Quantitative Human Paleogenetics: What can Ancient DNA Tell us About Complex Trait Evolution?

Evan K. Irving-Pease[1]*, Rasa Muktupavela[1], Michael Dannemann[2] and Fernando Racimo[1]*

[1] Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark, [2] Center for Genomics, Evolution and Medicine, Institute of Genomics, University of Tartu, Tartu, Estonia

Genetic association data from national biobanks and large-scale association studies have provided new prospects for understanding the genetic evolution of complex traits and diseases in humans. In turn, genomes from ancient human archaeological remains are now easier than ever to obtain, and provide a direct window into changes in frequencies of trait-associated alleles in the past. This has generated a new wave of studies aiming to analyse the genetic component of traits in historic and prehistoric times using ancient DNA, and to determine whether any such traits were subject to natural selection. In humans, however, issues about the portability and robustness of complex trait inference across different populations are particularly concerning when predictions are extended to individuals that died thousands of years ago, and for which little, if any, phenotypic validation is possible. In this review, we discuss the advantages of incorporating ancient genomes into studies of trait-associated variants, the need for models that can better accommodate ancient genomes into quantitative genetic frameworks, and the existing limits to inferences about complex trait evolution, particularly with respect to past populations.

Keywords: aDNA, paleogenetics, GWAS, polygenic adaptation, complex traits

INTRODUCTION

The last decade has seen dramatic advances in our understanding of the genetic architecture of polygenic traits (Visscher et al., 2017). The advent of genome-wide association studies (GWAS), with large sample sizes and deep phenotyping of individuals, has led to the identification of thousands of loci associated with complex traits and diseases (MacArthur et al., 2017; Bycroft et al., 2018; Buniello et al., 2019). The resulting associations, and their inferred effect sizes, have enabled the development of so-called polygenic risk scores (PRS), which summarise either the additive genetic contribution of single nucleotide polymorphisms (SNPs) to a quantitative trait (e.g., height), or the increase in probability of a binary trait (e.g., major coronary heart disease) (Dudbridge, 2013). For some well-characterised medical traits, like cardiovascular disease, the predictive value of PRS has led to their adoption in clinical settings (Knowles and Ashley, 2018); however, the accuracy of PRS remains limited to populations closely related to the original GWAS cohort (Martin et al., 2019) and can vary within populations due to age, sex and socioeconomic status (Mostafavi et al., 2020). Ancient genomics has yielded considerable insights into natural selection on large-effect variants (Malaspinas, 2016; Dehasque et al., 2020), and an increasing number of studies are also now utilizing ancient genomes to learn about polygenic adaptation; the process by which natural selection acts on a trait with a large number of genetic loci, leading to changes in allele frequencies

at many sites across the genome. Among these studies, the most commonly inferred complex traits are pigmentation and standing height.

ANCIENT DNA AND COMPLEX TRAIT GENOMICS

Skin, hair and eye pigmentation are among the least polygenic complex traits; though more than a hundred pigmentation-associated loci have been found, their heritability is largely dominated by large-effect common SNPs (Sulem et al., 2007; Eiberg et al., 2008; Han et al., 2008; Sturm et al., 2008; Hider et al., 2013; Liu et al., 2015; O'Connor et al., 2019). Additionally, several of these variants have signatures of past selective sweeps detectable in present-day genomes (Lao et al., 2007; Sabeti et al., 2007; Pickrell et al., 2009; Rocha, 2020). Nevertheless, genomic analyses in previously understudied populations—like sub-Saharan African groups—suggest that perhaps hundreds of skin pigmentation alleles of small effect remain to be found (Martin et al., 2017b). Similarly, recent studies have shown that eye pigmentation is far more polygenic than previous thought (Simcoe et al., 2021). Recent quantitative and molecular genomic studies are painting an increasingly complex picture of the architecture of these traits, featuring more considerable roles for epistasis, pleiotropy and small-effect variants than were previously assumed (for an extensive review of skin pigmentation, see Quillen et al., 2019).

Recently, ancient DNA (aDNA) studies have attempted to reconstruct pigmentation phenotypes in ancient human populations, although the extent to which these predictions are accurate remains uncertain. These reconstructions have been mostly focused on ancient individuals from Western Eurasia, due to the relatively higher abundance of SNP-phenotype associations from European-centric studies, and the poor portability of gene-trait associations to more distantly related populations (Martin et al., 2017a, 2019). For example, Olalde et al. (2014) queried pigmentation-associated SNPs in genomes of Mesolithic hunter-gatherer remains from western and central Eurasia, and suggested that the lighter skin colour characteristic of Europeans today was not widely present in the continent before the Neolithic. González-Fortes et al. (2017) analysed Mesolithic and Eneolithic genomes from central Europe, and inferred dark hair, brown eyes and dark skin pigmentation for the Mesolithic individuals and dark hair, light eyes, and lighter skin pigmentation for an Eneolithic individual. Similarly, Brace et al. (2019) inferred pigmentation phenotypes for Mesolithic and Neolithic genomes from western Europe, and reported that the so-called "Cheddar Man," a Mesolithic individual found in England, had blue/green eyes and dark to black skin, in contrast to later Neolithic individuals with dark to intermediate skin pigmentation. Contrastingly, Günther et al. (2018) found elevated frequencies of light skin pigmentation alleles in individuals from the Scandinavian Mesolithic, suggestive of early environmental adaptation to life at higher latitudes. These reconstructions have also been carried out in individuals with no skeletal remains; for example, Jensen et al. (2019) used pigmentation-associated SNPs to infer the skin, hair and eye

colour of a female individual whose DNA was preserved in a piece of birch tar "chewing gum."

Some aDNA studies have sought to systematically investigate how pigmentation-associated variants were introduced and evolved in the European continent. Wilde et al. (2014) was one of the first studies to provide aDNA-based evidence that skin, hair, and eye pigmentation-associated alleles have been under strong positive selection in Europe over the past 5,000 years. The first large-scale population genomic studies (Allentoft et al., 2015; Haak et al., 2015; Mathieson et al., 2015) showed that major effect alleles associated with light eye colour likely rose in frequency in Europe before alleles associated with light skin pigmentation. More recently, Ju and Mathieson (2021) argued that the increase in light skin pigmentation in Europeans was primarily driven by strong selection at a small proportion of pigmentation-associated loci with large effect sizes. When testing for polygenic adaptation using an aggregation of all known pigmentation-associated variants, they did not detect a statistically significant signature of selection.

The other trait that has shared comparable prominence with pigmentation in the aDNA literature is standing height. In contrast to pigmentation, the genetic architecture of height is highly polygenic (Yang et al., 2015; Bycroft et al., 2018; Yengo et al., 2018). The heritability of this trait is dominated by a large number of alleles with small effect sizes, and shows strong evidence for negative selection in present-day populations (O'Connor et al., 2019). Studies of the genetic component of height in ancient populations have shown that ancient West Eurasian populations were, on average, more highly differentiated for this trait than present-day West Eurasian populations, and more so than one would predict from genetic drift alone (Mathieson et al., 2015; Martiniano et al., 2017; Cox et al., 2019). Cox et al. (2019) compared predicted genetic changes in height in ancient populations to inferred height changes estimated via skeletal remains. They concluded that the changes in inferred standing height were partially predicted by genetics; with both measures remaining relatively constant between the Mesolithic and Neolithic, and increasing between the Neolithic and Bronze Age. A follow-up study by Cox et al. (2021) used polygenic scores for height to show that PRS predicts 6.8% of the observed variance in femur length in ancient skeletons, after controlling for other variables. This is approximately one quarter of the predictive accuracy of PRS in present-day populations; which the authors attribute to the low-coverage aDNA data used in their study. Contrastingly, Marciniak et al. (2021) used the discordance between PRS for height, calculated from aDNA, and height inferred from the corresponding skeletal remains, to argue that Neolithic individuals were shorter than expected due to either poorer nutrition or increased disease burden, relative to hunter-gatherer populations.

However, the inference of standing height from skeletal remains is not without its own problems. Both Cox et al. (2021) and Marciniak et al. (2021) used the method developed by Ruff et al. (2012) to estimate stature from skeletal remains. Nevertheless, their respective estimates of stature—based on femur length—varied between some of the individuals included in both studies. Where multiple skeletal elements were available for ancient individuals, Marciniak et al. (2021) also produced

separate stature estimates from femur, tibia, humerus and radius length, which varied substantially within some individuals; highlighting the uncertainty in estimates of stature from skeletal remains.

INFERRING COMPLEX TRAITS IN ARCHAIC HOMINIDS

The availability of genome sequences from archaic humans, like Neanderthals and Denisovans, has greatly expanded our understanding of their demographic history and interactions with modern humans (Meyer et al., 2012; Prüfer et al., 2014, 2017). However, little is known about complex traits in archaic humans, besides what can be inferred directly from their skeletal remains. In the case of Denisovans, such remains are presently limited to a few teeth, a mandible and other small bone fragments, making it difficult to make confident inferences of their biology (Meyer et al., 2012; Sawyer et al., 2015; Slon et al., 2017; Chen et al., 2019). However, past admixture events with archaic human groups have left a genetic legacy in present-day people, providing a possible inroad to study archaic human biology (Sankararaman et al., 2012). Today, around 2% of the genomes of non-African humans are known to be descended from Neandertals, and an additional ∼5% of the genomes of people in Oceania can be traced back to Denisovans (Sankararaman et al., 2014, 2016; Vernot and Akey, 2014; Vernot et al., 2016).

Knowledge about admixture between archaic and modern humans has led to a recent flurry of exploratory studies concerning the potential impact of archaic variants on complex traits in present-day populations. Various approaches have been used to identify introgressed archaic DNA putatively under positive selection in modern humans (Khrameeva et al., 2014; Sankararaman et al., 2014, 2016; Vernot and Akey, 2014; Perry et al., 2015; Gittelman et al., 2016; Vernot et al., 2016; Racimo et al., 2017b). Overall, these studies have shown that archaic DNA is linked to pathways related to metabolism, as well as skin and hair morphology. Via association studies, Neanderthal variants in specific loci have been shown to influence several disease and immune traits, as well as skin and hair colour, behavioural traits, skull shape, pain perception and reproduction (Sankararaman et al., 2014; Dannemann et al., 2016; Sams et al., 2016; Gunz et al., 2019; Skov et al., 2020; Zeberg and Pääbo, 2020, 2021; Zeberg et al., 2020a,b).

Additionally, comparisons between the combined phenotypic effects of Neandertal variants and frequency-matched non-archaic variants have revealed that Neanderthal DNA is over-proportionally associated with neurological and behavioural phenotypes, as well as viral immune responses and type 2 diabetes (Quach et al., 2016; Simonti et al., 2016; Dannemann and Kelso, 2017; Dannemann, 2021). These groups of phenotypes may be linked to environmental factors, such as ultraviolet light exposure, pathogen prevalence and climate, that substantially differed between Africa and Eurasia. It has been suggested that the over-proportional contribution of Neandertal DNA to immunity and behavioural traits in present-day humans might be a reflection of adaptive processes in Neandertals to these environmental differences. In comparison, much less is known

about the impact of Denisovan DNA on complex traits, because limited phenotypic data are presently available from present-day populations. However, individual Denisovan-like haplotypes found in high frequencies in some human populations have been associated with high altitude adaptation and fat metabolism (Huerta-Sánchez et al., 2014; Racimo et al., 2017a).

One key limitation to these approaches is that only about 40–50% of the Neandertal genome can be recovered in present-day humans, and therefore discoverable in such analyses (Sankararaman et al., 2014; Vernot and Akey, 2014; Skov et al., 2020). Furthermore, the majority of tested cohorts used for such studies are of European ancestry, which limits analyses to archaic variants present in these populations. This is particularly notable since Neandertal phenotype associations in European and Asian populations have been shown to contain population-specific archaic variants (Dannemann, 2021). It has also been shown that negative selection, soon after admixture, has played an important role in removing some of the missing segments of archaic DNA (Harris and Nielsen, 2016; Juric et al., 2016; Petr et al., 2019). It is therefore possible that missing segments of archaic DNA had strong phenotypic effects. For archaic DNA that does persist in present-day populations, much of it is segregating at low allele frequencies, making it difficult to confidently link it to phenotypic effects.

Furthermore, it remains questionable how transferable any phenotypic associations are between modern and archaic humans, given the difficulties of transferring associations between present-day populations (Martin et al., 2017a; Duncan et al., 2019). All of the above studies have used gene-trait association information from analyses carried out in modern humans. It remains undetermined if the phenotypic effects of archaic DNA in present-day populations are a reliable proxy for phenotypic effects in archaic humans themselves.

Recent studies have also aimed to predict the phenotypic effects of archaic DNA without relying on introgression in present-day populations (see **Figure 1**). Colbran et al. (2019) used a machine learning algorithm, trained on genetic variation in present-day humans, to infer putative regulatory effects on variation present only in Neandertal genomes. Gokhman et al. (2020a,b) used aDNA damage patterns to infer a DNA methylation map of the Denisovan genome, and linked the inferred regulatory patterns to loss-of-function phenotypes, in order to predict their skeletal morphology and vocal and facial anatomy. It remains to be seen how successful these approaches are at predicting archaic human phenotypes. A possible inroad into validation could rest on functional assays for testing and evaluating the phenotypic impact of archaic DNA (Dannemann et al., 2020; Dannemann and Gallego Romero, 2021; Trujillo et al., 2021).

THE CHALLENGE OF DETECTING POLYGENIC ADAPTATION IN ANCIENT POPULATIONS

Perhaps the most fascinating question about the evolution of complex traits in humans is whether they were subject to natural selection. Current methods to detect polygenic
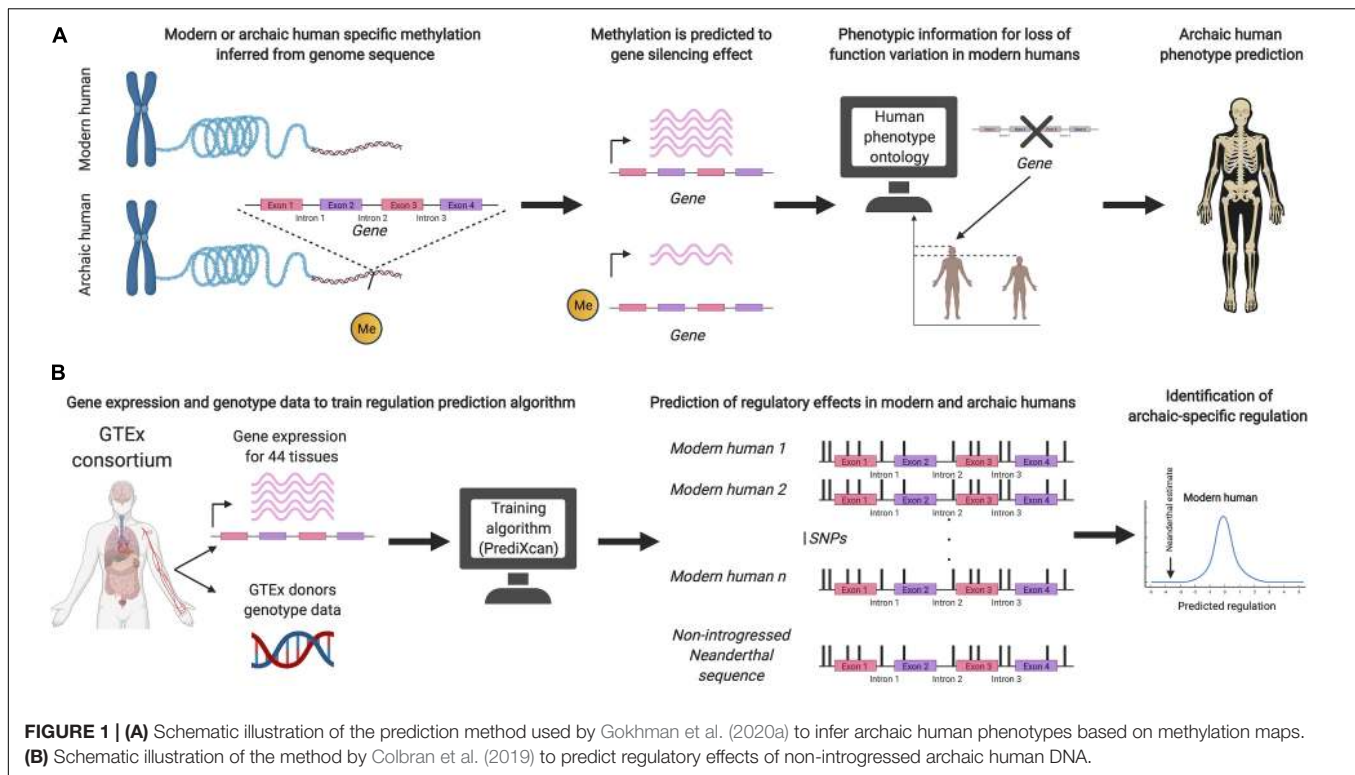
FIGURE 1 | (A) Schematic illustration of the prediction method used by Gokhman et al. (2020a) to infer archaic human phenotypes based on methylation maps. **(B)** Schematic illustration of the method by Colbran et al. (2019) to predict regulatory effects of non-introgressed archaic human DNA.
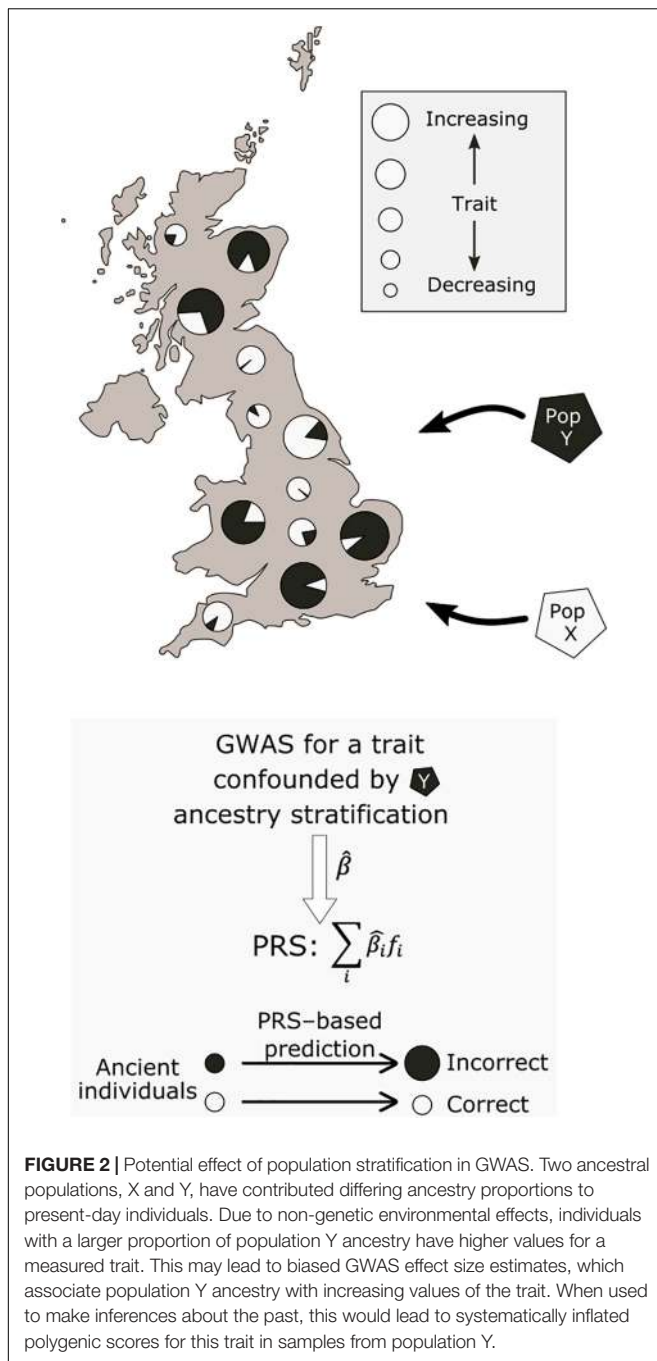
adaptation have mainly focused on present-day populations; using either differences between populations, or variation within them, to identify polygenic adaptation. For example, Berg and Coop (2014) developed a method that identifies over-dispersion of genetic values among populations, compared to a null distribution expected under a model of drift; which Racimo et al. (2018) extended to work with admixture graphs. Field et al. (2016) used the distribution of singletons around trait-associated SNPs, and Uricchio et al. (2019) used the joint distribution of variant effect sizes and derived allele frequencies (DAF). Whichever method is used, significant caveats must be addressed before attributing differences in such scores to polygenic adaptation (Novembre and Barton, 2018; Coop, 2019; Rosenberg et al., 2019). Most of these issues affect both present-day and ancient populations, but many are especially problematic when working with ancient genomes.

A prominently reported example of polygenic adaptation is that of selection for increasing height across a north-south gradient in Europe (Turchin et al., 2012; Berg and Coop, 2014; Robinson et al., 2015; Zoledziewska et al., 2015; Guo et al., 2018; Racimo et al., 2018; Berg et al., 2019b; Chen et al., 2020). Most studies which described this signal based their analyses on effect size estimates from the GIANT consortium, a GWAS meta-analysis encompassing 79 separate studies (Wood et al., 2014). Concerningly, follow-up work using the larger and more homogeneous UK Biobank cohort failed to replicate the signal of polygenic adaptation for height (Berg et al., 2019a; Sohail et al., 2019). A recent systematic comparison across a range of GWAS cohorts has further shown that the results of these tests are highly dependent on the ancestry composition of the

cohort used to obtain the effect size estimates (Refoyo-Martínez et al., 2021). These analyses showed that residual stratification in GWAS meta- and mega-analyses can result in inflated effect size estimates that, in turn, can lead to spurious signals of selection. The effects of this residual stratification may be exacerbated for ancient populations with non-uniform relatedness to present-day GWAS cohorts (see **Figure 2**).

Residual stratification is a major concern for GWAS, even among a relatively homogeneous cohort like the UK Biobank. Zaidi and Mathieson (2020) used simulations to show that fine-scale recent demography can confound GWAS which has been corrected for stratification using common variants only. Failure to adequately correct for localised population structure can lead to spurious associations between a trait and low-frequency variants that happen to be common in areas of atypical environmental effect. This finding is problematic as most GWAS have been conducted on either SNP array data, or on genomes imputed from SNP array data (Visscher et al., 2017). For example, GWAS summary statistics from the UK Biobank are based on imputed genomes (Bycroft et al., 2018). A limitation of this approach is that the accuracy of imputed genotypes are inversely correlated with the minor allele frequencies (MAF) of variants in the reference panel. Additionally, rare variants that are not segregating in the reference panel cannot be imputed at all. As a result, imputed genomes are specifically depleted in the rare variants needed to adjust for stratification from recent demography.

For large sample sizes, low-frequency variants (MAF ≤ 0.05) make a significant contribution to the heritability of many complex traits (Mancuso et al., 2016; Hartman et al., 2019), but

FIGURE 2 | Potential effect of population stratification in GWAS. Two ancestral populations, X and Y, have contributed differing ancestry proportions to present-day individuals. Due to non-genetic environmental effects, individuals with a larger proportion of population Y ancestry have higher values for a measured trait. This may lead to biased GWAS effect size estimates, which associate population Y ancestry with increasing values of the trait. When used to make inferences about the past, this would lead to systematically inflated polygenic scores for this trait in samples from population Y.

the role of rare variants is less well established. Both empirical and simulation studies have shown that for traits under either negative or stabilising selection, there is an inverse correlation between effect size and MAF (Simons et al., 2018; Schoech et al., 2019; Durvasula and Lohmueller, 2021). For the many traits thought to be under negative selection (O'Connor et al., 2019), large effect variants that are rare in present-day populations may have had higher allele frequencies in ancient populations due to selection. This makes polygenic scores for ancient individuals especially sensitive to bias from GWAS effect size estimates

ascertained from common variants only. Conversely, where present-day rare variants with large-effect sizes are known, higher frequencies in ancient populations would result in more accurate PRS predictions, due to their larger contribution to the overall genetic variance.

A recent analysis indicated that a substantial component of the unidentified heritability for anthropometric traits like height and BMI lies within large effect rare variants, some with MAF as low as 0.01% (Wainschtein et al., 2019). However, using GWAS to recover variant associations for SNPs as rare as this would require hundreds of thousands of whole-genomes, substantially exceeding the largest whole-genome GWAS published to date (e.g., Taliun et al., 2021). The consequence of this missing heritability may be particularly acute for trait prediction in ancient samples, as large-effect rare variants which contributed to variability in the past may no longer be segregating in present-day populations. Indeed, simulations suggest that the genetic architecture of complex traits is highly specific to each population, and that negative selection enriches for private variants, which contribute to a substantial component of the heritability of each trait (Durvasula and Lohmueller, 2021). Empirical studies have also identified that functionally important regions, including conserved and regulatory regions, are enriched for population-specific effect sizes, and that this pattern may have been driven by directional selection (Shi et al., 2021).

In addition to these issues, the majority of SNP associations inferred from GWAS are likely not the causal alleles. Instead, GWAS predominantly identifies SNPs which are in high linkage disequilibrium (LD) with causal alleles. Most GWAS also assume a model in which all complex trait heritability is additive and well tagged by SNPs segregating in the cohort; although some GWAS do include non-additive models (e.g., Guindo-Martínez et al., 2021). Consequently, effect size estimates are contingent on the LD structure of the cohort in which they were ascertained. Due to recombination, this LD structure decays through time, and is reshaped by the population history in which selection processes are embedded.

Over the last decade, paleogenomic studies have repeatedly demonstrated that the evolutionary histories of human populations are characterized by recurrent episodes of divergence, expansion, migration and admixture (reviewed in Pickrell and Reich, 2014; Skoglund and Mathieson, 2018). For example, in West Eurasia, four major ancestry groups have contributed to the majority of present-day genetic variation (Jones et al., 2015). As such, the LD structure of present-day British individuals—which underpins effect size estimates from the UK Biobank—was substantially different prior to the Bronze Age, when the most recent of these major admixture episodes occurred (Allentoft et al., 2015; Haak et al., 2015). To improve ancestral trait prediction, new methods which explicitly model the haplotype structure of both ancient populations and present-day GWAS cohorts are needed.

In aggregate, these issues combine to substantially diminish the portability of polygenic scores between populations. Indeed, in present-day populations, the predictive accuracy of PRS degrades approximately linearly with increasing genetic distance from the cohort used to ascertain the GWAS (Scutari et al., 2016;

Martin et al., 2017a, 2019; Kim et al., 2018; Bitarello and Mathieson, 2020; Mostafavi et al., 2020; Majara et al., 2021). Even within a single ancestry group, the correlation between PRS calculated from different discovery GWAS shows considerable variance (Schultz et al., 2021). However, the extent to which the issue of PRS portability also affects ancient populations, which are either partially or directly ancestral to the GWAS cohort, are yet to be determined.

In cases where a robust signal of polygenic adaptation can be identified, care must still be taken when interpreting which trait was actually subject to directional selection. Due to the highly polygenic nature of most complex traits, there is a high rate of genetic correlation between phenotypes (Shi et al., 2017; Ning et al., 2020). This can occur when correlated traits share causal alleles (i.e., pleiotropy) or where casual alleles are in high LD with each other. Consequently, selection acting on one specific trait can generate a spurious signal of polygenic adaptation for multiple genetically correlated traits. Recently, Stern et al. (2021) developed a method for conditional testing of polygenic adaptation to address this problem. When considered in a joint test, previously identified signals of selection for educational attainment and hair colour in British individuals were significantly attenuated by the signal of selection for skin pigmentation (Stern et al., 2021). However, this approach can only untangle genetic correlations between traits which have been measured in GWAS cohorts, leaving open the possibility that selection is acting on an unobserved yet correlated trait. Indeed, many GWAS traits are either coarse proxy measures with substantial socio-economic confounding (e.g., educational attainment), or narrow physiological measurements (e.g., levels of potassium in urine); neither of which are likely to have been direct targets of polygenic adaptation. In practice, the truly adaptive phenotype is rarely directly observable, and all measured traits are genetically correlated proxies at various levels of abstraction.

LIMITATIONS AND CAVEATS SPECIFIC TO ANCIENT DNA

In addition to all of the general issues and caveats discussed above, working with ancient DNA also involves a range of issues that are particular to the degraded nature of the data; such as post-mortem damage, generally low average sequence coverage, short fragment lengths, reference bias, and microbial and human contamination (Gilbert et al., 2005; Dabney et al., 2013; Renaud et al., 2019; Peyrégne and Prüfer, 2020). All of these factors affect our ability to correctly infer ancient genotypes; and therefore, to construct accurate polygenic scores or infer polygenic adaptation.

A common strategy for dealing with the low endogenous fraction of aDNA libraries is to use in-solution hybridisation capture to retrieve specific loci, or a set of predetermined SNPs (Avila-Arcos et al., 2011; Cruz-Dávalos et al., 2017). This approach has substantial advantages in on-target efficiency, at the cost of ascertainment bias. For example, in the case of the popular "1240k" capture array, targeted SNPs were predominantly ascertained in present-day individuals (Fu et al.,

2015; Haak et al., 2015). Consequently, an unknown fraction of the true ancestral variation is lost during capture. This is further exacerbated by the generally low coverage of most aDNA libraries; for which a common practice is to draw a read at random along each position in the genome, to infer "pseudo-haploid" genotypes. When used to compute polygenic scores for ancient populations, only a subset of GWAS variants can be used, which substantially reduces predictive accuracy. Cox et al. (2021) estimate that the combined effect of low-coverage and pseudo-haploid genotypes reduced their predictive accuracy by approximately 75%, when compared to present-day data.

An alternative approach is to perform low-coverage shotgun sequencing, followed by imputation, using a large reference panel (Ausmees et al., 2019; Hui et al., 2020). This has the dual advantages of reducing ascertainment bias and increasing the number of GWAS variants available to calculate polygenic scores. However, imputation itself introduces a new source of bias, particularly if the reference panel is not representative of the ancestries found in the low-coverage samples. Nevertheless, the level of imputation bias can be empirically estimated by downsampling high-coverage aDNA libraries and testing imputed genotypes against direct observations (e.g., Margaryan et al., 2020). Where a suitable reference panel exists, recently developed methods for imputation from low-coverage sequencing data (Davies et al., 2021; Rubinacci et al., 2021) show great promise for ancient DNA studies (e.g., Clemente et al., 2021).

Even under ideal conditions, in which exact polygenic scores for ancient populations are known *a priori*, interpreting differences in mean PRS between groups still requires careful consideration. For many polygenic traits, the variance between population means is lower than the variance within populations. As a result, differences in population level polygenic scores have limited predictive value for inferring the physiology or behaviour of individual people in the past. Genetics plays only a partial role in shaping phenotypic diversity, and differences in polygenic scores between individuals, or populations, does not automatically translate into differences in the expressed phenotype. Indeed, for some complex traits, an inverse correlation has been observed; in which polygenic scores have been steadily decreasing over recent decades, whilst the measured phenotype has been increasing [e.g., educational attainment (Kong et al., 2017; Abdellaoui et al., 2019)]. This highlights the substantial role of environmental variation in shaping phenotypic diversity. For ancient populations, we must also consider the wide variation in culture, diet, health, social organisation and climate which will have mediated any potential differences in population level polygenic scores. Furthermore, ancient populations are likely to have experienced a heterogeneous range of selective pressures. What we observe in present-day populations is not the result of a single directional process, but instead represents a mosaic of haplotypes which were shaped by different fitness landscapes, at varying levels of temporal depth.

Lastly, in most cases, we cannot directly observe phenotypes in the ancient individuals whose genomes have been studied. This greatly limits our ability to compare the genetically predicted value of a trait to its expressed phenotype, raising the

question: are predictions of most ancient phenotypes inherently unverifiable? For well-preserved traits, like standing height, there is considerable variability in estimates produced from different skeletal elements and between different studies (Cox et al., 2021; Marciniak et al., 2021). For traits that do not preserve well in the archaeological record, the prospects of validation are much poorer. These include not only soft tissue measurements (e.g., pigmentation or haemoglobin counts), but also personality and mental health traits that require an individual to be alive to be properly measured or diagnosed. Furthermore, some phenotypes are non-sensical outside of a modern context. Whilst it is possible to build a polygenic score for "time spent watching television" (UK Biobank code: 1070), it is not clear how to interpret any potential differences one might find between Mesolithic hunter-gatherers and Neolithic farmers. This problem extends more generally to all phenotypes which have strong gene–environment interactions, in which the expression of the trait may have been substantively different in the past due to diverse environmental conditions (e.g., the interplay between BMI and diet).

PROSPECTS FOR THE FUTURE

The growth in the number of ancient genomes currently shows little signs of slowing, nor does the increasing availability of gene-trait association data. Predictably, efforts to perform trait predictions in ancient individuals will also continue to grow. We believe that increased emphasis on limitations and caveats in the way we study and communicate these findings will enable a better understanding of what we can and cannot predict with existing models.

As a working assumption, polygenic scores from any single GWAS should be considered unreliable in an ancient trait reconstruction analysis. Researchers should only trust observed signals of trait evolution if those patterns hold across multiple independent GWAS (e.g., Chen et al., 2020), and preferably where each of these GWAS has been performed on a large cohort with homogeneous ancestry (Refoyo-Martínez et al., 2021).

We also need to better understand how well GWAS effect size estimates, ascertained in present-day populations, generalise to ancient populations that are only partially ancestral to the GWAS cohort. One approach to this would be to use simulations, under a plausible demographic scenario, to explore how the predictive accuracy of PRS degrades through time and across the boundaries of major ancestral migrations.

Traits that are preserved in the fossil record can provide a degree of partial benchmarking (Cox et al., 2019, 2021); however, the genetic components of variation are often only partially explained by polygenic scores, and environmental components almost always play large roles in expressed trait variation, often dwarfing the contribution of polygenic scores. Furthermore, only a few—largely osteological—traits are well preserved over time, so these comparisons will always be limited in scope.

That being said, there are several promising avenues of research that could serve to improve genetic trait prediction in ancient populations. An existing approach to improve the portability of PRS across ancestries is to prioritise variants with predicted functional roles (Amariuta et al., 2020; Weissbrod et al., 2020). This approach aims to improve PRS portability in present-day populations by reducing the fraction of spurious associations due to the cohort specific LD structure of the GWAS reference panel. Another promising approach is to jointly model PRS using GWAS summary statistics from multiple populations (Márquez-Luna et al., 2017; Ruan et al., 2021; Turley et al., 2021). By including information from genetically distant groups, these methods can account for the variance in effect sizes inferred between GWAS cohorts. This multi-ancestry approach holds particular promise for ancient populations, as it may help to identify variant associations which are segregating in only a subset of present-day populations, but which were more widespread in the past.

These studies also underscore the importance of studying the ancestral haplotype backgrounds on which beneficial, deleterious or neutral alleles spread. Recent studies have shown that tests of selection on individual loci can gain power by explicitly modelling patterns of ancestry across the genome (Pierron et al., 2018; Hamid et al., 2021). Strong selective signals might be masked by post-selection admixture processes, but might become evident once the ancestry of the selected haplotypes is explicitly modelled (Souilmi et al., 2020). This phenomenon is also likely to affect polygenic adaptation studies, particularly when the degree of correlation between genetic score differences and differences in ancestral haplotype backgrounds is expected to be high, for example, after admixture between populations that have been evolving in isolation for long periods of time.

A promising avenue of research is developing around new methods for approximately inferring ancestral recombination graphs (ARG) via the construction of tree sequences (Kelleher et al., 2019; Speidel et al., 2019), which have recently been extended to incorporate non-contemporaneous sampling (Speidel et al., 2021; Wohns et al., 2021). An ARG is a model which contains a detailed description of the genealogical relationships in a set of samples, including the full history of gene trees, ancestral haplotypes and recombination events which relate the samples to each other at every site in the genome (Griffiths and Marjoram, 1997). One potential advantage of an ARG is that it may be used to help mitigate issues with the portability of polygenic scores. By building an ARG composed of both ancient samples and the present-day cohorts used to ascertain the GWAS associations, one could potentially determine which haplotypes are shared between the GWAS cohort and the ancient populations; thereby reducing effect size bias in populations that are only partially ancestral to the GWAS cohort.

Another area in which ancient genomes offer unique potential is in detecting polygenic adaptation in response to environmental change. The time-series nature of ancient genomes provides the potential for the incorporation of paleoclimate reconstructions (e.g., Brown et al., 2018) into tests of polygenic adaptation, in a manner that is not possible with present-day data alone.

Ultimately, the ancient genomics community must come to terms with the limitations of genetic hindcasting. Ancient

genomes provide an unprecedented window into our past, but this window is often blurry and distorted. There is still a lot of information waiting to be obtained from ancient DNA, and some of the blurriness might ultimately come into focus as computational methods continue to improve. But we must also accept the fact that many aspects of past human biology—including physical characteristics and disease susceptibility—might be irrevocably lost to the tides of history. Ancient genome sequences are, after all, molecular fossils: imperfect and degraded records of lives that ceased to exist long ago.

AUTHOR CONTRIBUTIONS

REFERENCES

Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K. E., Nivard, M. G., Veul, L., et al. (2019). Genetic correlates of social stratification in Great Britain. *Nat. Hum. Behav.* 3, 1332–1342. doi: 10.1038/s41562-019-0757-5

Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172. doi: 10.1038/nature14507

Amariuta, T., Ishigaki, K., Sugishita, H., Ohta, T., Koido, M., Dey, K. K., et al. (2020). Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nat. Genet.* 52, 1346–1354. doi: 10.1038/s41588-020-00740-8

Ausmees, K., Sanchez-Quinto, F., Jakobsson, M., and Nettelblad, C. (2019). *An empirical evaluation of genotype imputation of ancient DNA*. Available Online at: http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A1367434&dswid= 5303 [Accessed January 16, 2021]

Avila-Arcos, M. C., Cappellini, E., Romero-Navarro, J. A., Wales, N., Moreno-Mayar, J. V., Rasmussen, M., et al. (2011). Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.* 1:74. doi: 10.1038/srep00074

Berg, J. J., and Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genet.* 10:e1004412. doi: 10.1371/journal.pgen.1004412

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., et al. (2019a). Reduced signal for polygenic adaptation of height in UK Biobank. *Elife* 8:47. doi: 10.7554/eLife.39725

Berg, J. J., Zhang, X., and Coop, G. (2019b). Polygenic adaptation has impacted multiple anthropometric traits. *bioRxiv* 2019:167551. doi: 10.1101/167551

Bitarello, B. D., and Mathieson, I. (2020). Polygenic scores for height in admixed populations. *G3* 10, 4027–4036. doi: 10.1534/g3.120.401658

Brace, S., Diekmann, Y., Booth, T. J., van Dorp, L., Faltyskova, Z., Rohland, N., et al. (2019). Ancient genomes indicate population replacement in Early Neolithic Britain. *Nat. Ecol. Evol.* 3, 765–771. doi: 10.1038/s41559-019-0871-9

Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C., and Haywood, A. M. (2018). PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci. Data* 5:180254. doi: 10.1038/sdata.2018.254

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi: 10.1093/nar/gky1120

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209. doi: 10.1038/s41586-018-0579-z

FUNDING

ACKNOWLEDGMENTS

Chen, F., Welker, F., Shen, C.-C., Bailey, S. E., Bergmann, I., Davis, S., et al. (2019). A late middle pleistocene denisovan mandible from the Tibetan Plateau. *Nature* 569, 409–412. doi: 10.1038/s41586-019-1139-x

Chen, M., Sidore, C., Akiyama, M., Ishigaki, K., Kamatani, Y., Schlessinger, D., et al. (2020). Evidence of polygenic adaptation in sardinia at height-associated loci ascertained from the Biobank Japan. *Am. J. Hum. Genet.* 107, 60–71. doi: 10.1016/j.ajhg.2020.05.014

Clemente, F., Unterländer, M., Dolgova, O., Amorim, C. E. G., Coroado-Santos, F., Neuenschwander, S., et al. (2021). The genomic history of the Aegean palatial civilizations. *Cell* 184, 2565–2586. doi: 10.1016/j.cell.2021.03.039

Colbran, L. L., Gamazon, E. R., Zhou, D., Evans, P., Cox, N. J., and Capra, J. A. (2019). Inferred divergent gene regulation in archaic hominins reveals potential phenotypic differences. *Nat. Ecol. Evol.* 3, 1598–1606. doi: 10.1038/s41559-019-0996-x

Coop, G. (2019). *Reading tea leaves? Polygenic scores and differences in traits among groups*. Available Online at: http://arxiv.org/abs/1909.00892

Cox, S. L., Moots, H., Stock, J. T., Shbat, A., Bitarello, B. D., Haak, W., et al. (2021). Predicting skeletal stature using ancient DNA. *bioRxiv* 2021:437877. doi: 10.1101/2021.03.31.437877

Cox, S. L., Ruff, C. B., Maier, R. M., and Mathieson, I. (2019). Genetic contributions to variation in human stature in prehistoric Europe. *Proc. Natl. Acad. Sci. U. S. A.* 116, 21484–21492. doi: 10.1073/pnas.1910606116

Cruz-Dávalos, D. I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., et al. (2017). Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* 17, 508–522. doi: 10.1111/1755-0998.12595

Dabney, J., Meyer, M., and Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harb. Perspect. Biol.* 5:a012567. doi: 10.1101/cshperspect.a012567

Dannemann, M. (2021). The population-specific impact of Neandertal introgression on human disease. *Genome Biol. Evol.* 13:evaa250. doi: 10.1093/gbe/evaa250

Dannemann, M., Andrés, A. M., and Kelso, J. (2016). Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human Toll-like receptors. *Am. J. Hum. Genet.* 98, 22–33. doi: 10.1016/j.ajhg.2015.11.015

Dannemann, M., and Gallego Romero, I. (2021). Harnessing pluripotent stem cells as models to decipher human evolution. *FEBS J.* 2021:15885. doi: 10.1111/febs.15885

Dannemann, M., He, Z., Heide, C., Vernot, B., Sidow, L., Kanton, S., et al. (2020). Human stem cell resources are an inroad to neandertal DNA functions. *Stem Cell Rep.* 15, 214–225. doi: 10.1016/j.stemcr.2020.05.018

Dannemann, M., and Kelso, J. (2017). The contribution of neanderthals to phenotypic variation in modern humans. *Am. J. Hum. Genet.* 101, 578–589. doi: 10.1016/j.ajhg.2017.09.010

Davies, R. W., Kucka, M., Su, D., Shi, S., Flanagan, M., Cunniff, C. M., et al. (2021). Rapid genotype imputation from sequence with reference panels. *Nat. Genet.* 53, 1–8. doi: 10.1038/s41588-021-00877-0

Dehasque, M., Ávila-Arcos, M. C., Díez-Del-Molino, D., Fumagalli, M., Guschanski, K., Lorenzen, E. D., et al. (2020). Inference of natural selection from ancient DNA. *Evol. Lett.* 4, 94–108. doi: 10.1002/evl3.165

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 9:e1003348. doi: 10.1371/journal.pgen.1003348

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., et al. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* 10, 1–9. doi: 10.1038/s41467-019-11112-0

Durvasula, A., and Lohmueller, K. E. (2021). Negative selection on complex traits limits phenotype prediction accuracy between populations. *Am. J. Hum. Genet.* 108, 620–631. doi: 10.1016/j.ajhg.2021.02.013

Eiberg, H., Troelsen, J., Nielsen, M., Mikkelsen, A., Mengel-From, J., Kjaer, K. W., et al. (2008). Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum. Genet.* 123, 177–187. doi: 10.1007/s00439-007-0460-x

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton, K. J., Golan, D., et al. (2016). Detection of human adaptation during the past 2000 years. *Science* 354, 760–764. doi: 10.1126/science.aag0776

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, 216–219. doi: 10.1038/nature14558

Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M., and Barnes, I. (2005). Assessing ancient DNA studies. *Trends Ecol. Evol.* 20, 541–544. doi: 10.1016/j.tree.2005.07.005

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. doi: 10.1016/j.cub.2016.10.041

Gokhman, D., Mishol, N., de Manuel, M., de Juan, D., Shuqrun, J., Meshorer, E., et al. (2020a). Reconstructing denisovan anatomy using DNA methylation maps. *Cell* 180:601. doi: 10.1016/j.cell.2020.01.020

Gokhman, D., Nissim-Rafinia, M., Agranat-Tamir, L., Housman, G., García-Pérez, R., Lizano, E., et al. (2020b). Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat. Commun.* 11:1189. doi: 10.1038/s41467-020-15020-6

González-Fortes, G., Jones, E. R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., et al. (2017). Paleogenomic evidence for multi-generational mixing between neolithic farmers and mesolithic hunter-gatherers in the lower danube basin. *Curr. Biol.* 27, 1801–1810. doi: 10.1016/j.cub.2017.05.023

Griffiths, R. C., and Marjoram, P. (1997). An ancestral recombination graph. *Instit. Mathemat. Appl.* 87:257.

Guindo-Martínez, M., Amela, R., Bonàs-Guarch, S., Puiggròs, M., Salvoro, C., Miguel-Escalada, I., et al. (2021). The impact of non-additive genetic associations on age-related complex diseases. *Nat. Commun.* 12:2436. doi: 10.1038/s41467-021-21952-4

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., et al. (2018). Population genomics of mesolithic scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* 16:e2003703. doi: 10.1371/journal.pbio.2003703

Gunz, P., Tilot, A. K., Wittfeld, K., Teumer, A., Shapland, C. Y., van Erp, T. G. M., et al. (2019). Neandertal introgression sheds light on modern human endocranial globularity. *Curr. Biol.* 29, 120–127. doi: 10.1016/j.cub.2018.10.065

Guo, J., Wu, Y., Zhu, Z., Zheng, Z., Trzaskowski, M., Zeng, J., et al. (2018). Global genetic differentiation of complex traits shaped by natural selection in humans. *Nat. Commun.* 9:1865. doi: 10.1038/s41467-018-04191-y

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211. doi: 10.1038/nature14317

Hamid, I., Korunes, K. L., Beleza, S., and Goldberg, A. (2021). Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. *Elife* 10:e63177. doi: 10.7554/eLife.63177

Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., et al. (2008). A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* 4:e1000074. doi: 10.1371/journal.pgen.1000074

Harris, K., and Nielsen, R. (2016). The genetic cost of neanderthal introgression. *Genetics* 203, 881–891. doi: 10.1534/genetics.116.186890

Hartman, K. A., Rashkin, S. R., Witte, J. S., and Hernandez, R. D. (2019). Imputed genomic data reveals a moderate effect of low frequency variants to the heritability of complex human traits. *bioRxiv* 2019:879916. doi: 10.1101/2019.12.18.879916

Hider, J. L., Gittelman, R. M., Shah, T., Edwards, M., Rosenbloom, A., Akey, J. M., et al. (2013). Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol. Biol.* 13:150. doi: 10.1186/1471-2148-13-150

Huerta-Sánchez, E., Jin, X., Asan Bianba, Z., Peter, B. M., Vinckenbosch, N., et al. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512, 194–197. doi: 10.1038/nature13408

Hui, R., D'Atanasio, E., Cassidy, L. M., Scheib, C. L., and Kivisild, T. (2020). Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Sci. Rep.* 10:18542. doi: 10.1038/s41598-020-75387-w

Jensen, T. Z. T., Niemann, J., Iversen, K. H., Fotakis, A. K., Gopalakrishnan, S., Vågene, ÅJ., et al. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nat. Commun.* 10:5520. doi: 10.1038/s41467-019-13549-9

Jones, E. R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* 6:8912. doi: 10.1038/ncomms9912

Ju, D., and Mathieson, I. (2021). The evolution of skin pigmentation-associated variation in West Eurasia. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2009227118. doi: 10.1073/pnas.2009227118

Juric, I., Aeschbacher, S., and Coop, G. (2016). The strength of selection against neanderthal introgression. *PLoS Genet.* 12:e1006340. doi: 10.1371/journal.pgen.1006340

Kelleher, J., Wong, Y., Wohns, A. W., Fadil, C., Albers, P. K., and McVean, G. (2019). Inferring whole-genome histories in large population datasets. *Nat. Genet.* 51, 1330–1338. doi: 10.1038/s41588-019-0483-y

Khrameeva, E. E., Bozek, K., He, L., Yan, Z., Jiang, X., Wei, Y., et al. (2014). Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nat. Commun.* 5:3584. doi: 10.1038/ncomms4584

Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J., and Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome Biol.* 19:179. doi: 10.1186/s13059-018-1561-7

Knowles, J. W., and Ashley, E. A. (2018). Cardiovascular disease: The rise of the genetic risk score. *PLoS Med.* 15:e1002546. doi: 10.1371/journal.pmed.1002546

Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., et al. (2017). Selection against variants in the genome associated with educational attainment. *Proc. Natl. Acad. Sci. U. S. A.* 114, E727–E732. doi: 10.1073/pnas.1612113114

Lao, O., de Gruijter, J. M., van Duijn, K., Navarro, A., and Kayser, M. (2007). Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann. Hum. Genet.* 71, 354–369. doi: 10.1111/j.1469-1809.2006.00341.x

Liu, F., Visser, M., Duffy, D. L., Hysi, P. G., Jacobs, L. C., Lao, O., et al. (2015). Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Hum. Genet.* 134, 823–835. doi: 10.1007/s00439-015-1559-0

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45, D896–D901. doi: 10.1093/nar/gkw1133

Majara, L., Kalungi, A., Koen, N., Zar, H., Stein, D. J., Kinyanda, E., et al. (2021). Low generalizability of polygenic scores in African populations due to genetic and environmental diversity. *bioRxiv* 2021:426453. doi: 10.1101/2021.01.12.426453

Malaspinas, A.-S. (2016). Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. *Mol. Ecol.* 25, 24–41. doi: 10.1111/mec.13492

Mancuso, N., Rohland, N., Rand, K. A., Tandon, A., Allen, A., Quinque, D., et al. (2016). The contribution of rare variation to prostate cancer heritability. *Nat. Genet.* 48, 30–35. doi: 10.1038/ng.3446

Marciniak, S., Bergey, C. M., Silva, A. M., Hałuszko, A., Furmanek, M., Veselka, B., et al. (2021). An integrative skeletal and paleogenomic analysis of prehistoric stature variation suggests relatively reduced health for early European farmers. *bioRxiv* 2021:437881. doi: 10.1101/2021.03.31.437881

Margaryan, A., Lawson, D. J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., et al. (2020). Population genomics of the Viking world. *Nature* 585, 390–396. doi: 10.1038/s41586-020-2688-8

Márquez-Luna, C., Loh, P.-R., South Asian Type 2 Diabetes (SAT2D) Consortium, Sigma Type 2 Diabetes Consortium, and Price, A. L. (2017). Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* 41, 811–823. doi: 10.1002/gepi.22083

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., et al. (2017a). Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* 100, 635–649. doi: 10.1016/j.ajhg.2017.03.004

Martin, A. R., Lin, M., Granka, J. M., Myrick, J. W., Liu, X., Sockell, A., et al. (2017b). An unexpectedly complex architecture for skin pigmentation in Africans. *Cell* 171, 1340–1353. doi: 10.1016/j.cell.2017.11.015

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., and Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591. doi: 10.1038/s41588-019-0379-x

Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., et al. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet.* 13:e1006852. doi: 10.1371/journal.pgen.1006852

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499–503. doi: 10.1038/nature16152

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222–226. doi: 10.1126/science.1224344

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., and Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *Elife* 9:e48376. doi: 10.7554/eLife.48376

Ning, Z., Pawitan, Y., and Shen, X. (2020). High-definition likelihood inference of genetic correlations across human complex traits. *Nat. Genet.* 52, 859–864. doi: 10.1038/s41588-020-0653-y

Novembre, J., and Barton, N. H. (2018). Tread lightly interpreting polygenic tests of selection. *Genetics* 208, 1351–1355. doi: 10.1534/genetics.118.300786

O'Connor, L. J., Schoech, A. P., Hormozdiari, F., Gazal, S., Patterson, N., and Price, A. L. (2019). Extreme polygenicity of complex traits is explained by negative selection. *Am. J. Hum. Genet.* 105, 456–476. doi: 10.1016/j.ajhg.2019.07.003

Olalde, I., Allentoft, M. E., Sánchez-Quinto, F., Santpere, G., Chiang, C. W. K., DeGiorgio, M., et al. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* 507, 225–228. doi: 10.1038/nature12960

Perry, G. H., Kistler, L., Kelaita, M. A., and Sams, A. J. (2015). Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. *J. Hum. Evol.* 79, 55–63. doi: 10.1016/j.jhevol.2014.10.018

Petr, M., Pääbo, S., Kelso, J., and Vernot, B. (2019). Limits of long-term selection against Neandertal introgression. *Proc. Natl. Acad. Sci. U. S. A.* 116, 1639–1644. doi: 10.1073/pnas.1814338116

Peyrégne, S., and Prüfer, K. (2020). Present-Day DNA contamination in ancient DNA datasets. *BioEssays* 42:2000081. doi: 10.1002/bies.202000081

Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., et al. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 19, 826–837. doi: 10.1101/gr.087577.108

Pickrell, J. K., and Reich, D. (2014). Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet.* 30, 377–389. doi: 10.1016/j.tig.2014.07.007

Pierron, D., Heiske, M., Razafindrazaka, H., Pereda-Loth, V., Sanchez, J., Alva, O., et al. (2018). Strong selection during the last millennium for African ancestry in the admixed population of Madagascar. *Nat. Commun.* 9:932. doi: 10.1038/s41467-018-03342-5

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., et al. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* 358, 655–658. doi: 10.1126/science.aao1887

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43–49. doi: 10.1038/nature12886

Quach, H., Rotival, M., Pothlichet, J., Loh, Y.-H. E., Dannemann, M., Zidane, N., et al. (2016). Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell* 167, 643–656. doi: 10.1016/j.cell.2016.09.024

Quillen, E. E., Norton, H. L., Parra, E. J., Lona-Durazo, F., Ang, K. C., Illiescu, F. M., et al. (2019). Shades of complexity: New perspectives on the evolution and genetic architecture of human skin. *Am. J. Phys. Anthropol.* 168, 4–26. doi: 10.1002/ajpa.23737

Racimo, F., Berg, J. J., and Pickrell, J. K. (2018). Detecting polygenic adaptation in admixture graphs. *Genetics* 208, 1565–1584. doi: 10.1534/genetics.117.300489

Racimo, F., Gokhman, D., Fumagalli, M., Ko, A., Hansen, T., Moltke, I., et al. (2017a). Archaic adaptive introgression in TBX15/WARS2. *Mol. Biol. Evol.* 34, 509–524. doi: 10.1093/molbev/msw283

Racimo, F., Marnetto, D., and Huerta-Sánchez, E. (2017b). Signatures of archaic adaptive introgression in present-day human populations. *Mol. Biol. Evol.* 34, 296–317. doi: 10.1093/molbev/msw216

Refoyo-Martínez, A., Liu, S., Jørgensen, A. M., Jin, X., Albrechtsen, A., Martin, A. R., et al. (2021). How robust are cross-population signatures of polygenic adaptation in humans? *bioRxiv* 2020:200030. doi: 10.1101/2020.07.13.200030

Renaud, G., Schubert, M., Sawyer, S., and Orlando, L. (2019). Authentication and assessment of contamination in ancient DNA. *Methods Mol. Biol.* 1963, 163–194. doi: 10.1007/978-1-4939-9176-1_17

Robinson, M. R., Hemani, G., Medina-Gomez, C., Mezzavilla, M., Esko, T., Shakhbazov, K., et al. (2015). Population genetic differentiation of height and body mass index across Europe. *Nat. Genet.* 47, 1357–1362. doi: 10.1038/ng.3401

Rocha, J. (2020). The evolutionary history of human skin pigmentation. *J. Mol. Evol.* 88, 77–87. doi: 10.1007/s00239-019-09902-7

Rosenberg, N. A., Edge, M. D., Pritchard, J. K., and Feldman, M. W. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evol. Med. Public Health* 2019, 26–34. doi: 10.1093/emph/eoy036

Ruan, Y., Anne Feng, Y.-C., Chen, C.-Y., Lam, M., Sawa, A., Martin, A. R., et al. (2021). Improving polygenic prediction in ancestrally diverse populations. *bioRxiv* 2020:20248738. doi: 10.1101/2020.12.27.20248738

Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* 53, 120–126. doi: 10.1038/s41588-020-00756-0

Ruff, C. B., Holt, B. M., Niskanen, M., Sladék, V., Berner, M., Garofalo, E., et al. (2012). Stature and body mass estimation from skeletal remains in the European Holocene. *Am. J. Phys. Anthropol.* 148, 601–617. doi: 10.1002/ajpa.22087

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. doi: 10.1038/nature06250

Sams, A. J., Dumaine, A., Nédélec, Y., Yotova, V., Alfieri, C., Tanner, J. E., et al. (2016). Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. *Genome Biol.* 17:246. doi: 10.1186/s13059-016-1098-6

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., et al. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507, 354–357. doi: 10.1038/nature12961

Sankararaman, S., Mallick, S., Patterson, N., and Reich, D. (2016). The combined landscape of denisovan and neanderthal ancestry in present-day humans. *Curr. Biol.* 26, 1241–1247. doi: 10.1016/j.cub.2016.03.037

Sankararaman, S., Patterson, N., Li, H., Pääbo, S., and Reich, D. (2012). The date of interbreeding between neandertals and modern humans. *PLoS Genet.* 8:e1002947. doi: 10.1371/journal.pgen.1002947

Sawyer, S., Renaud, G., Viola, B., Hublin, J.-J., Gansauge, M.-T., Shunkov, M. V., et al. (2015). Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl. Acad. Sci. U. S. A.* 112, 15696–15700. doi: 10.1073/pnas.1519905112

Schoech, A. P., Jordan, D. M., Loh, P.-R., Gazal, S., O'Connor, L. J., Balick, D. J., et al. (2019). Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun.* 10:790. doi: 10.1038/s41467-019-08424-6

Schultz, L. M., Merikangas, A. K., Ruparel, K., Jacquemont, S., Glahn, D. C., Gur, R. E., et al. (2021). Stability of polygenic scores across discovery genome-wide association studies. *bioRxiv* 2021:449060. doi: 10.1101/2021.06.18.449060

Scutari, M., Mackay, I., and Balding, D. (2016). Using genetic distance to infer the accuracy of genomic prediction. *PLoS Genet.* 12:e1006288. doi: 10.1371/journal.pgen.1006288

Shi, H., Gazal, S., Kanai, M., Koch, E. M., Schoech, A. P., Siewert, K. M., et al. (2021). Population-specific causal disease effect sizes in functionally important regions

impacted by selection. *Nat. Commun.* 12:1098. doi: 10.1038/s41467-021-21 286-1

Shi, H., Mancuso, N., Spendlove, S., and Pasaniuc, B. (2017). Local genetic correlation gives insights into the shared genetic architecture of complex traits. *Am. J. Hum. Genet.* 101, 737–751. doi: 10.1016/j.ajhg.2017.09.022

Simcoe, M., Valdes, A., Liu, F., Furlotte, N. A., Evans, D. M., Hemani, G., et al. (2021). Genome-wide association study in almost 195,000 individuals identifies 50 previously unidentified genetic loci for eye color. *Sci. Adv.* 7:eabd1239. doi: 10.1126/sciadv.abd1239

Simons, Y. B., Bullaughey, K., Hudson, R. R., and Sella, G. (2018). A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.* 16:e2002985. doi: 10.1371/journal.pbio.2002985

Simonti, C. N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D. S., Chisholm, R. L., et al. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. *Science* 351, 737–741. doi: 10.1126/science.aad2149

Skoglund, P., and Mathieson, I. (2018). Ancient genomics of modern humans: The first decade. *Annu. Rev. Genom. Hum. Genet.* 19, 381–404. doi: 10.1146/annurev-genom-083117-021749

Skov, L., Coll Macià, M., Sveinbjörnsson, G., Mafessoni, F., Lucotte, E. A., Einarsdóttir, M. S., et al. (2020). The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* 582, 78–83. doi: 10.1038/s41586-020-2225-9

Slon, V., Viola, B., Renaud, G., Gansauge, M.-T., Benazzi, S., Sawyer, S., et al. (2017). A fourth denisovan individual. *Sci. Adv.* 3:e1700186. doi: 10.1126/sciadv.1700186

Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., et al. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *Elife* 8:e39702. doi: 10.7554/eLife.39702

Souilmi, Y., Tobler, R., Johar, A., Williams, M., Grey, S. T., Schmidt, J., et al. (2020). Ancient human genomes reveal a hidden history of strong selection in Eurasia. *bioRxiv* 2020:021006. doi: 10.1101/2020.04.01.021006

Speidel, L., Cassidy, L., Davies, R. W., Hellenthal, G., Skoglund, P., and Myers, S. R. (2021). Inferring population histories for ancient genomes using genome-wide genealogies. *bioRxiv* 2021:431573. doi: 10.1101/2021.02.17.431573

Speidel, L., Forest, M., Shi, S., and Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* 51, 1321–1329. doi: 10.1038/s41588-019-0484-x

Stern, A. J., Speidel, L., Zaitlen, N. A., and Nielsen, R. (2021). Disentangling selection on genetically correlated polygenic traits via whole-genome genealogies. *Am. J. Hum. Genet.* 108, 219–239. doi: 10.1016/j.ajhg.2020.12.005

Sturm, R. A., Duffy, D. L., Zhao, Z. Z., Leite, F. P. N., Stark, M. S., Hayward, N. K., et al. (2008). A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am. J. Hum. Genet.* 82, 424–431. doi: 10.1016/j.ajhg.2007.11.005

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., et al. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* 39, 1443–1452. doi: 10.1038/ng.2007.13

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299. doi: 10.1038/s41586-021-03205-y

Trujillo, C. A., Rice, E. S., Schaefer, N. K., Chaim, I. A., Wheeler, E. C., Madrigal, A. A., et al. (2021). Reintroduction of the archaic variant of NOVA1 in cortical organoids alters neurodevelopment. *Science* 371:eaax2537. doi: 10.1126/science.aax2537

Turchin, M. C., Genetic Investigation of ANthropometric Traits (Giant) Consortium, Chiang, C. W. K., Palmer, C. D., Sankararaman, S., Reich, D., et al. (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44, 1015–1019. doi: 10.1038/ng.2368

Turley, P., Martin, A. R., Goldman, G., Li, H., Kanai, M., Walters, R. K., et al. (2021). Multi-Ancestry Meta-Analysis yields novel genetic discoveries and ancestry-specific associations. *bioRxiv* 2021:441003. doi: 10.1101/2021.04.23.441003

Uricchio, L. H., Kitano, H. C., Gusev, A., and Zaitlen, N. A. (2019). An evolutionary compass for detecting signals of polygenic selection and mutational bias. *Evol. Lett.* 3, 69–79. doi: 10.1002/evl3.97

Vernot, B., and Akey, J. M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. *Science* 343, 1017–1021. doi: 10.1126/science.1245938

Vernot, B., Tucci, S., Kelso, J., Schraiber, J. G., Wolf, A. B., Gittelman, R. M., et al. (2016). Excavating neandertal and denisovan DNA from the genomes of Melanesian individuals. *Science* 352, 235–239. doi: 10.1126/science.aad9416

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 Years of GWAS discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* 101, 5–22. doi: 10.1016/j.ajhg.2017.06.005

Wainschtein, P., Jain, D. P., Yengo, L., Zheng, Z., TOPMed Anthropometry Working Group, Trans-Omics for Precision Medicine Consortium, et al. (2019). Recovery of trait heritability from whole genome sequence data. *bioRxiv* 2019:588020. doi: 10.1101/588020

Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., et al. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* 52, 1355–1363. doi: 10.1038/s41588-020-00735-5

Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M., et al. (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4832–4837. doi: 10.1073/pnas.1316513111

Wohns, A. W., Wong, Y., Jeffery, B., Akbari, A., Mallick, S., Pinhasi, R., et al. (2021). A unified genealogy of modern and ancient genomes. *bioRxiv* 2021:431497. doi: 10.1101/2021.02.16.431497

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46, 1173–1186. doi: 10.1038/ng.3097

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., Lee, S. H., et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* 47, 1114–1120. doi: 10.1038/ng.3390

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* 27, 3641–3649. doi: 10.1093/hmg/ddy271

Zaidi, A. A., and Mathieson, I. (2020). Demographic history mediates the effect of stratification on polygenic scores. *Elife* 9:e61548. doi: 10.7554/eLife.61548

Zeberg, H., Dannemann, M., Sahlholm, K., Tsuo, K., Maricic, T., Wiebe, V., et al. (2020a). A Neanderthal sodium channel increases pain sensitivity in present-day humans. *Curr. Biol.* 30, 3465–3469. doi: 10.1016/j.cub.2020.06.045

Zeberg, H., Kelso, J., and Pääbo, S. (2020b). The neandertal progesterone receptor. *Mol. Biol. Evol.* 37, 2655–2660. doi: 10.1093/molbev/msaa119

Zeberg, H., and Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 587, 610–612. doi: 10.1038/s41586-020-2818-3

Zeberg, H., and Pääbo, S. (2021). A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. *Proc. Natl. Acad. Sci. U. S. A.* 2021:118. doi: 10.1073/pnas.2026309118

Zoledziewska, M., UK10K Consortium, Sidore, C., Chiang, C. W. K., Sanna, S., Mulas, A., et al. (2015). Height-reducing variants and selection for short stature in Sardinia. *Nat. Genet.* 47, 1352–1356. doi: 10.1038/ng.3403