease associations, the two identified associations show opposite effects, one risk-increasing and the second risk-decreasing. These contrasting effects suggest that Neandertal DNA is not influencing this disease in a strictly one-directional manner. The allele frequencies of these two aSNPs was <1% in the UK biobank, limiting the power to assess whether these variants would show the same effects in Europeans. However, no additional type 2 diabetes-associated aSNPs have been detected in the UK biobank, a result that might indicate that the impact of Neandertal DNA on this disease could be population-specific. This hypothesis finds additional support in this study by the overproportional association of aSNPs with type 2 diabetes in the Biobank Japan when compared with association with frequency-matched nonarchaic variants, a result that was not replicated in the UK biobank. Previous studies have demonstrated that the characteristics of type 2 diabetes differ between populations. For example, in Asian populations, a lower onset and BMI has been observed in affected people compared with Europeans (Yoon et al. 2006; Ma and Chan 2013). In addition, type 2 diabetes in Asians was more often associated with developing other comorbidities. These results likely indicate that the genetics underlying this disease differ between populations, which could also be the reason why the enrichment results for Neandertal associations is population-specific. Particularly, the differences in fat distribution between Asians and Europeans have been linked to differences in type 2 diabetes prevalence. Previous studies have linked Neandertal DNA to fat-related phenotypes (Khrameeva et al. 2014; Racimo 2016; Dannemann and Kelso 2017), providing a possible explanation of its link to affect type 2 diabetes.

The enrichment results for type 2 diabetes were unique among all tested diseases in both biobank cohorts. This observation is consistent with previous studies in European cohorts that have shown that most tested disease and nondisease phenotypes show no difference in their number of associations with Neandertal and non-Neandertal variants (Simonti et al. 2016; Dannemann and Kelso 2017). However, also in Europeans, some specific phenotypes like neurological and behavioral traits were shown to have an overproportional amount of associations with Neandertal variants, that is comparable to the enrichment result for type 2 diabetes in the Japan cohort. Future studies analyzing Asian

cohorts with behavioral and neurological data will enable us to evaluate whether the enrichment results for these phenotypes in Europeans are population-specific, similar to type 2 diabetes in Japanese, or shared between populations. It is important to note that the reported enrichment results are based on a method that was designed to allow for a comparison of phenotypic effects of archaic variants to those of their nonarchaic counterparts. And while the allele-frequency matching of archaic and nonarchaic variants and LD-pruning steps correct for frequency differences and differences in haplotype structure and diversity between archaic and nonarchaic DNA, there are other differences that this method does not accounted for. For example, the number of archaic variants of an archaic haplotype is typically higher and it is difficult to estimate how this difference would affect any phenotypic effects. Yet, matching archaic and nonarchaic haplotypes by length and number of SNPs would likely introduce other unwanted biases. In its current form, the method testing whether highly linked phenotype-associated SNPs include more often Neandertal DNA. If that is the case Neandertal DNA has to—at least in part—contribute true positive associations. This setup at the same time allows to address the aforementioned problems when matching archaic and nonarchaic variants, which makes this method still a valuable tool to screen for phenotypes with an unusual number of associations with Neandertal DNA. In order to understand and increase the robustness in such enrichment results, it is therefore necessary to replicate them in other cohorts and with orthogonal approaches and continue to improve the design of methods to compare the phenotypic effects of archaic DNA with an appropriate null model.

Nevertheless, the association results presented in this study underscore the importance to study more genetically diverse phenotype cohorts to investigate evolutionary processes like archaic admixture to understand how much of their phenotypic effects are driven by population-specific archaic DNA. This study is certainly just one of many to come that expand their analyses to Asian cohorts and help to complete the picture of the phenotypic legacy of Neandertal DNA in modern humans. In the future, phenotypic data from populations with high levels of Denisovan DNA will allow us to study the phenotypic genome-wide effects of Denisovan variants as well to compare these effects with those of Neandertal variants. In addition, this study primarily focuses on the phenotypic impact of genetic variation that has been added to the modern human gene pool by Neandertal admixture. In order to investigate the full phenotypic role of archaic DNA, future studies will have to incorporate other factors such as the modification of the allele frequencies of nonarchaic variants by archaic admixture to fully understand the phenotypic legacy of Neandertal and Denisovan admixture. However, future studies investigating genetically diverse cohorts will not only help

to expand our knowledge of how archaic DNA influences phenotype variation in modern humans it will also enable us to test for the robustness of previous results. For example, there is a notable difference in the number of cases in the UK Biobank compared with the Biobank Japan for most tested diseases, with all but three diseases showing less cases in the European cohort (supplementary table S1, Supplementary Material online), providing a possible explanation why this study finds less than half of the number of significant association for archaic variants in the UK Biobank compared with those detected in the Biobank Japan (table 1). In a hypothetical example of a relative genotype disease risk of 1.5 and a significance level of 10^{-8}, 12 of the analyzed 66 GWAS do not provide sufficient power (<0.8) at any allele frequency between 1% and 50%, frequencies most aSNPs segregate at in present-day populations (supplementary fig. S4B and C, Supplementary Material online), to provide robust associations, with 8 of them being UK Biobank GWAS (supplementary fig. S4A, Supplementary Material online). Using the same hypothetical example, only 30 of the 66 GWAS provide a power of at least 0.8 for variants at 5%, suggesting that particularly for those GWAS with low sample sizes, variants with relative genotype risks of 1.5 or lower are likely not detected. In addition to power, differences in population structure between cohorts of different ancestries can bias any comparative analysis. For example, previous studies have pointed out how population structure in the UK Biobank can lead to biased phenotypic inferences (Berg et al. 2019; Haworth et al. 2019). Analyzing cohorts with sufficient and matching sample sizes and phenotyping will be crucial to increase comparability between biobanks and evaluate previous findings for their robustness.

Finally, it is also important to keep in mind that studies like this one, report associations with Neandertal DNA, that will still need to be subject to functional testing to investigate whether the archaic variants are the functional basis for the phenotypic effects. In addition, the reported associations all required an aSNPs as the top associated variant. It is important to note that even this conservative step doesn't guarantee that any of these variants are actually causal (Schaid et al. 2018). Howerever, four of the eight associated loci and their corresponding aSNPs in this study could be replicated in at least one additional cohort, thus making them a robust and confident set of Neandertal variants that affect disease variation. However, some disease associations with both case numbers and aSNPs with allele frequencies in both cohorts that allow for a robust association inference did not replicate between cohorts. Studying these and other Neandertal introgressed DNA in vitro will help us to determine whether it is the archaic variants that are actually causal and understand their molecular mechanisms. The output of such experiments can

form the basis to study Neandertal-specific biology, and help us to ascertain how it influences modern humans today.

Materials and Methods

Genome-Wide Association Maps

Biobank Japan Project

This study is using summary statistics from GWAS maps for 40 diseases (supplementary table S1, Supplementary Material online) that have recently been made publicly available by the Biobank Japan Project (Nagai et al. 2017; Ishigaki et al. 2020). These GWAS have been conducted using 212,453 Japanese individuals and the detailed genotype and association analysis information is available in Ishigaki et al. (2020). The genome-wide association maps included samples and variants that have passed quality control filters provided by the Biobank Japan. Briefly, individuals have been genotyped using combinations of three custom arrays (Illumina HumanOmniExpressExome BeadChip, Illumina 356 HumanOmniExpress, and HumanExome BeadChips). Subsequently, samples have been excluded based on quality measures which included minimum genotype call rates (0.98), non-East Asians ancestry inferred by principal component analysis. In addition, genomic variants have been excluded if they did not meet a minimum call rate (0.99), had P values for Hardy–Weinberg equilibrium $<1.0 \times 10^{-6}$, less than five heterozygous calls or a concordance rate of $>99.5\%$ between arrays-based variants and whole-genome sequencing-based variants measured in a subset of 939 samples. Further, additional genotypes have been imputed based on 1,000 Genome Project samples (phase 3) as a reference. Variants with an imputation quality of $R^2 < 0.7$ have been excluded.

The genome-wide association analyses by the Biobank Japan have been conducted using a generalized linear mixed model following a two-step approach. The first step generated a null logistic mixed model with covariates that was applied to the genotype data. The covariates included age, the top five ancestry principal components, and sex, the latter only for diseases with both, males and females, tested. In the second step, a single variant association test based on the imputed variant dosages has been performed.

UK Biobank

This study uses GWAS summary statistics from the UK biobanks (Sudlow et al. 2015; Bycroft et al. 2018) for 26 diseases that have matching data for any of the 40 diseases from the Biobank Japan (supplementary table S1, Supplementary Material online). Details on the analysis can be found on the lab's home page (http://www.nealelab.is/uk-biobank/) and at http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas and http://www.nealelab.is/blog/2019/10/24/updating-snp-heritability-results-from-4236-phenotypes-in-uk-biobank. Briefly, the GWAS include 361,194 individuals that have passed quality controls provided by the UK biobank. These quality controls have excluded samples with non-British ancestry based on principal component-based ancestry estimates, that are related, had sex chromosome aneuploidies or withdrawn from the biobank. In addition, the association analyses have been restricted to ~10.8 million SNPs with a minor allele frequency $>0.1\%$, a Hardy–Weinberg equilibrium P value $>1 \times 10^{-10}$ in the 337,199 individuals and an UK biobank provided INFO score >0.8. The association studies have been performed using PHESANT (Millard et al. 2018) including as covariates the first 20 PCs, sex, age, age^2, sex $\times$ age, and sex $\times$ age^2 for diseases with male and female samples and only the nonsex covariates of this model for sex-specific diseases.

Additional to the filters used by the GWAS provided by the Biobank Japan and UK biobank, the analyses in this study were restricted to autosomal variants with a minimal minor allele frequency of 1% in the individuals included in a given disease cohort.

Disease Associations with SNPs Likely of Neandertal Origin

Definition of Putative Neandertal SNPs in GWAS Cohorts

In order to define SNPs that are likely of Neandertal origin in each of the GWAS cohorts, this study used previously generated Neandertal haplotype information from the 1,000 Genomes Project (Dannemann et al. 2020). These haplotypes have been identified using an approach that combines allele-sharing patterns of variants between non-Africans and Neandertals and a haplotype length that is more likely the result of admixture than of ILS. Due to the admixture between modern humans and Neandertals, ~55,000 years ago, the average length of Neandertal haplotypes in modern humans is expected to be longer than the length of segments that are the result of ILS (Sankararaman et al. 2012). In order to identify putative Neandertal haplotypes and formally test that the detected haplotypes are not the results of ILS, this approach employs a two-step model. First, this method considers variants as potential candidates for Neandertal introgression if they 1) are present in both at least one 1,000 Genome non-African and in a homozygous state in the Vindija Neandertal (Prüfer et al. 2017) and 2) absent in the 1,000 Genomes Project (phase 3; 1000 Genomes Project Consortium 2012) Yoruba population. The Vindija Neandertal has been chosen as a reference individual because of its genetically closer relationship to the introgressing Neandertal population when compared with the Altai Neandertal (Prüfer et al. 2014, 2017). Additionally, the Yoruba population has been selected in this analysis because of its low Neandertal ancestry, which has been shown to be the lowest among all 1,000 Genomes populations (Prüfer et al. 2017). The identified variants that have been identified based on this methodology are referred to as aSNPs.

In the second step of this method, the lengths of genomic regions where such variants consequently appear in a given 1,000 Genomes individual have been tested for their probability to be consistent with ILS. This approach to detect putatively introgressed regions, previously implemented by Huerta-Sánchez et al. (2014) and modified using updated ages of the Neandertal individuals in Dannemann et al. (2016) has been applied to detect regions in all non-African 1,000 Genomes Project individuals (phase 3). The test requires the calculation of the expected length of ILS segments given the local recombination rate. Based on the haplotype length, a P value is calculated that reflects the probability for a segment of that length or longer to be consistent with ILS. The test has been performed using two recombination maps (Kong et al. 2002; International HapMap 3 Consortium et al. 2010). Haplotypes with a Benjamini–Hochberg method corrected P value <0.05 based on at least one of these recombination maps have been considered to be putatively introgressed Neandertal haplotypes. The final list of aSNPs has then been defined as the subset of the original list of aSNPs from step one that have been found to form one of these haplotypes (Dannemann et al. 2020). I used these aSNPs as marker variants for Neandertal introgression throughout the analyses in this manuscript.

Significant Disease Associations with Putative Neandertal Variants

All aSNPs with an association P value $<10^{-8}$ in a given cohort were defined as significant associations (supplementary table S2, Supplementary Material online). To provide additional confidence in an actual link between the putative introgressed variants and the disease, the associated aSNPs were further required to have no nonarchaic SNP in LD ($r^2 > 0.2$) with a stronger association in the surrounding genomic region (± 1 Mb, supplementary figs. S1 and S2 and table S2, Supplementary Material online). All associated aSNPs in a given region that showed high levels of LD ($r^2 > 0.5$) were then combined into an aSNP cluster with a tag aSNP, which was defined as the aSNP among this cluster that had the lowest association P value. The r^2 calculations were performed with plink (parameters: –ld-window-r2 0.5 –ld-window 99999) (Purcell et al. 2007) using the 1,000 Genomes Project British (GBR) and Japanese (JPT) populations, each chosen to match the UK and Japan biobank cohorts, respectively. All aSNPs with $r^2 > 0.5$ to a given tag aSNP were considered to be linked to Neandertal haplotypes contributing to the association signal in the region (table 1). For those diseases that were available in both biobanks, the significant tag aSNPs together with its aSNPs in LD were tested for their association scores in the complementary biobank. The most significant aSNP and its association statistics in the complementary biobank are shown in table 1. Finally, aSNPs that modify the

amino acid sequence were detected using the variant effect predictor (McLaren et al. 2016).

Likely Archaic Source of Significantly Associated Neandertal-Like Variants

In addition to Neandertal DNA, individuals from East Asian populations have been described to carry low levels ($<0.5\%$) of Denisovan DNA (Reich et al. 2010, 2011; Meyer et al. 2012). In order to test whether some of the aSNPs with a significant association are possibly the result of Denisovan admixture, genomic regions of all 1,000 Genomes haplotypes with and without the aSNPs were tested for their sequence similarity to the genome sequences of the Altai and Vindija Neandertal and the Denisovan individuals. The range of the tested genomic regions was defined by the most distant aSNPs that show $r^2 > 0.5$ with the top associated aSNP. The results of this sequence comparison are shown in supplementary figure S3, Supplementary Material online.

Comparison of the Phenotypic Effects of Putative Neandertal Variants to Those of Nonarchaic Variants

Sets of Neandertal variants in present-day human genomes have been shown to have high levels of LD among them due to the relatively recent admixture between modern humans and Neandertals $\sim 55,000$ years ago (Sankararaman et al. 2012). In addition, most of the Neandertal variants found in people today are segregating at low frequencies, which are consistent with the $\sim 2\%$ of Neandertal DNA in non-Africans today. When comparing the number of disease associations with Neandertal variants to the number of associations for nonarchaic variants, the features of Neandertal DNA need to be addressed in order to eliminate any bias from the comparison. Previous studies have implemented a two-step approach to address these differences between Neandertal and nonarchaic DNA that this study used as well (Quach et al. 2016; Simonti et al. 2016; Dannemann and Kelso 2017). In the first step, genomic variants in each biobank cohort were combined into sets of variants that show an LD of $r^2 > 0.5$ among them. For each such set, a random tag SNP was selected to represent the variants in LD. For the sets containing aSNPs, a random aSNP is chosen. Sets without aSNPs have a random SNP representing them. SNPs without other variants in LD represent themselves. In the second step, for each disease, the total number of tag aSNPs with an association P value below a significance cutoff was calculated. This number was then compared with the total number of 1,000 sets of nonarchaic tag SNPs with P value $<$cutoff. For this analysis, six different P value cutoffs were tested: 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8} (supplementary table S3, Supplementary Material online). Each of these random sets was sampled to match the tag aSNPs in amount and frequency. The frequencies were obtained from the analyzed disease cohort. With a decreasing

detection power for variants of lower minor allele frequencies, this step was designed to provide equal statistical association power between aSNPs and nonarchaic variants. The results of this enrichment analysis are shown in supplementary table S3, Supplementary Material online.

Power Calculation

Power calculation were performed using the GAS power Calculator (http://csg.sph.umich.edu/abecasis/gas_power_calculator/) (Johnson and Abecasis 2017) applying a relative genotype risk of 1.5, the case.control numbers for each of the 66 GWAS analyzed in this study (supplementary table S1, Supplementary Material online), a significance level of 10^{-8}, and varying allele frequencies of between 1% and 50%.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Data Availability

GWAS data are available at: Biobank Japan: http://jenger. riken.jp/en/result and UK Biobank: http://www.nealelab.is/uk-biobank.

Literature Cited

1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.

Barkett M, Gilmore TD. 1999. Control of apoptosis by Rel/NF-kappaB transcription factors. Oncogene 18(49):6910–6924.

Berg JJ, et al. 2019. Reduced signal for polygenic adaptation of height in UK Biobank. Elife 8:e39725.

Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell 169(7):1177–1186.

Bycroft C, et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562(7726):203–209.

Castaneda JM, et al. 2017. *TCTE1* is a conserved component of the dynein regulatory complex and is required for motility and metabolism in mouse spermatozoa. Proc Natl Acad Sci U S A. 114(27):E5370–E5378.

Dannemann M, Andrés AM, Kelso J. 2016. Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human Toll-like receptors. Am J Hum Genet. 98(1):22–33.

Dannemann M, et al. 2020. Human stem cell resources are an inroad to Neandertal DNA functions. Stem Cell Rep. 15(1):214–225.

Dannemann M, Kelso J. 2017. The contribution of Neanderthals to phenotypic variation in modern humans. Am J Hum Genet. 101(4):578–589.

Dannemann M, Prüfer K, Kelso J. 2017. Functional implications of Neandertal introgression in modern humans. Genome Biol. 18(1):61.

Dannemann M, Racimo F. 2018. Something old, something borrowed: admixture and adaptation in human evolution. Curr Opin Genet Dev. 53:1–8.

Deschamps M, et al. 2016. Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. Am J Hum Genet. 98(1):5–21.

Eeles RA, et al. 2009. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat Genet. 41(10):1116–1121.

Enard D, Petrov DA. 2018. Evidence that RNA viruses drove adaptive introgression between Neanderthals and modern humans. Cell 175(2):360–371.e13.

Gittelman RM, et al. 2016. Archaic hominin admixture facilitated adaptation to out-of-Africa environments. Curr Biol. 26(24):3375–3382.

Gudmundsson J, et al. 2010. Genetic correction of PSA values using sequence variants associated with PSA levels. Sci Transl Med. 2(62):62ra92.

Gunz P, et al. 2019. Neandertal introgression sheds light on modern human endocranial globularity. Curr Biol. 29(5):895.

Harris K, Nielsen R. 2016. The genetic cost of Neanderthal introgression. Genetics 203(2):881–891.

Haworth S, et al. 2019. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. Nat Commun. 10(1):333.

Huerta-Sánchez E, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194–197.

International HapMap 3 Consortium, et al. 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467:52–58.

Ishigaki K, et al. 2020. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat Genet. 52(7):669–679.

Johnson JL, Abecasis GR. 2017. GAS power calculator: web-based power calculator for genetic association studies. Cold Spring Harb Lab. 10.1101/164343.

Juric I, Aeschbacher S, Coop G. 2016. The strength of selection against Neanderthal introgression. PLoS Genet. 12(11):e1006340.

Khrameeva EE, et al. 2014. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. Nat Commun. 5:3584.

Kong A, et al. 2002. A high-resolution recombination map of the human genome. Nat Genet. 31(3):241–247.

Liu X, Li YI, Pritchard JK. 2019. Trans effects on gene expression can drive omnigenic inheritance. Cell 177(4):1022–1034.e6.

Ma RCW, Chan JCN. 2013. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 1281(1):64–91.

McLaren W, et al. 2016. The Ensembl variant effect predictor. Genome Biol. 17(1):122.

Meyer M, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338(6104):222–226.

Millard LAC, Davies NM, Gaunt TR, Davey Smith G, Tilling K. 2018. Software application profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. Int J Epidemiol. 47(1):29–35.

Mills MC, Rahal C. 2019. A scientometric review of genome-wide association studies. Commun Biol. 2(1):9.

Myouzen K, et al. 2012. Functional variants in *NFKBIE* and *RTKN2* involved in activation of the NF-κB pathway are associated with rheumatoid arthritis in Japanese. PLoS Genet. 8(9):e1002949.

Nagai A, et al. 2017. Overview of the BioBank Japan Project: study design and profile. J Epidemiol. 27(3):S2–S8.

Petr M, Pääbo S, Kelso J, Vernot B. 2019. Limits of long-term selection against Neandertal introgression. Proc Natl Acad Sci U S A. 116(5):1639–1644.

Prüfer K, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505(7481):43–49.

Prüfer K, et al. 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science 358(6363):655–658.

Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3):559–575.

Qin P, Stoneking M. 2015. Denisovan ancestry in east Eurasian and native American populations. Mol Biol Evol. 32(10):2665–2674.

Quach H, et al. 2016. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. Cell 167(3):643–656.e17.

Racimo F. 2016. Testing for ancient selection using cross-population allele frequency differentiation. Genetics 202(2):733–750.

Racimo F, et al. 2016. Archaic adaptive introgression in TBX15/WARS2. Mol Biol Evol. 34(3):509–524.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 16(6):359–371.

Reich D, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468(7327):1053–1060.

Reich D, et al. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am J Hum Genet. 89(4):516–528.

Sams AJ, et al. 2016. Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. Genome Biol. 17(1):246.

Sankararaman S, et al. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507(7492):354–357.

Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. Curr Biol. 26(9):1241–1247.

Sankararaman S, Patterson N, Li H, Pääbo S, Reich D. 2012. The date of interbreeding between Neandertals and modern humans. PLoS Genet. 8(10):e1002947.

Schaid DJ, Chen W, Larson NB. 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet. 19(8):491–504.

SIGMA Type 2 Diabetes Consortium, et al. 2014. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 506:97–101.

Simonti CN, et al. 2016. The phenotypic legacy of admixture between modern humans and Neandertals. Science 351(6274):737–741.

Sirugo G, Williams SM, Tishkoff SA. 2019. The missing diversity in human genetic studies. Cell 177(4):1080.

Skov L, et al. 2020. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. Nature 582(7810):78–83.

Spracklen CN, et al. 2020. Identification of type 2 diabetes loci in 433,540 East Asian individuals. Nature 50:1505.

Sudlow C, 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(3):e1001779.

Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343(6174):1017–1021.

Vernot B, et al. 2016. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science 352(6282):235–239.

Wall JD, Yoshihara Caldeira Brandt D. 2016. Archaic admixture in human history. Curr Opin Genet Dev. 41:93–97.

Wojcik GL, et al. 2019. Genetic analyses of diverse populations improves discovery for complex traits. Nature 570(7762):514–518.

Yoon K-H, et al. 2006. Epidemic obesity and type 2 diabetes in Asia. Lancet 368(9548):1681–1688.

Associate editor: Satta Yoko